



Universidad de Oviedo
Departamento de Matemáticas

**Análisis de la incertidumbre en
problemas inversos geofísicos no
lineales de alta dimensión**

*Uncertainty analysis in nonlinear,
high-dimensional geophysical
inverse problems*

Tesis Doctoral

José Luis García Pallero

PROGRAMA DE DOCTORADO EN MATEMÁTICAS Y ESTADÍSTICA

Programa de doctorado en Matemáticas y Estadística

Departamento de Matemáticas de la Universidad de Oviedo

Análisis de la incertidumbre en problemas inversos geofísicos no lineales de alta dimensión

Uncertainty analysis in nonlinear, high-dimensional geophysical inverse problems

*Memoria presentada por **José Luis García Pallero** en el Departamento de Matemáticas para optar al grado de Doctor por la Universidad de Oviedo, bajo la dirección del **Dr. D. Juan Luis Fernández Martínez**.*

Miembros titulares del tribunal:

Presidente **Dr. Sylvain Bonvalot**, Laboratoire GET, BGI, Université de Toulouse
Secretario **Dr. Paulino José García Nieto**, Universidad de Oviedo
Vocal **Dr. José Jaime Gómez Hernández**, Universitat Politècnica de València

Miembros suplentes del tribunal:

Presidenta **Dra. Colette Sirieix**, I2M, Université de Bordeaux
Secretaria **Dra. María Zulima Fernández Muñiz**, Universidad de Oviedo
Vocal **Dr. José Durany Castrillo**, Universidade de Vigo

Lugar y fecha de defensa: Oviedo, 28 de julio de 2016

Este trabajo está dedicado a todos los que fueron mis profesores en el Colegio Público de Tremañes (Gijón), en especial a José Alberto Villanueva Álvarez y a José Luis Rodríguez Montes.

Agradecimientos

Si bien en apariencia el apartado de agradecimientos es el de más fácil escritura, doy fe de que es sólo eso: mera fachada. Sí que afirmo que es el que más satisfacción deja, lo cual compensa todo lo demás. Y es que no es tarea sencilla la de concentrar en unos pocos párrafos las palabras de gratitud que debo a todas las personas que, de una u otra forma y a lo largo de los años, han colaborado, bien sea con discusiones técnicas y/o apoyo moral, a la finalización de este trabajo.

En primer lugar me gustaría mostrar mi agradecimiento al Dr. Juan Luis Fernández Martínez, director de la Tesis, por todo el tiempo y esfuerzo dedicados a que esta obra haya salido adelante. Su ayuda y guía en las investigaciones realizadas tienen un valor incalculable. Agradezco también la ayuda recibida por parte de los doctores María Zulima Fernández Muñoz y Luis Mariano Pedruelo González, de cuyos trabajos y resultados esta Tesis es deudora y, en muchos aspectos, continuación natural.

Durante algunos meses de los años 2014 y 2015 realicé dos estancias de investigación en la École des Mines d'Albi-Carmaux (Francia). Es obligado y un placer dejar aquí constancia escrita del amabilísimo trato que recibí por parte de todo el personal de la citada institución, muy especialmente en lo que al Dr. Olivier Fudym respecta, coautor en dos artículos incluidos en esta Tesis y con el que espero seguir colaborando en el futuro. Tampoco puedo dejar de mencionar al Dr. Sylvain Bonvalot, del Laboratoire GET y director del Bureau Gravimétrique International, quien ha proporcionado los datos gravimétricos utilizados en los ejemplos reales tratados en esta Tesis y cuya implicación en la interpretación de los resultados obtenidos no hace más que elevar la calidad final del trabajo.

En la actualidad soy Ayudante en la ETSI en Topografía, Geodesia y Cartografía de la Universidad Politécnica de Madrid. Quiero agradecer las palabras de ánimo que mis compañeros me dedicaban cada vez que me los encontraba por los pasillos, tras el consabido *¿cómo va la Tesis?!* Mención especial en este apartado merece la Dra. Rosa María García Blanco, con quien comparto asignaturas y a la que el calificativo de madre se le queda corto, ya que ha sido para mí mucho más que eso desde mi entrada en la Escuela. Con el Dr. Juan Francisco Prieto Morín he pasado innumerables horas (y las que nos quedan) haciendo cálculos, revisando artículos, hablando de Geodesia, de lo divino y de lo humano, y en él he encontrado apoyo en varios momentos importantes. Quiero también mencionar a mis colegas Sandra Martínez, Teresa Fernández Pareja, Jorge Gaspar, Ramón Alcarria, Alejandra Staller, Antonio Vázquez y Rufino Pérez, ya que, de una forma u otra, todos han aportado su granito de arena.

Y cómo no recordar a los que fueron mis compañeros en el antiguo Instituto de Astronomía y Geodesia, con quienes pasé largos ratos discutiendo y aprendiendo. Por todo ello les doy las gracias a Pablo José González Méndez, a María Charco Romero y a Pedro Galán del Sastre (apoyos fundamentales en un momento delicado), a Alicia Arjona Almodóvar, a Gonzalo Bardenas Manchado y a Beatriz Moreno Monge. Y también a Javier Mejuto González, perseguidor de sueños profesional que un día atracará en el puerto seguro y abrigado que merece.

Contents

Agradecimientos	5
Contents	7
List of Figures	9
Resumen	11
Abstract	13
1 Introduction and state of the art	15
1.1 Introduction	15
1.2 The cost function topography of an inverse problem	17
1.3 The Particle Swarm Optimization (PSO) algorithm	19
1.4 Introduction to gravity inversion in sedimentary basins	22
1.4.1 Basin as a trapezoidal figure	23
1.4.2 Basin as an irregular polygon	24
1.4.3 Basin as a set of simple geometrical bodies	25
1.4.4 Methods based on the direct problem solution	27
1.4.5 Global optimization methods	28
2 Thesis contributions	31
2.1 The effect of noise and Tikhonov regularization in inverse problems	31
2.1.1 The effect of noise in linear inverse problems	32
2.1.2 The effect of Tikhonov regularization in linear inverse problems	34
2.1.3 The effect of noise in nonlinear inverse problems	37
2.1.4 The effect of Tikhonov regularization in nonlinear inverse problems	38
2.2 Gravity inversion in sedimentary basins using PSO	40
2.2.1 Domain discretization	40
2.2.2 Density contrast treatment	40
2.2.3 Search space, constraints, and inversion	41
2.2.4 Inverse problem solution and uncertainty assessment	46
3 Articles published in peer-reviewed journals	49
3.1 From Bayes to Tarantola: New insights to understand uncertainty in inverse problems	51

3.2	The effect of noise and Tikhonov's regularization in inverse problems. Part I: The linear case	67
3.3	The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case	81
3.4	Gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization	93
3.5	3D gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization	109
4	Other publications	133
4.1	Noise, regularization and uncertainty: new insights for linear and nonlinear inverse problems	135
4.2	The effect of the noise and the regularization in inverse problems. Geophysical implications	147
5	Conclusiones y líneas futuras	153
5.1	Conclusiones	153
5.2	Líneas futuras de investigación	154
6	Conclusions and future research	157
6.1	Conclusions	157
6.2	Future research	158
A	Reuse permissions and impact factor reports	161
	Bibliography	187

List of Figures

1.1	Sedimentary basin approximation by a trapezoidal figure	24
1.2	Sedimentary basin as an irregular polygon	24
1.3	Sedimentary basin as a set of rectangles	25
1.4	Sedimentary basin as a set of prisms	26
2.1	Noise and regularization effects in inverse linear problems	36
2.2	Noise and regularization effects in inverse nonlinear problems	39
2.3	Search space limits for a synthetic noise-free 2D model	42
2.4	Noise-free 2D gravity inversion without constraints using PSO	43
2.5	Noise-free 2D gravity inversion with relative constraints using PSO	44
2.6	Search space limits for a synthetic 2D model with white noise	45
2.7	Noisy 2D gravity inversion with relative constraints using PSO and cost function	46
2.8	Convergence and dispersion curves	47
2.9	Noisy 2D inverse problem solution using relative constraints	48

Resumen

En esta Tesis Doctoral se aborda el análisis de incertidumbre de la solución en problemas geofísicos lineales y no lineales en espacios altamente dimensionales. En una primera parte teórica se justifica el análisis de incertidumbre desde un punto de vista determinista mediante técnicas provenientes del álgebra lineal y de la teoría de optimización. En particular, en el caso de los problemas inversos lineales se demuestra que el condicionamiento del sistema lineal está relacionado con la excentricidad de la región de equivalencia. En el caso de los problemas inversos no lineales se justifica por qué las técnicas de linealización sólo proporcionan un análisis de incertidumbre en el entorno de la solución adoptada, pudiendo conducir a soluciones erróneas en el caso de que la información a priori utilizada sea incorrecta.

Se realiza un análisis teórico exhaustivo del efecto del ruido y de la regularización de Tikhonov en la topografía de la función objetivo en los casos de los problemas inversos lineales y no lineales, mostrando analíticamente que en ambos se ve deformada dicha topografía. El ruido la deforma de un modo homogéneo en el caso de los problemas lineales y no homogéneo en los problemas inversos no lineales. En el caso de la regularización de Tikhonov la deformación es anisotrópica, actuando de forma diferente en cada componente del modelo solución.

En segundo lugar, se ha desarrollado un método de inversión en 2D y 3D para la resolución (estimación del mejor modelo y su análisis de incertidumbre) del problema gravimétrico inverso en cuencas sedimentarias, en su formulación no lineal para la determinación de la interfase de separación sedimentos-basamento, basado en el algoritmo Particle Swarm Optimization. Se han obtenido muy buenos resultados con ejemplos sintéticos y reales, tanto en la estimación del modelo de mejor ajuste como en el análisis de su incertidumbre, lo que convierte al algoritmo PSO en una alternativa seria para la resolución de este tipo de problemas.

Palabras clave: Problemas inversos, inversión lineal y no lineal, análisis de incertidumbre, ruido, regularización, Particle Swarm Optimization, problema gravimétrico inverso, cuenca sedimentaria.

Abstract

This PhD Thesis tackles the uncertainty analysis of the solution of linear and nonlinear geophysical inverse problems in high-dimensional spaces. In a first theoretical part, the uncertainty analysis from a deterministic and from the optimization theory point of view is justified. Particularly, it is demonstrated that in linear inverse problems, the linear system conditioning is related to the equivalence region eccentricity. In the case of nonlinear inverse problems, the reason why the linearization techniques only provide a local uncertainty analysis, and can lead to wrong solutions if the a priori information used is not correct, is demonstrated.

An exhaustive theoretical analysis of the effects of noise and Tikhonov regularization in the objective function topography for linear and nonlinear inverse problems is developed, analytically showing that both effects deform the aforementioned topography. Noise deforms it in a homogeneous way in the case of linear inverse problems, while the deformation is non-homogeneous for nonlinear problems. In the case of Tikhonov regularization, the deformation is anisotropic, affecting each component of the solution model in a different way.

Secondly, a method for 2D and 3D nonlinear gravity inversion in sedimentary basins (in the nonlinear form for determination of the interface sediments-basement) has been developed, comprising the best model estimation and its uncertainty analysis and based on the Particle Swarm Optimization algorithm. Very good results have been obtained, both with synthetic and real examples, regarding the best model estimation and its uncertainty analysis, which makes the PSO algorithm a serious alternative for the solution of this kind of problems.

Keywords: Inverse problems, linear and nonlinear inversion, uncertainty analysis, noise, regularization, Particle Swarm Optimization, gravimetric inverse problem, sedimentary basin.

Introduction and state of the art

1.1 Introduction

The concept of ill-posedness was proposed by J. Hadamard in 1902, and involves problems that either have no solution, have an infinity of solutions, or have an unstable (ill-conditioning) solution. This character is translated to the cost function topography as the existence of one or various regions that contain the equivalent models, which are those that predict the observed data at the same level of tolerance and are compatible with the prior information that it is at disposal (Fernández-Martínez et al., 2012a). These equivalent regions, from the geometrical point of view, are valleys in the model space, rectilinear or curvilinear (depending on whether the problem is linear or nonlinear, respectively) and with almost null gradients. The uncertainty analysis of the adopted solution for a problem consists of adequately mapping these valleys, which is always a necessary task as the regularization techniques do not provoke the disappearance of the equivalent model sets, and instead they only stabilize the inversion (Tikhonov and Arsenin, 1977; Fernández-Martínez et al., 2012a).

On the other hand, inverse problems are a special kind of optimization problems since the cost function involves observed (direct or indirect) data, that is always affected by noise. It is well known that, specially for ill-conditioned linear inverse problems, the noise is amplified back to the model parameters through the Moore-Penrose's pseudoinverse operator, generating spurious solutions if regularization techniques are not used. In the case of nonlinear inverse problems, which are very important in Geophysics and industrial applications, the noise deforms in a non-homogeneous way the cost function topography, decreasing the size of good misfit regions and increasing the medium misfit ones. As a result, the global optimum search becomes a harder task, but medium misfit zones are easier to find. This explains the success of global search algorithms in solving problems with few parameters and relatively computational inexpensive direct problems (Fernández-Martínez et al., 2010a; García González, 2011).

The study of the effect of noise in the inverse problem solution has a paramount importance since its influence will determine in a crucial way the quality and uncertainty of the obtained result, which will in turn influence the future decision-making steps in real applications. Traditionally, the uncertainty analysis of inverse problems has been addressed from a probabilistic (Bayesian) point of view, mainly through the use of random search methods, which deal with the uncertainty considering it the result of a stochastic process (Mosegaard and Tarantola, 1995; Gouveia and Scales, 1998). Nevertheless, the application of this technique is not suitable for industrial applications due to the direct's problem high dimensional character and/or the high computational cost, which are common in this kind of environments. In addition, the use of random search methods for the uncertainty analysis is not the best approach as the problem

has a clear algebraic structure. Finally, the linear techniques based on the inverse problem linearization only provide a local uncertainty analysis (around the adopted solution) when they are applied to nonlinear inverse problems. Therefore, the uncertainty analysis of nonlinear inverse problems in high dimensional spaces and with computationally highly expensive direct problems is still an open field in scientific research, although significant steps have been taken using a deterministic analysis of the cost function topography (Fernández-Martínez et al., 2012a) and through dimension reduction techniques (Fernández Muñiz, 2012) and global search methods (Fernández-Martínez et al., 2011).

The two main objectives of this PhD Thesis are:

1. To perform an exhaustive analysis of the effect of observational noise and zero-order Tikhonov regularization in the cost function topography for linear and nonlinear inverse problems. Expressions for the quantification of the noise and regularization effects in the cost function topography will be developed for linear and nonlinear problems. This analysis, although theoretical at this stage, is valuable in order to get important information for the future design of inversion methods, and might impact how the uncertainty analysis is performed, since noise in data is always present and good prior models are not always available.
2. To develop a method for nonlinear gravity inversion in sedimentary basins (for 2D and 3D environments), comprising the best model estimation and its uncertainty analysis, based on the Particle Swarm Optimization (PSO) algorithm. The gravity inverse problem is an essential and widely used tool for the Earth crust study at regional and local scales, as well as in mineral exploration. As a potential-field based problem, the gravity inverse method has an inherent non-uniqueness, so regularization techniques are commonly used when local methods are used for its solution. In this work, a global search approach (PSO) is tested, obtaining good results in the best model estimation, as well as concerning its uncertainty assessment.

This PhD Thesis is presented in the form of Thesis as Compendium of Publications (*Tesis como compendio de publicaciones*), so the most important part of this report are the chapters dedicated to the publications themselves, although a general introduction is also necessary. Chapter 1 contains a general introduction and the state of the art of the main points discussed in this PhD Thesis. Section 1.2 presents a short introduction to the cost function topography of the inverse problems. Section 1.3 gives a general background of the Particle Swarm Optimization (PSO) algorithm, as it was the method used in the gravity inversion problem solution. In section 1.4 a general introduction about gravity inversion in sedimentary basins is given. Several classical methods are enumerated and their advantages and disadvantages are discussed. Chapter 2 presents a summary of the contributions of this PhD Thesis. Section 2.1 contains a general introduction to the analysis of noise and Tikhonov regularization effects in linear and nonlinear inverse problems, whose detailed development is presented in articles 1, 2 and 3, and in publications 1 and 2. Finally, in section 2.2 the gravity inversion method developed in this PhD Thesis is presented in a general form, being articles 4 and 5 the publications where the new technique is presented in detail. Chapter 3 contains the full text of articles 1, 2, 3, 4, and 5, while chapter 4 presents publications 1 and 2 (two congress contributions). Finally, chapter 6 presents the conclusions and future research lines, and appendix A contains other documentation necessary in order to fulfill the Universidad de Oviedo regulations for the form of Thesis as Compendium of Publications.

1.2 The cost function topography of an inverse problem

Most inverse problems in applied sciences can be written in discrete form as

$$\mathbf{F}(\mathbf{m}) \simeq \mathbf{d}^{obs}, \quad (1.1)$$

where $\mathbf{d}^{obs} \in \mathbb{R}^s$ is the vector of observed data, always, whatever the problem, affected by noise; $\mathbf{m} = (m_1, m_2, \dots, m_n)^T \in M \subset \mathbb{R}^n$ is the set of discrete parameters to be found, that belongs to a set of admissible models M defined in terms of some prior knowledge; and $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^s$ represents the forward problem (linear or nonlinear) between the model and data spaces. The symbol \simeq in equation (1.1) means that this relation might not be exact, so the referred problem, that consists of finding the model or models \mathbf{m} , whose predictions $\mathbf{F}(\mathbf{m})$ accurately match the observed data \mathbf{d}^{obs} , might not have a solution. This fact is related to the ill-posedness character of inverse problems, that is, either the inverse problem does not admit solution, or the solution exists but is not unique, or, finally, the solution exists and is unique, but unstable, i. e., the solution does not depend continuously on the observed data. This incompatibility is also caused by the fact that the operator \mathbf{F} is only an approximation of the reality (modelling errors), although in this PhD Thesis the effect of the noise in data and that of the regularization will be the only ones taken into account.

The cost function is usually defined (and in this PhD Thesis it will be used in that way) as the squared Euclidean norm of the residual vector in the form

$$c(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2, \quad (1.2)$$

although other norms can be used. The limit of the region of equivalence (the region where all contained models fit the observations below a tolerance level of value tol) is defined, using the Euclidean norm, as

$$\|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 = tol^2. \quad (1.3)$$

[Fernández-Martínez et al. \(2012a\)](#) analyzed the topography of the cost function in linear and nonlinear inverse problems in order to describe its general characteristics.

For linear inverse problems, [Fernández-Martínez et al. \(2012a\)](#) show that equation (1.3) corresponds to a hyper-quadric with kernel matrix $\mathbf{F}^T \mathbf{F}$. This hyper-quadric can be a hyper-ellipsoid or a elliptic hyper-cylinder, depending on the singular values of \mathbf{F} : it will be a hyper-ellipsoid if all singular values are positive or an elliptic hyper-cylinder if any of them are zero. In the case of purely overdetermined (or even-determined) linear problems with $s \geq n$ and $\text{rank}(\mathbf{F}) = n$, the hyper-quadric is a hyper-ellipsoid centered at the least squares solution of the system, where its axes are oriented following the orthogonal vectors contained in the \mathbf{V} base defined by the singular value decomposition of \mathbf{F} , i. e., $\mathbf{F} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. The lengths of the axes are defined as $1/\lambda_k$, where λ_k are the singular values, which implies that the model parameters in the \mathbf{V} base are more uncertain in the directions of smaller singular values. In this case, it can be concluded that the region containing the equivalent models is a bounded and simply connected set.

In the case of underdetermined linear systems, where the number of equations is less than the number of unknowns and all the equations are linearly independent, the hyper-quadric of equivalence is an elliptic hyper-cylinder with infinitely long axes, corresponding to the null singular values, in the directions spanning the kernel of the forward operator \mathbf{F} . In this case, the least squares solution is the minimum norm solution. The region containing the equivalent models is an unbounded and simply connected set.

For rank-deficient overdetermined linear systems the rank of \mathbf{F} is reduced to the number of independent equations, so it is lower than the number of parameters. The region of linear equivalence becomes unbounded in the directions of the \mathbf{V} base that spans the null space of \mathbf{F} . The hyper-quadric degenerates in this case to an elliptic hyper-cylinder (or to a set of planes in the limiting case). Rank-deficient underdetermined linear systems are a subclass of underdetermined problems. The least squares solution in this case coincides with the minimum norm solution of the ensemble of least squares solutions. As in the previous case, the region of equivalence is an elliptic hyper-cylinder.

In conclusion (Fernández-Martínez et al., 2012a), the region of equivalence for linear problems is a simply connected set and no other isolated island or local minimums exist in the misfit function landscape, as it is apparent from the topological features of the hyper-ellipsoids and elliptical hyper-cylinders. This region is bounded only if the inverse problem is purely overdetermined or even-determined. In the other cases, the shape of the cost function corresponds to a flat elongated valley in the directions of the null space of the forward operator \mathbf{F} . An important point is the fact that the problem is linear implies that the valley is straight; since the hyper-quadric is the exact region of equivalence, the valley direction can not change in the model space. In any case, practical linear inverse problems in Science and Engineering are ill-conditioned, due to the small singular values of the forward operator \mathbf{F} , causing instabilities in the inversion, and amplifying the noise in data. Tikhonov regularization (Tikhonov and Arsenin, 1977) is the most commonly used method of regularization for ill-posed problems, stabilizing the solution by adding an extra term to the cost functional presented in equation (1.2), becoming (for the zero-order Tikhonov regularization)

$$c^r(\mathbf{m}) = \|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2, \quad (1.4)$$

where ε^2 is the so-called damping parameter, generally close to zero, and \mathbf{m}^{ref} is a reference model. The main effect of the Tikhonov regularization is that the hyper-quadric becomes a hyper-ellipsoid whose longer axes are $1/\varepsilon$ in the directions of the \mathbf{V} vectors spanning the null space of \mathbf{F} , limiting then the length of the misfit valley, causing the region of equivalence to become bounded.

For nonlinear inverse problems, the same kind of reasoning that was applied for linear inverse problems to characterize local regions of equivalence around a certain model \mathbf{m}_0 belonging to the nonlinear region of equivalence can be used, although due to the nonlinearity of \mathbf{F} the concept of null space of \mathbf{F} can not be strictly used, since it always implies linearity. Nevertheless, to a first-order approximation the analysis holds for the null space of the Jacobian of \mathbf{F} around a model \mathbf{m}_0 , $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}$. Although higher order terms are involved in the topography of the cost function in nonlinear problems, the first-order analysis illuminates the main similarities and differences between the regions of equivalence in linear and nonlinear inverse problems.

As for the linear problems, the equivalence region for the local environment (the tangent plane) of \mathbf{m}_0 is a hyper-ellipsoid or an elliptical hyper-cylinder, depending on the rank of $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}$. But this analysis depends on the used model \mathbf{m}_0 , which makes the Jacobian of \mathbf{F} to change continuously with the point \mathbf{m}_0 . This makes the local hyper-quadric change its orientation for each \mathbf{m}_* , depicting a meandering valley or the so-called croissant-shape, depending on the complexity of the inverse problem. Also, the model space of nonlinear inverse problems might contain different unconnected islands of low misfit. Finally, Fernández-Martínez et al. (2012a) concluded that uncertainty in nonlinear inverse problems need not to be treated as random processes, since equivalence regions have a natural algebraic interpretation.

1.3 The Particle Swarm Optimization (PSO) algorithm

Particle Swarm Optimization (PSO) has been the algorithm used in the second part of this PhD Thesis for the solution (best model selection and its uncertainty assessment) of the proposed application inverse problem (see sections 1.4 and 2.2, and articles 4 and 5). Therefore, a short and general description of the method is given here. PSO (Kennedy and Eberhart, 1995) is a global optimization algorithm inspired by the behavior of bird flocks and fish schools searching for food, where a swarm of particles (models) explores the space of possible solutions in order to optimize a given cost function related to an inverse problem, which in this work is the data prediction error expression (1.2) particularized for the inverse gravity problem. In this case a full family of Particle swarm optimizers is used, formed by the so called GPSO, CC-PSO, CP-PSO, RR-PSO and PP-PSO members. The cloud versions of these algorithms (Fernández-Martínez et al., 2010a) are used, that is, no parameter tuning of the inertia, local and global accelerations is needed, since each of the particles in the swarm have their own PSO parameters, that are automatically chosen from a set of PSO parameters that are located in the neighborhood of the corresponding second order stability regions. These mathematical results make PSO to be a very unique algorithm, different from other heuristic approaches.

PSO applied to optimization problems is very simple: individuals, or particles, are represented by vectors whose length is the number of degrees of freedom of the optimization problem. To start, a population of particles is initialized with random positions \mathbf{x}^0 and velocities \mathbf{v}^0 , and the same objective function (expression (1.2), for example) is used to compute the fitness (or prediction error) of each particle in the swarm. As time advances, the position and velocity of each particle is updated as a function of its fitness and the corresponding values of its neighbors. At time-step $k + 1$, the PSO algorithm updates the individuals positions \mathbf{x}^{k+1} and velocities \mathbf{v}^{k+1} as

$$\begin{cases} \mathbf{v}_i^{k+1} = \omega \mathbf{v}_i^k + \phi_1 (\mathbf{g}^k - \mathbf{x}_i^k) + \phi_2 (\mathbf{l}_i^k - \mathbf{x}_i^k), \\ \mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1}, \end{cases} \quad (1.5)$$

where \mathbf{g}^k is the global best position in the whole swarm, \mathbf{l}_i^k is the i -th particle's best position, $\phi_1 = r_1 a_g$ and $\phi_2 = r_2 a_l$ are the random global and local accelerations, and ω is a real constant called inertia weight. Finally, r_1 and r_2 are random variables uniformly distributed in $(0, 1)$ to weight the global and local acceleration constants a_g and a_l . Tuning of the PSO parameters implies providing suitable values for the inertia constant ω and for the local and global accelerations a_g and a_l , in order to achieve exploration and also convergence towards the low misfits regions. As it was pointed before, in this PhD Thesis the cloud versions of the algorithm (Fernández-Martínez et al., 2010a) are used, therefore no selection of the ω , a_g and a_l PSO parameters is needed.

The PSO algorithm, as it was originally proposed in equation (1.5), is not dimensionally correct, since velocities and accelerations, and positions and velocities are directly added. The solution to this simple enigma comes from physical analysis. PSO can be interpreted as a double stochastic gradient algorithm in the model space, and is the particular case of the generalized PSO (GPSO) algorithm (Fernández-Martínez and García-Gonzalo, 2008) for $k = t$ and $\Delta t = 1$, which has in general a fast convergence behavior (García González, 2011):

$$\begin{cases} \mathbf{v}(t + \Delta t) = \mathbf{v}(t) [1 - (1 - \omega)\Delta t] + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{l}(t) - \mathbf{x}(t)], \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t)\Delta t. \end{cases} \quad (1.6)$$

The GPSO algorithm was derived using a mechanical analogy: a damped mass-spring system with unit mass, damping factor $1 - \omega$, and total stiffness constant $\phi = \phi_1 + \phi_2$, i. e., the so-called PSO continuous model

$$\begin{cases} \mathbf{x}''(t) + (1 - \omega)\mathbf{x}'(t) + \phi\mathbf{x}(t) = \phi_1\mathbf{g}(t) + \phi_2\mathbf{l}(t), \\ \mathbf{x}(0) = \mathbf{x}_0, \\ \mathbf{x}'(0) = \mathbf{v}_0, \end{cases} \quad (1.7)$$

where $\mathbf{x}(t), t \in \mathbb{R}$ is the position of any particle of the swarm in continuous time. In this model, particles interact through the local and global attractors $\mathbf{l}(t)$ and $\mathbf{g}(t)$, and mean particle trajectories oscillate around the particle position (Fernández-Martínez and García-Gonzalo, 2008)

$$\mathbf{o}_i(t) = \frac{\phi_1\mathbf{g}(t) + \phi_2\mathbf{l}_i(t)}{\phi_1 + \phi_2}. \quad (1.8)$$

The full stochastic analysis of the PSO continuous and discrete models (GPSO) was performed in a collection of 3 papers by Fernández-Martínez and García-Gonzalo (2008, 2009, 2011b). This analysis shed light on the relationship between PSO convergence and the PSO parameters tuning: good PSO parameters (ω, a_g and a_l) are located in the neighborhood of the upper border of the second order stability region. Therefore, no magical tuning points with five decimal figures exist: stochastic stability and convergence are intimately related in the PSO case. Using the above mentioned mechanical analogy a full family of Particle Swarm optimizers was also derived (Fernández-Martínez and García-Gonzalo, 2009, 2012; García-Gonzalo et al., 2014). Numerical analysis using different analytical benchmark functions have shown that the most performing algorithm of the family in terms of the balance of its exploration/exploitation capabilities is RR-PSO, which is the PSO family member obtained by adopting both regressive finite difference schemes of the PSO continuous model (equation (1.7)) for acceleration and velocity.

The RR-PSO algorithm can be written as

$$\begin{cases} \mathbf{v}(t + \Delta t) = \frac{\mathbf{v}(t) + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{l}(t) - \mathbf{x}(t)]}{1 + (1 - \omega)\Delta t + \phi \Delta t^2}, \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t)\Delta t, \end{cases} \quad (1.9)$$

where $(t, \Delta t) \in \mathbb{R}$. This algorithm has unbounded regions of first and second order stochastic stability. It has been numerically shown (Fernández-Martínez and García-Gonzalo, 2012) that good parameters sets in the RR-PSO case are concentrated around the line $\bar{\phi} = 3(\omega - 3/2)$, mainly for inertia values $\omega > 2$. This straight line is independent of the cost function that is optimized, and remains invariant when the number of optimization parameters increases. Furthermore, this line is located in a region of medium attenuation and very high frequency for the swarm particle trajectories. This last property confers to RR-PSO a good balance between exploration and exploitation, allowing for a very efficient and explorative search around the oscillation center of each particle in the swarm.

Two other explorative members of the PSO family are also used in this PhD Thesis (see Fernández-Martínez and García-Gonzalo (2009, 2012) for details):

- The CP-PSO:

$$\begin{cases} \mathbf{v}(t + \Delta t) = \frac{(1 + \phi \Delta t^2)\mathbf{v}(t) + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{l}(t) - \mathbf{x}(t)]}{1 + (1 - \omega)\Delta t}, \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t. \end{cases} \quad (1.10)$$

- The PP-PSO:

$$\begin{cases} \mathbf{v}(t + \Delta t) = [1 - (1 - \omega)\Delta t] \mathbf{v}(t) + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{l}(t) - \mathbf{x}(t)], \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t. \end{cases} \quad (1.11)$$

The CP-PSO is the most explorative member of the family (García González, 2011), which is an interesting option for performing nonlinear uncertainty analysis and exploring the cost function landscape. PP-PSO has the same velocity rule than GPSO, but the positions of the particles are written in time t , instead of $t + 1$. PP-PSO is more explorative than GPSO, having a lower convergence rate.

Finally, the CC-PSO algorithm was also used, which, as the GPSO, has a fast convergence rate (García González, 2011) and can be written as (Fernández-Martínez et al., 2010a)

$$\begin{cases} \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \left\{ \frac{[2 + (\omega - 1)\Delta t]\mathbf{v}(t) + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{l}(t) - \mathbf{x}(t)]}{2} \right\} \Delta t, \\ \mathbf{v}(t + \Delta t) = \frac{2 + (\omega - 1)\Delta t}{2 + (1 - \omega)\Delta t} \mathbf{v}(t) + \sum_{k=0}^1 \left\{ \frac{\phi_1 \Delta t [\mathbf{g}(t + k\Delta t) - \mathbf{x}(t + k\Delta t)] + \phi_2 \Delta t [\mathbf{l}(t + k\Delta t) - \mathbf{x}(t + k\Delta t)]}{2 + (1 - \omega)\Delta t} \right\}, \end{cases} \quad (1.12)$$

where $(t, \Delta t) \in \mathbb{R}$.

The use of PSO in applied geophysics has been introduced in Fernández-Álvarez et al. (2006), Shaw and Srivastava (2007) and Fernández-Martínez et al. (2010a,b). Particularly, Fernández-Martínez et al. (2010a) presented the application of Generalized PSO (GPSO), CC-PSO and CP-PSO to the solution and appraisal of a 1D-DC resistivity inverse problem, justifying in each case how to perform the parameter tuning. Nevertheless, in gravity inversion (the applied problem of this PhD Thesis) the PSO algorithm has barely been used. Sanyi et al. (2009) presented some simple synthetic examples using buried cylinders (with depths and density contrast as unknowns), and compared the results with the ones obtained via Levenberg-Marquardt, genetic algorithm (GA), simulated annealing (SA), and Ant Colony Optimization (ACO) approaches. The conclusion was that PSO outperforms the rest of the algorithms, showing higher convergence rates for PSO and ACO with respect to GA and SA. Toushmalani (2013b,a) used PSO to invert the parameters of a fault (thickness of the sheet, left and right distances to the middle of the sheet, and the fault's angle) using gravity anomalies, comparing the results with the Levenberg-Marquardt method. His general conclusion was that PSO provides better agreement between the predicted model anomaly and the observations than using the Levenberg-Marquardt technique. Sweilam et al. (2008) also compared the PSO algorithm with function stretching (SPSO) with genetic algorithms for synthetic and real (a salt dome) inverse gravimetric problems, finding similar solutions for both algorithms, but with the advantage for SPSO of having less parameters to tune than the GA method. These papers that are devoted to the use of PSO in gravity inversion as a global optimizer, comparing its performance with classical inversion techniques, miss the most important feature of global optimization algorithms, which is their capability to perform an approximate nonlinear uncertainty analysis (sampling of the nonlinear uncertainty region(s) of the solution that has been adopted. Also, a clear description about the importance of a correct tuning of the PSO parameters Fernández-Martínez et al. (2010a,b, 2012b) is usually absent.

1.4 Introduction to gravity inversion in sedimentary basins¹

Estimation, using surface gravity measurements, of the horizontal separation interface between two media which have different densities (where the upper one, embedded in the lower one, has a density $\rho_u < \rho_l$, as it is common in sedimentary basins) is a nonlinear inverse problem. Gravity inversion in this kind of environments is a classical tool, widely used in Geophysics, in oil and gas prospecting, hydrogeology, glaciology, etc. (Dobrin, 1960; Nettleton, 1976; Telford et al., 1976; Parker, 1994; Blakely, 1995). The gravity inverse problem, as a potential field based method, has an inherent non-uniqueness in their solution, which leads to an infinity number of mass distributions generating the same gravitational signal (Skeels, 1947; Al-Chalabi, 1971; Hinze et al., 2013; Zhdanov, 2015). It is therefore mandatory to introduce some kind of constraint(s) and to take into account external information (as borehole data, seismic profile information, contrasted prior models, etc.) in order to enclose the final solution.

Once the gravity is observed (in this PhD Thesis only its norm is considered, regardless of its vector character) and corrected (time-dependent signals and atmospheric effects), measurements are transformed into anomalies Δg through the classic and well-known free-air, Bouguer plate and terrain reductions, plus the subtraction of the normal gravity field signal (Telford et al., 1976; Blakely, 1995). These values, called *complete Bouguer anomalies*, contain information about the crust-mantle interface and the mass distribution in the wide range of depths for the Earth crust, so their influence is considered as regional. Although some sedimentary basins whose area covers thousands of square kilometers exist, their depths are commonly small compared to the Earth crust thickness, so for a correct problem interpretation it is necessary to isolate the part of Δg generated only by the sediments. This part is commonly called *local* or *residual anomaly*, while the part of the signal due to the regional and deep mass distributions is known as *regional trend*. These components can be isolated as a previous step or during the inversion process itself, and a variety of methods exist in order to perform this task, as, for example, the polynomial fitting of curves or surfaces (Beltrão et al., 1991), the use of minimum curvature surfaces (Mickus et al., 1991) or through spectral techniques (Xu et al., 2009), among others. In this PhD Thesis, only the polynomial method has been used.

Residual anomalies Δg on sedimentary basins have in the most common cases negative values, as the anomalous mass derived from a contrast density sediments-basement $\Delta\rho = \rho_s - \rho_b < 0$ is also negative. These anomalies are directly related to the basin depth (sediments layer thickness), and they constitute the fundamental observable for the inversion task. They can be also classified (assuming that the time-dependent signals and atmospheric effects were correctly removed, introducing only negligible level of noise) in various components:

1. The anomaly due to the sediments itself.
2. The observational noise, always present.
3. The noise introduced in gravity reduction for complete Bouguer anomaly computation.
4. The noise introduced in the regional trend suppression.

Noise introduced in parts 2, 3 and 4 contributes to the previously highlighted inverse problem non-uniqueness, so it is must always be taken into account in the results interpretation step.

¹Gravity inversion in sedimentary basins is the main application treated in this PhD Thesis, as a continuation of the author's Master Degree. In the future, other geophysical inverse problems will follow.

The most widely used techniques for the solution of this inverse problem are based on local optimization methods, either by the linearization of the problem and the use of linear algebra techniques, or by the sequential application of the direct formulation. Global search methods have also been proposed. For the basin modeling several alternatives have been proposed, as its approximation as regular geometric bodies, irregular polygons or the discretization in rectangles or prisms. Regarding the density contrast $\Delta\rho$ between the sediments and the basement, the nonlinear approach for the problem, where the interface depths are the unknowns, imply that $\Delta\rho$ must be postulated, either as a constant value or being variable with depth.

Finally, with regards to the dimensions of the space the problem can be proposed as 2D, 2.5D or 3D. When a dimension of an anomalous body is much larger than the other two (Nettleton (1976) suggests that a $2\times$ or $3\times$ factor is enough), its gravity anomaly computation can be solved using a 2D approximation (Pick et al., 1973; Nettleton, 1976; Telford et al., 1976). This situation is common in sedimentary basins, where their horizontal extensions are generally much larger than their depth, so profiles perpendicular to the largest dimension can be used in order to use a 2D formulation. The 2.5D approach is used when the strike length is not large enough to be considered infinite and using the 2D approximation is not valid (Chakravarthi and Sundararajan, 2006). Although in this latter case a 3D approach could be used, this implies an increase in the parameters number.

In the next sections, some parametrization examples will be presented. The density contrast approximations and solution methods will be explained in order to get general conclusions for comparison purposes with the methodology that has been proposed in this PhD Thesis.

1.4.1 Basin as a trapezoidal figure

Rao (1990) presents a bidimensional inversion method based on the basin approximation as a trapezoidal figure, as it can be seen in figure 1.1. The polygon has the points A , B , C and D as vertices, and the origin O is situated at the terrain surface, above the vertical of the bottom side CD middle point (distance which is stated as $2T$). Z_1 and Z_2 are the upper and lower side depths (Z axis is considered positive downwards). As in an asymmetrical trapezoid the minimum (or maximum if the density contrast is positive) of the generated gravity anomaly is not located in the vertical of its geometrical center, a new unknown² D is added, which expresses the distance from a reference point D to the origin O . The model is expressed by the parameters $\mathbf{m} = [Z_1, Z_2, T, \theta_1, \theta_2, D]$, which are the unknowns to estimate in the inversion process. r_1 , r_2 , r_3 , r_4 , ϕ_1 , ϕ_2 , ϕ_3 and ϕ_4 are auxiliary variables needed in the attraction computation of the basin over a point $P(x, z)$.

In this case, the sediments-basement density contrast variation, $\Delta\rho = \rho_s - \rho_b$, is considered variable following a quadratic function with depth as

$$\Delta\rho(Z) = a_0 + a_1Z + a_2Z^2, \quad (1.13)$$

where a_0 is the density contrast at the terrain surface, and a_1 and a_2 are variation parameters, which can be determined (adjusted via least squares or other technique) based on data measured at boreholes, for example. The formulation for the attraction of the basin model over an exterior point at height $Z = 0$, and whose density contrast follows equation (1.13), can be expressed as

$$\Delta g_P = F(x_p, \Delta\rho(Z), Z_1, Z_2, T, \theta_1, \theta_2, D). \quad (1.14)$$

²Although letter D was also used for a vertex, it is the one used by Rao (1990).

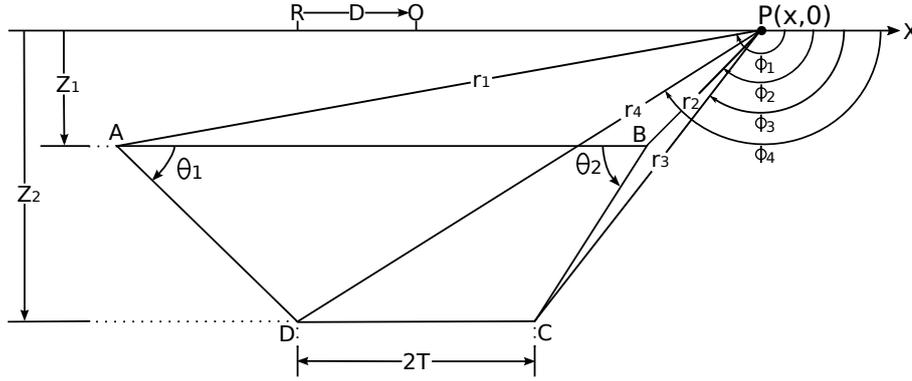


Figure 1.1: Rao (1990) model of sedimentary basin approximation by an asymmetrical trapezoidal figure.

Given a set of N observations at the terrain surface, considered it at height $Z = 0$, the minimization of

$$f(\mathbf{m}) = \sum_{i=1}^N [\Delta g_{P_i}^{obs} - \Delta g_{P_i}^{com}]^2 = \|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2, \quad (1.15)$$

is considered, where $\Delta g_{P_i}^{obs}$ are the observed anomalies (previously refined of regional trend) and $\Delta g_{P_i}^{com}$ are the computed anomalies using equation (1.14) at each observed point. Starting with initial values for Z_1 , Z_2 , T , θ_1 , θ_2 and D , the Marquardt algorithm (Marquardt, 1963) is employed for the problem solution. Regarding the uncertainty assessment, Rao (1990) does not cover this topic explicitly, except for the solution of the problem using different initial values for the parameters and their qualitative evaluation. It should be remarked that this is a very low-dimensional problem.

1.4.2 Basin as an irregular polygon

A sedimentary basin parametrization via an irregular polygon composed by an arbitrary number of sides is a commonly used technique (see for example Al-Chalabi (1972); Chakravarthi et al. (2001)). Figure 1.2 shows an approximation of the model, where a polygon composed by N vertices, each of them along the vertical of each observed point (although this is not mandatory), depicts the sediments-basement interface. The upper side of the polygon is the terrain surface, generally considered planar and at height $Z = 0$.

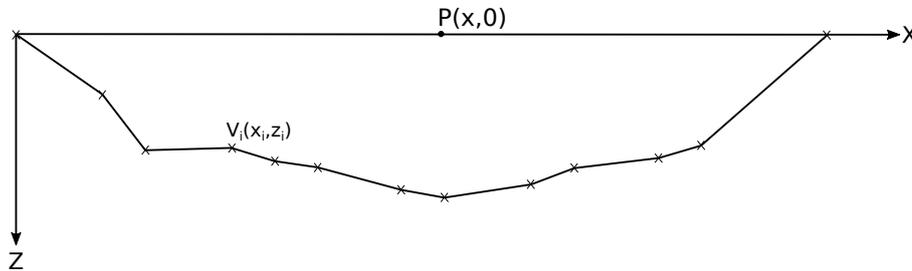


Figure 1.2: Sedimentary basin as an irregular polygon composed by an arbitrary number of sides, as presented in Chakravarthi et al. (2001).

Starting with an initial model, the inversion consists in the optimum position determination along the vertical of each N vertices V_i , in a way that the functional (1.15) be minimized (using

the corresponding expression for Δg^{com}). Al-Chalabi (1972) uses a constant density contrast $\Delta\rho$, proposes to include as unknowns the coefficients of a polynomial in order to fit the regional trend, and also outlines the possibility of fixing the position of one or more polygon vertices if their depths are known a priori. On the other hand, in Chakravarthi et al. (2001) a model for the density contrast is used, which follows the formulation of Rao et al. (1993, 1994):

$$\Delta\rho(z) = \frac{\Delta\rho_0^3}{(\Delta\rho_0 - \alpha z)^2}, \quad (1.16)$$

where $\Delta\rho_0$ is the density contrast at the terrain surface, α a density variation parameter, and z the depth. As in equation (1.13), this variation parameter can be experimentally adjusted when density variation data with depth is at disposal.

Al-Chalabi (1972) proposes the problem solution through gradient based local optimization methods or using «local search» techniques, pointing that this last methods, although slower than the local optimization ones, are more robust when the initial model is not good and is located far from the optimum. Chakravarthi et al. (2001) use the Marquardt algorithm (Marquardt, 1963) to perform the inversion. In neither case uncertainty assessment is performed. Chakravarthi et al. (2001) only compare the obtained depth in one of the proposed examples with a value from a borehole, while Al-Chalabi (1972) only emphasizes the misfit between the model residuals and the noise attributed to the observations, although he indicates that this fact is not necessarily indicative of a geologically plausible solution, pointing the importance of using all the external information at disposal in order to constraint the problem.

1.4.3 Basin as a set of simple geometrical bodies

Another commonly used method for the basin parametrization in 2D uses juxtaposition of rectangles (see for example Silva et al. (2006, 2009)). As it can be seen in figure 1.3, the rectangles are situated in such a way that their upper sides are at the terrain surface level, considered plane and at height $Z = 0$, although that is not necessarily mandatory. For the sake of simplicity, it is common to set the horizontal dimensions to be equal for all rectangles. Then, the bottom sides depict the sediments-basement interface.

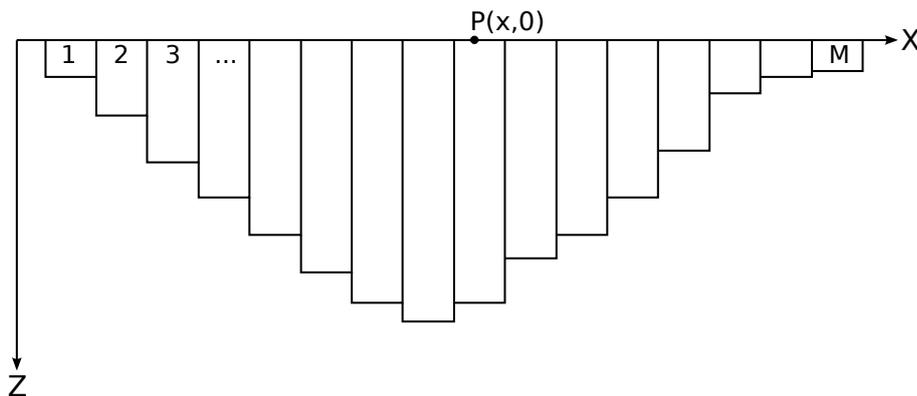


Figure 1.3: Sedimentary basin model as an accretion of rectangles.

This method can also be applied to 3D environments if prisms are used instead of rectangles (see for example Barbosa et al. (1997, 1999); Chakravarthi and Sundararajan (2007)). Figure 1.4 shows schematically the 3D model where, as in the 2D case, the bottom sides of the M prisms

depict the sediments-basement interface. In both cases, 2D and 3D, it is not necessary that the observed points be gridded, although [Barbosa et al. \(1999\)](#) and [Chakravarthi and Sundararajan \(2007\)](#) have used this regular distribution.

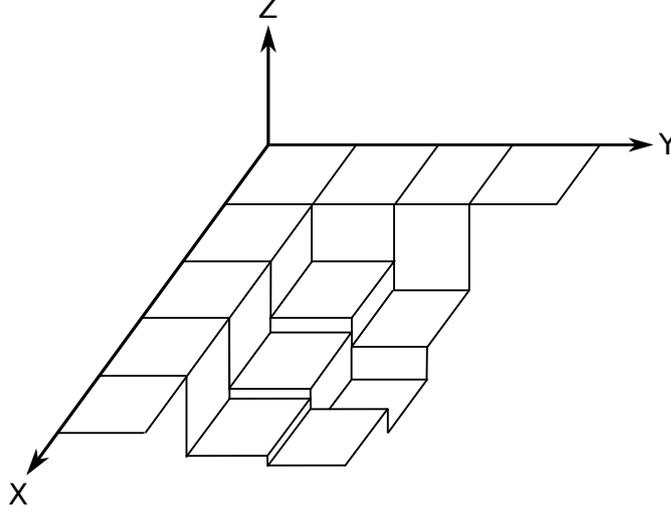


Figure 1.4: Sedimentary basin model as an accretion of right rectangular prisms.

Let N the number of observed gravity points, $P_i(x_i, 0)$ for the 2D case and $P_i(x_i, y_i, 0)$ for the 3D one, with $i = 1 \dots N$, which are situated at the rectangles or prisms upper sides. The gravity anomaly generated by the model over each point follows the equation

$$\Delta g_{P_i} = \sum_{j=1}^M F(\Delta\rho(z), z_j, r_{ij}), \quad (1.17)$$

where $\Delta\rho(z)$ is the density contrast, constant or variable with depth, z_j is the depth of each prism (the problem unknowns), and r_{ij} is the horizontal vector determining the position of each rectangle or prism related to each observation point. [Chakravarthi and Sundararajan \(2007\)](#) include a second degree polynomial surface as regional trend, so equation (1.17) is then

$$\Delta g_{P_i} = \sum_{j=1}^M F(\Delta\rho(z), z_j, r_{ij}) + Ax_i^2 + By_i^2 + Cx_i y_i + Dx_i + Ey_i + F, \quad (1.18)$$

where A, B, C, D, E and F are the polynomial parameters to be fitted.

Regarding the density contrast, [Barbosa et al. \(1997, 1999\)](#) used a constant value, [Chakravarthi and Sundararajan \(2007\)](#) used the law referred in equation (1.16), and [Silva et al. \(2006, 2009\)](#) employed the hyperbolic law of [Litinsky \(1989\)](#):

$$\Delta\rho(z) = \frac{\Delta\rho_0 \beta^2}{(\beta + z)^2}, \quad (1.19)$$

where $\Delta\rho_0$ is the density contrast at the surface and β a variation parameter. In any case, $\Delta\rho_0$, α and β can be also considered as additional model parameters at the inversion.

[Chakravarthi and Sundararajan \(2007\)](#) propose the minimization of

$$f(\mathbf{z}) = \sum_{i=1}^N [\Delta g_{P_i}^{obs} - \Delta g_{P_i}^{com}]^2 = \|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2, \quad (1.20)$$

where $\Delta g_{P_i}^{com}$ is computed using equation (1.18). The problem is solved using the Marquardt algorithm (Marquardt, 1963), and there is not imposed any external constraint, except the fact that all prisms depth defining the basin perimeter are set to 0, which guarantees in all cases more equations than unknowns. The authors compared the inversion results with known depths from boreholes, but there is no uncertainty assessment.

Silva et al. (2006, 2009) used constraints in order to impose smooth conditions to the sediments-basement interface for stabilizing the inversion process. The objective function was

$$f^r(\mathbf{z}) = \frac{1}{M-1} \sum_{j=1}^{M-1} (z_j - z_{j+1}) = \frac{1}{M-1} \|\mathbf{Rz}\|_2^2, \quad (1.21)$$

where the superscript r means *relative constraints*, as they are called in Barbosa et al. (1997), and \mathbf{R} is a regularization matrix designed to look for smooth solutions. The objective function is in this case

$$f(\mathbf{z}) = \frac{1}{N} \|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2 + \mu^r \frac{1}{M-1} \|\mathbf{Rz}\|_2^2, \quad (1.22)$$

where μ^r is a balance factor between the two parts in the equation. Silva et al. (2006, 2009) solve the problem via the Marquardt algorithm and do not perform any uncertainty assessment for the adopted solution.

Barbosa et al. (1997, 1999) add to equation (1.22) a new kind of constraint called *absolute constraints* in order to take into account known depths, if they exist, at discrete points. The objective function is then

$$f(\mathbf{z}) = \frac{1}{N} \|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2 + \mu^r \frac{1}{M-1} \|\mathbf{Rz}\|_2^2 + \mu^a \frac{1}{L} \|\mathbf{z}^{obs} - \mathbf{Az}\|_2^2, \quad (1.23)$$

where the vector \mathbf{z}^{obs} stores nonzero values in the positions corresponding to prisms with known depth and the matrix \mathbf{A} (with dimensions $L \times M$) has as many rows as known depths, and each of them stores the value 1 in the corresponding column. The values μ^a and μ^r are regularization parameters. Barbosa et al. (1999) introduced in the part relative to the relative constraints a weight matrix \mathbf{W} in the form

$$f^r(\mathbf{z}) = \frac{1}{M-1} \|\mathbf{WRz}\|_2^2, \quad (1.24)$$

so that the influence of each individual constraint can be managed. All these approaches were solved via the Marquardt method and, as in previous cases, no uncertainty assessment was performed.

1.4.4 Methods based on the direct problem solution

The main drawbacks of local optimization techniques are their dependency on the initial model that has been adopted, besides the numerical problems that could appear in the systems of equations solution. Inversion methods based on the sole repetition of the direct problem, without the need of any linear (or linearized) equations system solution, have also been proposed. Although they are based on the direct problem solution, they can not be classified as global optimization methods because they do not explore the model space in a global way. In general, although the implementation and computation scheme are simpler than for the local optimization methods, the dependency of an initial model is still present, so the obtained solution can be a local minima if the initial model is far from the true one.

[Bott \(1960\)](#) presented a bidimensional inversion method where the basin is modeled as a set of rectangles, as it was explained in section 1.4.3, and uses a constant density contrast $\Delta\rho$. Given a set of N gravity points (with regional trend removed) and dividing the basin into M rectangles, the inversion procedure is composed by the following steps:

1. A first approximation \mathbf{H}_1 to the rectangles' thickness is computed using the Bouguer plate formula $\Delta g_i = 2\pi G \Delta\rho H_i$ for $i = 1, \dots, M$, where G is the gravitational constant. If the observed points do not coincide with the prisms centers (in the horizontal plane), the anomalies must be interpolated.
2. Using the estimated initial model, the anomalies at each observed point Δg_i^{com} are computed. Using these anomalies and the calculated ones, the residuals are computed as $\delta\Delta g_i = \Delta g_i^{obs} - \Delta g_i^{com}$.
3. The Bouguer's formula is applied to the residuals as $\delta\Delta g_i = 2\pi G \Delta\rho \delta H_i$, and the model is upgraded as $\mathbf{H}_{n+1} = \mathbf{H}_n + \delta\mathbf{H}$.
4. The algorithm comes back to step 2, where the residuals are one again computed using the last estimated model as reference. [Bott \(1960\)](#) stops the algorithm after a fixed number of iterations, although another kind of criteria could be used, as those based on the relative differences between iteration in the evaluation of functionals as the one expressed in (1.20), for example.

This method has been used over the years with only minor changes. [Murthy and Rao \(1979\)](#) proposed the same method as [Bott \(1960\)](#), but using the exponential law for the density contrast variation with depth

$$\Delta\rho(z) = \Delta\rho_0 e^{-\lambda z}, \quad (1.25)$$

where $\Delta\rho_0$ is the density contrast at the terrain surface and λ is a variation parameter. The gravitational attraction over an external point due to a rectangle whose density follows the law (1.25) has not a closed expression in the spatial domain, so the attraction is approximated by discretizing the rectangle along the vertical direction and interpolating the density contrast for each slice according to the expressed law of variation. [Rao \(1986\)](#) used the variation law stated in equation (1.13), while in [Chakravarthi \(1995\)](#) the equation (1.16) is employed.

The method can also be used in 3D environments, using prisms as discretization figures instead of rectangles, as can be seen in [Chakravarthi and Sundararajan \(2004\)](#); [Chakravarthi et al. \(2013\)](#); [Chen and Zhang \(2015\)](#). In any case, 2D or 3D, the different authors do not perform any uncertainty assessment beyond the comparison between the estimated best model and known depth values at discrete points.

1.4.5 Global optimization methods

[Boschetti et al. \(1997\)](#) present a genetic algorithms ([Goldberg, 1989](#); [Holland, 1992](#)) application to joint inversion of gravity and magnetic data in sedimentary basins where, besides the optimal model, its uncertainty assessment is given. The 2D basin modeling is based on the approximation presented in section 1.4.2, being the unknowns the positions (along the vertical line) of the points defining the sediments-basement interface. The paper presents a synthetic example, which considers joint gravity and magnetic inversion and treats the density increment $\Delta\rho$ and magnetic susceptibility Δk also as unknowns. It also presents a real example, composed only by gravity data, where the density contrast is considered a known parameter. A computation

scheme is proposed in order to accelerate the processing time and to avoid working with a high number of parameters in the initial stages of the inversion. This scheme is based on the discretization of the model at low resolution in the initial iterations, and gradually increase the resolution as the number of iterations advances (Boschetti et al., 1995). For the joint inversion model, the objective function is

$$f(\mathbf{m}) = W_{\Delta k} \frac{\|\Delta \mathbf{k}^{obs} - \Delta \mathbf{k}^{com}\|_2^2}{M} + W_{\Delta g} \frac{\|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2}{N}, \quad (1.26)$$

where M and N are the number of points with magnetic and gravity observations, respectively, and $W_{\Delta k}$ and $W_{\Delta g}$ are weighting parameters controlling the influence of each type of data. The model and uncertainty estimation procedure works as follows:

1. All generated models which produce a value for the objective function below a given tolerance are selected. Some models, although numerically valid, are not geologically admissible, but they are also taken into account (this fact will be later discussed in section 2.2).
2. Each point position is averaged using all previously selected models. The final set of averaged points is adopted as problem solution. This method has as disadvantage that this *best model* could not generate the minimum value for the objective function among all computed candidates.
3. For each point, its standard deviation is computed using all selected models, and this value is imposed as its depth uncertainty. This procedure has as disadvantage that the extreme models, composed by the shallowest and deepest points, could be an uncertainty value bigger than the one adopted as tolerance.

In Parker (1999), a short chapter is devoted to the 3D inversion problem using prisms as discretization elements. The density contrast is considered as variable with depth following the law (specific for the area under study)

$$\rho(z) = 1.95 + 0.26z, \quad (1.27)$$

where 1.95 g/cm^3 is the density at the terrain surface and 0.26 g/cm^4 corresponds to its variation. The objective function to minimize is the same as the one shown in equation (1.20) and no uncertainty assessment is performed. In words of Parker (1999):

No smoothness constraint or regularization of any type were applied to the objective function, and the inversion appears somewhat unstable in regions with sparse data coverage. However, the same major features can be identified in all of the inversion results.

This fact, the *unstable* solutions, as in the case of Boschetti et al. (1997), will be addressed in detail in section 2.2.

Global optimization methods have been also employed in gravity inversion problems for the determination of other kind of structures than sedimentary basins. González Montesinos (1999); Montesinos et al. (2005) presented a method based on genetic algorithms focused on the work with complex and isolated structures. Camacho et al. (2002, 2007, 2011), inspired by René (1986), developed a method based on systematic exploration of the model space by accumulation of prisms, mainly focused on the inversion in complex environments with positive and negative density contrast, although the technique has been also tested in sedimentary basins (Camacho et al., 2012, 2015).

Thesis contributions

2.1 The effect of noise and Tikhonov regularization in inverse problems¹

The solution of inverse problems has an intrinsic uncertainty that is mainly caused by noise in data, incomplete data sampling and simplified physics. In the first part of this PhD Thesis (articles 1, 2 and 3, and publications 1 and 2), an analysis of the effect produced by noise and zero-order Tikhonov regularization in linear and nonlinear inverse problems is performed. In the case of ill-conditioned inverse problems the noise is amplified back to the model parameters through the generalized inverse operator. Ill-conditioning might provide spurious solutions (with no physical meaning) if no regularization techniques are used. For a given misfit tolerance the models that fit the observed data are called equivalent, and are located in a region of the model space that is bounded in the case of linear inverse problems by a linear hyper-quadric surface for a given tolerance error. The noise shifts the solution found by least squares methods and deforms homogeneously the topography of the cost function, while Tikhonov regularization transforms the linear hyper-quadric from an elliptical cylinder to a very oblong hyper-ellipsoid in the directions that originally spanned the kernel of the linear forward operator.

As in the linear case, noise shifts the solution found by nonlinear optimization, but in the nonlinear approach the cost function topography is deformed in a non-homogeneously way. Under certain conditions, noise might decrease the size of the regions of very low misfits, and at the same time increase the size of regions of medium misfits. As a result of that, the inverse problem solution becomes more difficult to be found by search methods, but locating the region (or regions) of medium misfit might be easier. The effect of the regularization in the linearized region of equivalence is similar to the linear case. Although the use of regularization techniques has a positive effect in decreasing the uncertainty of the solution that has been found, it unfortunately does not provoke the disappearance of the nonlinear equivalent models. This knowledge is very important because noise in data is always present, and the possibility of having at disposal good prior information is rare. Therefore, the nonlinear uncertainty analysis and solution appraisal is always needed. Also, the linearized region of equivalence does not coincide with the nonlinear region of equivalence, so an a posteriori uncertainty analysis based on the first region will not produce correct results.

¹This section summarizes the results presented in articles 1, 2 and 3, and in publications 1 and 2.

2.1.1 The effect of noise in linear inverse problems

The inverse problem presented in equation (1.1) is in general solved for a linear operator \mathbf{F} via linear least squares (LLS) methods finding

$$\mathbf{m}^\dagger : c(\mathbf{m}^\dagger) = \min_{\mathbf{m} \in \mathbb{R}^n} \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2, \quad (2.1)$$

where the generalized solution \mathbf{m}^\dagger of this optimization problem is provided by the Moore-Penrose pseudo-inverse (see for instance Aster et al. (2012)).

Let it be so that $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta\mathbf{d}$, where \mathbf{d}^{true} is the data that would be measured in absence of noise (noise-free data), and \mathbf{m}^{true} is the hypothetical model that has generated these data, that is, $\mathbf{F}(\mathbf{m}^{true}) = \mathbf{d}^{true}$. Then,

$$\mathbf{m}^\dagger = \mathbf{F}^\dagger \mathbf{d}^{obs} = \mathbf{F}^\dagger (\mathbf{d}^{true} + \delta\mathbf{d}) = \mathbf{m}^{true} + \Delta\mathbf{m}, \quad (2.2)$$

with

$$\Delta\mathbf{m} = \mathbf{F}^\dagger \delta\mathbf{d} = \mathbf{V}\Sigma^\dagger \mathbf{U}^T \delta\mathbf{d} = \mathbf{V}\Sigma^\dagger \delta\mathbf{d}_U = \sum_{k=1}^r \frac{\delta d_{Uk}}{\mu_k} \mathbf{v}_k, \quad (2.3)$$

where $r = \text{rank}(\mathbf{F})$, and $\delta\mathbf{d}_U = \mathbf{U}^T \delta\mathbf{d}$ is the noise term $\delta\mathbf{d}$ referred to the \mathbf{U} orthonormal base, provided by the singular value decomposition of the forward operator $\mathbf{F} = \mathbf{U}\Sigma\mathbf{V}^T$. Equation (2.3) can be split into two parts: (i) the one that provides the stable part, and (ii) the one that causes noise in data to be amplified back into the model parameters, providing different equivalent solutions. This second unstable part is associated with the \mathbf{v}_k vectors of the \mathbf{V} base corresponding to the vanishing singular values $\mu_k \rightarrow 0$. This effect is called ill-conditioning of the discrete linear inverse problem. So, in presence of noise, the solution found \mathbf{m}^\dagger will never coincide with true solution \mathbf{m}^{true} , that has generated the observed data. The ill-conditioned character of the discrete inverse problems does not only cause the numerical determination of the model of lower misfit to be unstable, but more important, it is the origin of the uncertainty problem. The condition number κ of the matrix \mathbf{F} is defined as the ratio of its maximum and minimum singular values:

$$\kappa(\mathbf{F}) = \frac{\mu_{\max}}{\mu_{\min}}, \quad (2.4)$$

and $\kappa(\mathbf{F})$ is related to the eccentricity of the linear equivalence region as follows:

$$\varepsilon = \sqrt{1 - \left(\frac{\mu_{\min}}{\mu_{\max}}\right)^2} = \sqrt{1 - \left(\frac{1}{\kappa(\mathbf{F})}\right)^2}. \quad (2.5)$$

Ill-conditioned problems are those whose region of equivalence eccentricity tends to 1 and the equivalent models are located in oblong and elongated rectilinear valleys of the cost function topography. Ill-conditioning should also be viewed as a geometric property of the linear equivalence region (hyper-quadric). Thus, the following alternative definition might be proposed: *A linear inverse problem $\mathbf{F}(\mathbf{m}) \simeq \mathbf{d}^{obs}$ is ill-conditioned if the eccentricity of its region of equivalence is close to 1* (see article 1 and publication 1 for details). This first paper was devoted to justify the uncertainty analysis of discrete inverse problems and it is a continuation of the PhD Thesis of María Zulima Fernández Muñiz (Fernández Muñiz, 2012).

Let's be examined more in detail how noise affects the region of equivalent solutions, defined as $L_{tol} = \{\mathbf{m} \in \mathbf{M} : \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 \leq tol^2\}$. As

$$\|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2 = \|\mathbf{F}\mathbf{m}\|_2^2 - 2(\mathbf{F}\mathbf{m})^T \mathbf{d}^{obs} + \|\mathbf{d}^{obs}\|_2^2, \quad (2.6)$$

the boundary of the L_{tol} region is the hyper-quadric containing the linear equivalent models predicting the observed data with error tolerance tol (Fernández-Martínez et al., 2012a)

$$\|\mathbf{Fm}\|_2^2 - 2(\mathbf{Fm})^T \mathbf{d}^{obs} + \|\mathbf{d}^{obs}\|_2^2 - tol^2 = 0, \quad (2.7)$$

or

$$\|\Sigma \mathbf{m}_V\|_2^2 - 2(\Sigma \mathbf{m}_V)^T \mathbf{d}_U^{obs} + \|\mathbf{d}_U^{obs}\|_2^2 - tol^2 = 0, \quad (2.8)$$

with $\mathbf{d}_U^{obs} = \mathbf{U}^T \mathbf{d}^{obs}$, and $\mathbf{m}_V = \mathbf{V}^T \mathbf{m}$. In this last case the hyper-quadric is referred to the \mathbf{U} and \mathbf{V} orthogonal bases provided by the singular value decomposition of \mathbf{F} .

Taking into account equations (2.7) and (2.8), the following can be observed:

1. Noise in data does not affect the matrix of the hyper-quadric, $\mathbf{F}^T \mathbf{F}$. Thus, the orientation of the linear hyper-quadric (\mathbf{V} principal axes) remains the same.
2. Noise perturbs the center of the hyper-quadric, which coincides with the least squares solution $\mathbf{m}_c = \mathbf{m}^\dagger$ of the linear inverse problem stated in equation (2.2).
3. Finally, noise in data deforms homogeneously the topography of the unperturbed cost function in the region of equivalent models, bounded by L_{tol} .

The last assertion can be proved analytically. Equation (2.8) can be written as (see article 2 for details)

$$\sum_{k=1}^r (\mu_k m_{V_k} - d_{U_k}^{obs})^2 = tol^2 - \sum_{k=r+1}^s d_{U_k}^{obs^2}, \quad (2.9)$$

where, if $r = \text{rank}(\mathbf{F}) < n$, the components m_{V_k} with $k = r+1, \dots, n$ are not determined by the linear system. Geometrically, the hyper-quadric becomes an elliptical cylinder in the directions of \mathbf{v}_k with $k = r+1, \dots, n$ (Fernández-Martínez et al., 2012a). Considering $d_{U_k}^{obs} = d_{U_k}^{true} + \delta d_{U_k}$, the hyper-quadric for the noise-free case is

$$\sum_{k=1}^r (\mu_k m_{V_k} - d_{U_k}^{true})^2 = tol^2 - \sum_{k=r+1}^s d_{U_k}^{true^2}. \quad (2.10)$$

The following condition has to be fulfilled for both hyper-quadrics, with and without noise, to have the same axes length

$$tol^* = \sqrt{tol^2 + \sum_{k=r+1}^s [(d_{U_k}^{true} + \delta d_{U_k})^2 - d_{U_k}^{true^2}]} = \sqrt{tol^2 + \sum_{k=r+1}^s (\delta d_{U_k}^2 + 2\delta d_{U_k} d_{U_k}^{true})}. \quad (2.11)$$

Calling $\Delta tol = \sum_{k=r+1}^s (\delta d_{U_k}^2 + 2\delta d_{U_k} d_{U_k}^{true})$,

1. If $\Delta tol > 0$, then $tol^* > tol$, that is, in presence of noise the region with the same size corresponds to a region of higher misfit. Thus, the regions of equivalent models will decrease in size. Particularly this will happen if $\sum_{k=r+1}^s \delta d_{U_k} d_{U_k}^{true} = 0$ as, for instance, in presence of white noise. In this case locating the equivalence region of low misfits ($tol \leq 10\%$ for instance) by means of global optimization methods becomes a harder task due to its smaller size. Conversely, finding the global optimum of the perturbed cost function via local optimization methods inside this region is easier. Although the noise $\delta \mathbf{d}$ does not affect the conditioning of the forward operator \mathbf{F} , it could be considered as

a kind of regularization since for the same value of the error tolerance the axes of the hyper-quadric become smaller. Finally when using global optimization methods under the presence of noise, the posterior search is usually performed in a region of higher misfit than the one corresponding to the noise-free case.

2. Conversely, if $\Delta tol < 0$, then $tol^* < tol$, and the region with the same size in presence of noise corresponds to a region of lower misfit. In this case, the regions of equivalent models will increase in size. This fact could be used to locate more easily the low misfit region by artificially adding an adequate colored noise to the observed data. This case could be happen in practice only for noise which is negatively correlated with the true data. Future research will be devoted to this subject.

These theoretical results are confirmed via a synthetic numerical example concerning a linear regression problem $y = \alpha x + \beta + \delta$, where (α, β) are the unknown parameters and δ represents the noise. The experiment consists of generating a dataset of 100 different (x_i, y_i) points, stating as true (noise-free) model parameters the values $\alpha_t = 4$ and $\beta_t = 2$. Two different levels of white Gaussian noise $\delta_1 \rightarrow N(0, 0.05)$ and $\delta_2 \rightarrow N(0, 0.075)$ were added to the noise-free data. Figure 2.1 (left) shows the 9% relative misfit contour lines for each case. According to equation (2.11), the original 9% equivalence region corresponds in size to the regions of 10.12% and 11.88% for δ_1 and δ_2 , respectively. Also, the least squares solutions are shifted to the models $(\alpha_1 = 3.90, \beta_1 = 2.14)$ and $(\alpha_2 = 3.84, \beta_2 = 2.25)$.

In conclusion, in linear inverse problems noise shifts the center of the linear region of equivalence (hyper-quadric), that coincides with the least squares solution, and deforms homogeneously the topography of the cost function. The term homogeneously refers to the fact that the deformation does not depend on the model \mathbf{m}_0 that is considered in this analysis.

2.1.2 The effect of Tikhonov regularization in linear inverse problems

Regularization refers to the mathematical techniques that are used to stabilize the inversion in linear and nonlinear problems. In this section, the role of the regularization and its effect on the cost function topography for linear inverse problems is analyzed, showing geometrically its effect on the estimated solution. Regularization in linear inverse problems can be performed by truncation and by damping (see article 2 for details). Zero-order Tikhonov regularization is the most common damping strategy, and consists of minimizing a linear combination of the data prediction misfit and the distance to a reference model \mathbf{m}^{ref} :

$$\mathbf{m}^\varepsilon : c^r(\mathbf{m}) = \min_{\mathbf{m} \in \mathbb{R}^n} \left[\left\| \mathbf{F}\mathbf{m} - \mathbf{d}^{obs} \right\|_2^2 + \varepsilon^2 \left\| \mathbf{m} - \mathbf{m}^{ref} \right\|_2^2 \right], \quad (2.12)$$

where ε^2 is the so-called damping parameter, generally close to zero.

Now, the normal equations system is

$$\left(\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I} \right) \cdot \left(\mathbf{m} - \mathbf{m}^{ref} \right) = \mathbf{F}^T \mathbf{d}^{obs}, \quad (2.13)$$

and the solution is computed via least squares as (see Aster et al. (2012) for example)

$$\mathbf{m}^\varepsilon = \mathbf{m}^{ref} + \left(\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I} \right)^{-1} \mathbf{F}^T \left(\mathbf{d}^{obs} - \mathbf{F}\mathbf{m}^{ref} \right), \quad (2.14)$$

where the operator $\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}$ is full rank, so $(\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I})^{-1}$ exists and is continuous. Equation (2.14) can be expressed using the singular value decomposition $\mathbf{F} = \mathbf{U}\Sigma\mathbf{V}^T$ as

$$\mathbf{m}_V^\varepsilon = \mathbf{m}_V^{ref} + \left(\Sigma^T \Sigma + \varepsilon^2 \mathbf{I} \right)^{-1} \Sigma^T \Delta \mathbf{d}_U^{ref}, \quad (2.15)$$

where $\Delta \mathbf{d}_U^{ref}$ is the reference model data misfit referred to the \mathbf{U} base, i. e.,

$$\Delta \mathbf{d}_U^{ref} = \mathbf{U}^T (\mathbf{d}^{obs} - \mathbf{F} \mathbf{m}^{ref}) = \mathbf{d}_U^{obs} - \Sigma \mathbf{m}_V^{ref}. \quad (2.16)$$

The solution can be finally expressed as (see article 2 for details)

$$\mathbf{m}_V^\varepsilon = \sum_{k=1}^r \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}}{\mu_k^2 + \varepsilon^2} \mathbf{v}_k + \sum_{k=r+1}^n m_{V_k}^{ref} \mathbf{v}_k, \quad (2.17)$$

where, as in the previous section, μ_k are the singular values stored in Σ .

All the vectors in the \mathbf{V} base play a role in the regularized solution \mathbf{m}^ε since the coordinates that originally resided in the kernel of \mathbf{F} are now informed by the reference model \mathbf{m}^{ref} . Also, the coordinates in the r first vectors of the \mathbf{V} base are linear combinations of the reference model coordinates $m_{V_k}^{ref}$ and the data d_{U_k} , being the weights $\varepsilon^2/(\mu_k^2 + \varepsilon^2)$ and $\mu_k/(\mu_k^2 + \varepsilon^2)$, respectively. Thus, the reference model serves to incorporate information from the kernel.

Referred to the \mathbf{V} base, the hyper-quadric for this problem can be written as

$$\mathbf{m}_V^T (\Sigma^T \Sigma^{-1} \varepsilon^2 \mathbf{I}) \mathbf{m}_V - 2 \mathbf{m}_V^T (\Sigma^T \mathbf{d}_U^{obs} + \varepsilon^2 \mathbf{m}_V^{ref}) = tol^2 - \mathbf{d}_U^{obs T} \mathbf{d}_U^{obs} - \varepsilon^2 \mathbf{m}_V^{ref T} \mathbf{m}_V^{ref}, \quad (2.18)$$

that can be reformulated as

$$\begin{aligned} & \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} m_{V_k} - \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \varepsilon^2 \sum_{k=r+1}^n (m_{V_k} - m_{V_k}^{ref})^2 = \\ & = tol^2 + \sum_{k=1}^r \frac{(\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs})^2}{\mu_k^2 + \varepsilon^2} - \|\mathbf{d}_U^{obs}\|_2^2 - \varepsilon^2 \sum_{k=1}^r m_{V_k}^{ref 2}. \end{aligned} \quad (2.19)$$

Finally, calling

$$\Delta_r = \sum_{k=1}^r \frac{(\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs})^2}{\mu_k^2 + \varepsilon^2} - \|\mathbf{d}_U^{obs}\|_2^2 - \varepsilon^2 \sum_{k=1}^r m_{V_k}^{ref 2}, \quad (2.20)$$

the axes of the hyper-quadric are

$$\left[\frac{\sqrt{tol^2 + \Delta_r}}{\sqrt{\mu_i^2 + \varepsilon^2}}, \left(\frac{\sqrt{tol^2 + \Delta_r}}{\varepsilon} \right)_j \right]_{\substack{i=1, \dots, r \\ j=r+1, \dots, n}}, \quad (2.21)$$

instead of

$$\left[\frac{\sqrt{tol^2 - \|\mathbf{d}_U^{obs}\|_2^2 + \sum_{k=1}^r d_{U_k}^{obs 2}}}{\mu_i}, \infty_j \right]_{\substack{i=1, \dots, r \\ j=r+1, \dots, n}}, \quad (2.22)$$

with no regularization. The regularization with a model of reference \mathbf{m}^{ref} prescribes the center of the hyper-quadric to $m_{V_j}^{ref}$ for the coordinates of the solution that originally resided in the kernel of \mathbf{F} , and bounds the axes in these \mathbf{v}_j directions to $\sqrt{tol^2 + \Delta_r}/\varepsilon$, instead of infinity. Accordingly, the equivalence region changes from an elliptical cylinder (with no regularization) to a very oblong hyper-ellipsoid in the directions of the kernel of \mathbf{F} .

The following relationship applies for both equivalent regions of tolerance tol^* (with regularization) and tol (without regularization) to have the same size (or axes)

$$\frac{\sqrt{tol^{*2} + \Delta_r}}{\sqrt{\mu_i^2 + \varepsilon^2}} = \frac{\sqrt{tol^2 - \sum_{k=r+1}^s d_{U_k}^{obs2}}}{\mu_i}, \quad i = 1, \dots, r, \quad (2.23)$$

where the right part comes from equation (2.9). Thus

$$tol^* = \sqrt{\frac{\mu_i^2 + \varepsilon^2}{\mu_i^2} \left(tol^2 - \sum_{k=r+1}^s d_{U_k}^{obs2} \right) - \Delta_r}, \quad (2.24)$$

which, taking into account the value of Δ_r can be expanded to (see article 2 for details)

$$tol^*(\mu_i) = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left(tol^2 - \sum_{k=r+1}^s d_{U_k}^{obs2} \right) + \varepsilon^2 \sum_{k=1}^r \frac{(d_{U_k}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}}. \quad (2.25)$$

In this case, tol^* depends on the axes index i , that is, the deformation is different for each axis of the hyper-quadric due to the effect of regularization. The values of $tol^*(\mu_i)$ depend on each singular value and, thus, the axes of the hyper-quadric associated to the biggest singular values (the directions of smaller uncertainty) might keep unchanged by the regularization. Figure 2.1 (right) shows the effect of the regularization in the synthetic linear regression case presented in section 2.1.1. The regularization parameter was set to $\varepsilon^2 = 0.5$, for which the regularized least squared solution found was $(\alpha = 3.84, \beta = 2.23)$. According to the presented analysis in this section, it is possible to observe that 9% equivalence region decreases in size, with respect to the not regularized case, mainly in the directions of maximum uncertainty, that are associated to the smallest singular values of the forward operator. Using equation (2.25), this anisotropic deformation can be computed, obtaining the values $tol_{\min}^* = 9.01\%$ and $tol_{\max}^* = 9.46\%$, and thus confirming the theoretical results.

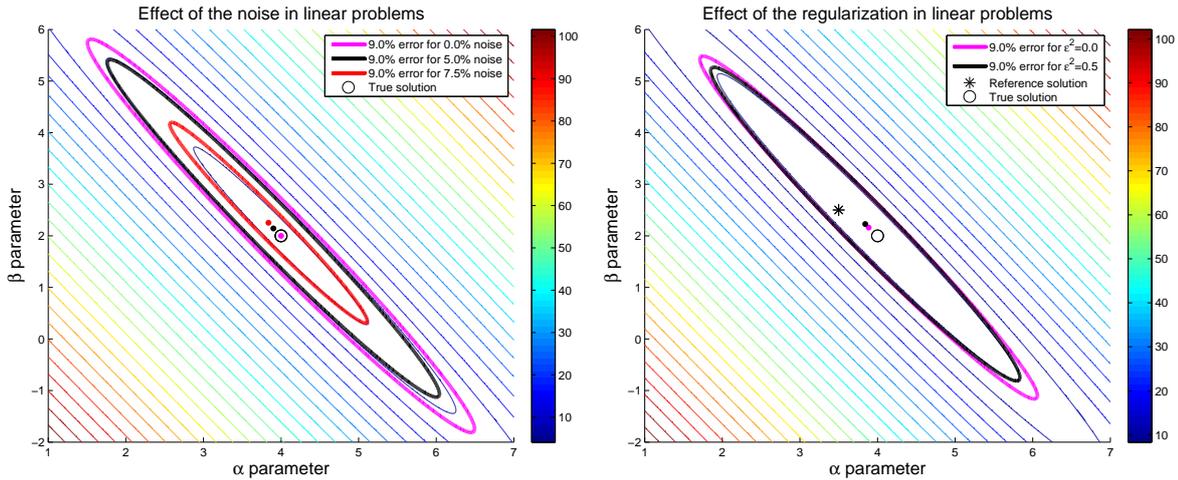


Figure 2.1: Different cost functions in linear inverse problems. Left: effect of different levels of noise. Right: effect of the zero-order Thikonov regularization.

In conclusion, in linear inverse problems the regularization deforms homogeneously (it does not depend on the model that has been considered) and anisotropically (it depends on the axes of the equivalence region) the region of equivalence.

2.1.3 The effect of noise in nonlinear inverse problems

It has been shown in [Fernández-Martínez et al. \(2012a\)](#) that in the case of nonlinear problems the nonlinear region of equivalence is one or more elongated valleys with curvilinear shapes and almost null gradients. Also, depending on the injectivity of the forward operator several basins of low misfit can coexist in the cost function landscape. To explore (see article 3 and publications 1 and 2 for details) how the noise affects the cost function landscape for this kind of problems a linearization of the forward operator \mathbf{F} in a model \mathbf{m}_0 , that belongs to the nonlinear equivalent region is adopted:

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_0) + \mathbf{J}\mathbf{F}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + o(\|\mathbf{m} - \mathbf{m}_0\|_2), \quad (2.26)$$

where, in this case, $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^s$ is a nonlinear operator and $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}$ is its Jacobian in \mathbf{m}_0 . The nonlinear equivalence region of value tol is defined as

$$V_{tol} = \left\{ \mathbf{m} \in \mathbf{M} : c(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 \leq tol^2 \right\}, \quad (2.27)$$

where $\mathbf{m} \in \mathbb{R}^n$, $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta\mathbf{d} \in \mathbb{R}^s$, and $\delta\mathbf{d}$ represents, as in section 2.1.1, the noise.

Substituting equation (2.26) into equation (2.27), the linearized hyper-quadric L_{tol} of value tol in \mathbf{m}_0 is

$$\Delta\mathbf{m}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + 2(\Delta\mathbf{d} - \delta\mathbf{d})^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + \|\Delta\mathbf{d} - \delta\mathbf{d}\|_2^2 < tol^2, \quad (2.28)$$

where $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ and $\Delta\mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}$, equation that can be written, taking into account the singular value decomposition $\mathbf{J}\mathbf{F}_{\mathbf{m}_0} = \mathbf{U}\Sigma_{\mathbf{m}_0}\mathbf{V}^T$, as

$$\Delta\mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + 2\Delta\mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + \|\Delta\mathbf{b}_U\|_2^2 < tol^2, \quad (2.29)$$

where $\Delta\mathbf{b}_U = \Delta\mathbf{d}_U - \delta\mathbf{d}_U = \mathbf{U}^T[\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true} - \delta\mathbf{d}]$ and $\Delta\mathbf{m}_V = \mathbf{V}^T \Delta\mathbf{m}$.

Equation (2.29) can be rewritten as (see article 3 for details)

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta b_{U_k})^2 = tol^2 - \sum_{k=r+1}^s \Delta b_{U_k}^2, \quad (2.30)$$

where r is the rank of $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}$ and μ_k the corresponding singular values. For the noise-free case, $\delta\mathbf{d} = 0$, the equation becomes

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta d_{U_k})^2 = tol^2 - \sum_{k=r+1}^s \Delta d_{U_k}^2. \quad (2.31)$$

The tol equivalent region in the noisy case will have the same size as the corresponding region for the noise-free case under the condition

$$tol^{*2}(\mathbf{m}_0) - \sum_{k=r+1}^s \Delta b_{U_k}^2 = tol^2(\mathbf{m}_0) - \sum_{k=r+1}^s \Delta d_{U_k}^2, \quad (2.32)$$

then

$$tol^*(\mathbf{m}_0) = \sqrt{tol^2(\mathbf{m}_0) + \sum_{k=r+1}^s (\delta d_{U_k}^2 - 2\Delta d_{U_k} \delta d_{U_k})}, \quad (2.33)$$

where the regions are called $tol^*(\mathbf{m}_0)$ and $tol(\mathbf{m}_0)$ due to its dependency on the model \mathbf{m}_0 adopted to linearize the forward operator. By continuity of the cost function there might be a neutral line where $tol^*(\mathbf{m}_0) = tol(\mathbf{m}_0)$, so no deformation takes place along it. Inside the region limited by this line $tol^*(\mathbf{m}_0) > tol(\mathbf{m}_0)$, and regions decrease in size in presence of noise, and outside, $tol^*(\mathbf{m}_0) < tol(\mathbf{m}_0)$, so regions increase in size. In conclusion, noise deforms the topography of the cost function non-homogeneously, decreasing regions of low misfit and increasing the regions of medium misfit, if certain conditions are fulfilled.

To show the effect of noise in the nonlinear inverse problem case, a simple synthetic example $y = \alpha(1 - e^{-\beta x}) + \delta$ was analyzed, generating synthetically a dataset with 100 points. In addition to the true noise-free solution ($\alpha_t = 20, \beta_t = 0.1$), the problem was solved also adding white noise of level $\delta \rightarrow N(0, 0.05)$. The model ($\alpha_0 = 19, \beta_0 = 0.15$) was set as initial model for the nonlinear least squares optimization. Figure 2.2 (left) shows the linearized and nonlinear equivalence regions for a relative error tolerance of 9% for these two cases. The figure background shows the contour lines of the corresponding noise-free linearized problem. The nonlinear equivalence regions show the croissant shape, which is typical of this kind of problems (Fernández-Martínez et al., 2012a), while the linearized regions represent the hyper-quadric around the solution that has been adopted. It is possible to observe that the linearized equivalence regions only span locally the full nonlinear uncertainty region, and also that both regions decrease in size as the noise level increases.

2.1.4 The effect of Tikhonov regularization in nonlinear inverse problems

The nonlinear equivalent region of value tol for the zero-order Tikhonov regularized inverse problem is

$$V_{tol} = \left\{ \mathbf{m} \in \mathbf{M} : c^r(\mathbf{m}) = \left\| \mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs} \right\|_2^2 + \varepsilon^2 \left\| \mathbf{m} - \mathbf{m}^{ref} \right\|_2^2 \leq tol^2 \right\}, \quad (2.34)$$

where, as in section 2.1.2, ε^2 is the regularization parameter and \mathbf{m}^{ref} is a reference model.

Adopting a linearization of $c^r(\mathbf{m})$ in a model \mathbf{m}_0 , located on the nonlinear equivalent region, it is possible to obtain the linearized region of equivalence for this case as

$$\Delta \mathbf{m}^T \mathbf{J} \mathbf{F}_{\mathbf{m}_0}^T \mathbf{J} \mathbf{F}_{\mathbf{m}_0} \Delta \mathbf{m} + 2(\Delta \mathbf{d} - \delta \mathbf{d})^T \mathbf{J} \mathbf{F}_{\mathbf{m}_0} \Delta \mathbf{m} + \varepsilon^2 \left\| \Delta \mathbf{m} - \Delta \mathbf{m}^{ref} \right\|_2^2 + \left\| \Delta \mathbf{d} - \delta \mathbf{d} \right\|_2^2 < tol^2, \quad (2.35)$$

where, as in previous section, $\mathbf{J} \mathbf{F}_{\mathbf{m}_0}$ is the Jacobian of the nonlinear forward operator \mathbf{F} in \mathbf{m}_0 , $\Delta \mathbf{m} = \mathbf{m} - \mathbf{m}_0$ and $\Delta \mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}$. Taking into account the singular value decomposition $\mathbf{J} \mathbf{F}_{\mathbf{m}_0} = \mathbf{U} \Sigma_{\mathbf{m}_0} \mathbf{V}^T$, equation (2.35) can be written as

$$\Delta \mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta \mathbf{m}_V + 2 \Delta \mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta \mathbf{m}_V + \varepsilon^2 \left\| \Delta \mathbf{m}_V - \Delta \mathbf{m}_V^{ref} \right\|_2^2 + \left\| \Delta \mathbf{b}_U \right\|_2^2 < tol^2, \quad (2.36)$$

where $\Delta \mathbf{b}_U = \Delta \mathbf{d}_U - \delta \mathbf{d}_U = \mathbf{U}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true} - \delta \mathbf{d}]$, $\Delta \mathbf{m}_V = \mathbf{V}^T \Delta \mathbf{m}$ and $\Delta \mathbf{m}_V^{ref} = \mathbf{V}^T \Delta \mathbf{m}^{ref}$, that can be reformulated as

$$\begin{aligned} & \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} \Delta m_{V_k} - \frac{\varepsilon^2 \Delta m_{V_k}^{ref} + \mu_k \Delta b_{U_k}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \varepsilon^2 \sum_{k=r+1}^n \left(\Delta m_{V_k} - \Delta m_{V_k}^{ref} \right)^2 = \quad (2.37) \\ & = tol^2 + \sum_{k=1}^r \frac{\left(\varepsilon^2 \Delta m_{V_k}^{ref} + \mu_k \Delta b_{U_k} \right)^2}{\mu_k^2 + \varepsilon^2} - \left\| \Delta \mathbf{b}_U \right\|_2^2 - \varepsilon^2 \sum_{k=1}^r \Delta m_{V_k}^{ref^2}. \end{aligned}$$

Performing a similar analysis to the one in the linear case, here for the linearized equivalent regions, with and without regularization, the following relationship is found:

$$tol_i^*(\mathbf{m}_0) = \sqrt{tol^2(\mathbf{m}_0) + \frac{\varepsilon^2}{\mu_i^2} \left(tol^2(\mathbf{m}_0) - \sum_{k=r+1}^s \Delta b_{U_k}^2 \right) + \varepsilon^2 \sum_{k=1}^r \frac{(\Delta b_{U_k} + \mu_k \Delta m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}}, \quad (2.38)$$

where tol_i^* is the tolerance in presence of regularization and tol without it.

Figure 2.2 (right) shows the effect of the regularization in the regression synthetic case presented in section 2.1.3. In this numerical example, the level of white noise was $\delta \rightarrow N(0, 0.05)$, the regularization parameter was set to ε^2 , the reference model was $(\alpha_r = 15, \beta_r = 0.07)$, and the initial model to perform the nonlinear least squares optimization was $(\alpha_0 = 19, \beta_0 = 0.15)$. The figure shows the 9% equivalence regions for the linearized and full nonlinear problems. It can be observed that the linearized equivalence regions decrease in size anisotropically as the regularization parameter increases. The size of the nonlinear equivalence region also decreases in size less rapidly than the linearized region.

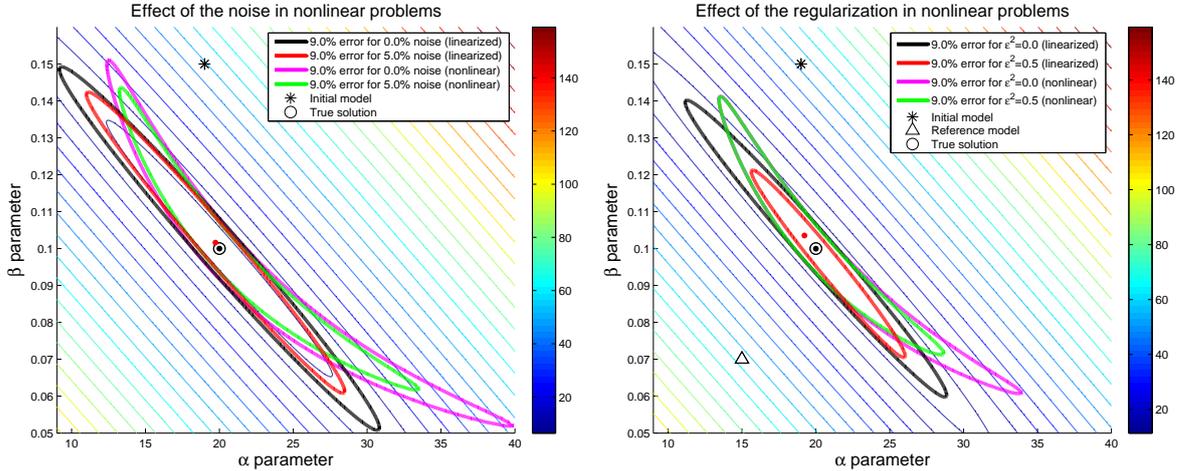


Figure 2.2: Different cost functions in nonlinear inverse problems. Left: effect of noise. Right: effect of the zero-order Tikhonov regularization.

The conclusions arising from this analysis are the following:

1. The center of the linearized hyper-quadratic provides the regularized solution of the linearized inverse problem in \mathbf{m}_0 .
2. The noise and the regularization deform nonlinearly the cost function topography in the neighborhood of \mathbf{m}_0 , acting differently in each model component. Thus, the regularization and the noise in data have similar effects on the cost function landscape, although some differences do exist, since the regularization also acts in an anisotropic fashion, deforming differently each of the axes of the linearized region of equivalence.
3. The nonlinear and linearized equivalent regions are completely different. It has been analytically proved that the nonlinear equivalent region has a meandering valley shape. Also, there could be other low misfit basins in the cost function landscape if the forward operator is not injective (Fernández-Martínez et al., 2012a). This is one of the main differences between the linear and nonlinear inverse problems. It can be easily understood

that in the non-injective case the regularization might provide different solutions located in different basins of the nonlinear equivalent region, depending on the reference model that has been adopted. These solutions might not coincide with the model parameters that have generated the observed data.

2.2 Gravity inversion in sedimentary basins using PSO²

One of the main results of this PhD Thesis is the proposal of a 2D and 3D gravity inversion method for sedimentary basins based on the *Particle Swarm optimization* (PSO) algorithm, which has been introduced in section 1.3. In this section, the developed technique (published in articles 4 and 5) is explained in detail.

2.2.1 Domain discretization

A model discretization based on the accretion of rectangles for 2D and prisms for 3D has been selected, as it was shown in figures 1.3 and 1.4. The main reason for this selection is the operational and programming simplicity, as well as the ease of increasing the model resolution by diminishing the size of the prisms in the horizontal plane.

Although for computation tasks it is not necessary to have equal rectangle dimensions, working with equal sizes facilitates the treatment of models and their posterior filtering, as it will be explained in section 2.2.3. Moreover, if the observed points are homogeneously distributed in the working area, the use of equal horizontal dimensions for the rectangles and prisms is fully justified. After several tests, it has been concluded that these horizontal dimensions can be selected as a factor between $0.5\times$ and $1.5\times$ of the mean separation among the observed points. In 3D cases, horizontal dimensions can also be different in the X and Y axes if the points distribution suggests it. Regarding the prisms' top sides, in this work they will be located at the terrain surface. Many sedimentary basins have flat surfaces, so situating all prisms at a reference level is a common practice, but there are other environments where the basin surface has an important slope, as in the example shown in article 4 in Atacama (Chile).

In the case of the gravity computations for the 2D model the used formulation is the classical one stated in Telford et al. (1976); Barbosa and Silva (1994), and is detailed in article 4. Regarding to the 3D model, two approximations (that can be seen in article 5) are used:

- For the adjacent area to the attracted point, i. e., for the prism on which the observed point is located, and its neighbors in a 3×3 grid, the exact attraction formula of Nagy et al. (2000) is used.
- For the rest of prisms the approximated formulation of MacMillan (1958) (also detailed in Fullea et al. (2008)) is used. This method helps to speed up computations and is accurate enough.

2.2.2 Density contrast treatment

The density contrast between sediments and basement, $\Delta\rho = \rho_s - \rho_b$, is a parameter of fundamental importance in sedimentary basins gravity inversion and in the results interpretation. In many nonlinear approaches, $\Delta\rho$ is a priori fixed and remains unchanged through iterations. Nevertheless, due to the compaction, the density of the sediments, ρ_s , generally increases with

²This section summarizes the results presented in articles 4 and 5.

depth and, therefore, the density contrast $\Delta\rho$ decreases. This effect should be taken into account in the inversion. As it was pointed in section 1.4, there are multiple density variation models, for which the gravity attraction formulation can be found in the literature, although some which are very useful, like the one presented in equation (1.25), have not closed expression in the space domain.

In this PhD Thesis a numerical approximation to the direct problem is proposed, which allows the use of any density distribution, analytical or not. This idea is inspired by the method used by Murthy and Rao (1979). Valid for 2D and 3D bodies, the method is based on the discretization of the rectangles or prisms into smaller bodies along the vertical. Then, the attraction of each sub rectangle or prism is computed, whose densities are considered constant (and can be obtained through interpolation in the used density distribution), and finally all contributions are added in order to obtain the attraction of the entire body. This method has the advantage of being able to be adapted to real density distributions, such as the ones coming from boreholes.

Table 2.1: Exact and approximate gravity anomalies, in μGal , generated by a rectangle of 250 m width and 1000 m depth, and variable density contrast with depth following equation (1.16) and parameters $\Delta\rho_0 = -500 \text{ kg/m}^3$ and $\alpha = 0.1 \text{ kg/m}^4$. The attracted point is situated at the middle position in the upper side.

Sub rectangles	Δg_e	Δg_a	$\Delta g_e - \Delta g_a$
5	-4619.722	-4600.226	-19.496
10	-4619.722	-4614.738	-4.984
25	-4619.722	-4618.923	-0.799
50	-4619.722	-4619.522	-0.200
100	-4619.722	-4619.672	-0.050

The minimum number of sub elements in which the original rectangle or prism should be divided in order to obtain an adequate precision for the attraction depends on the function variation parameters, the basin average depth and the data uncertainty. Nevertheless, if borehole data are available visual inspection can give valuable information about the right thickness. Table 2.1 shows the gravity anomaly values Δg produced by a rectangle of 250 m width, 1000 m depth and variable density contrast following equation (1.16) with parameters $\Delta\rho_0 = -500 \text{ kg/m}^3$ and $\alpha = 0.1 \text{ kg/m}^4$, where the attracted point is situated in the middle of the rectangle top side. In order to properly select the partition size, it is necessary to take the estimated uncertainty of the observations into account, and compare it to the error contribution due to the prisms partition approximation.

2.2.3 Search space, constraints, and inversion

The proposed inversion method includes simultaneous polynomial regional trend estimation in 2D and 3D. As it was pointed out in section 1.4, the polynomial degree selection depends on various factors, mainly the basin extension. Generally, unless there is conclusive evidence for using higher degrees, it is desirable to use low degree polynomials, such as the line and the plane in 2D and 3D, respectively. The main reason for that is the fact that high degree polynomials can absorb part of the anomaly generated by the basin itself.

The parameters searching space definition is a fundamental step in any optimization method based on global search algorithms. For the problem proposed in this PhD Thesis, the model is composed by:

1. The regional trend polynomial coefficients.
2. The rectangles or prisms bottom heights in which the working area is divided.

The search limits for the polynomial parameters (supposing a 2D problem —3D generalization is straightforward— and a trend modeling as $\Delta g_i^{reg} = A + Bx_i$, where x_i is the planimetric coordinate of each working point) can be computed via multiple least squares fitting, randomly perturbing the observations by an amount at the level of the expected anomaly noise. From the estimated A and B sets the extreme values A^{\min} , A^{\max} , B^{\min} and B^{\max} are selected, and these extreme quantities define the search space for the regional trend factors. Finally, a first approximation to the regional trend (any set of parameters computed in the previous step can be used) is subtracted from the original observations in order to obtain the initial anomaly residual values for the interface sediments-basement search limits computation.

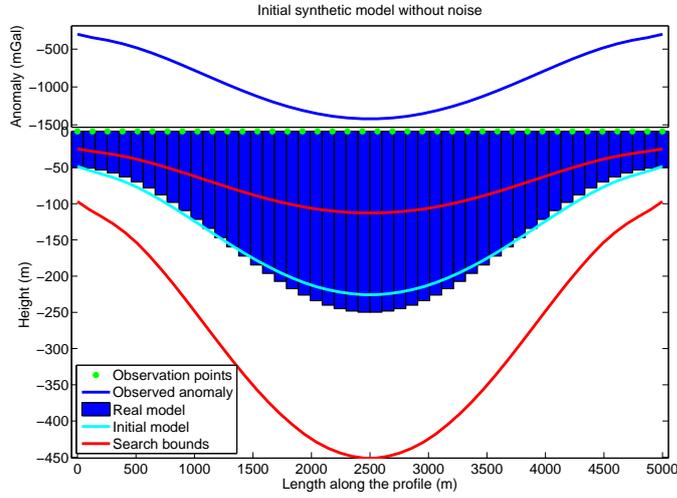


Figure 2.3: Search space limits (red lines) for a synthetic 2D gravity inversion problem. The applied factors to the approximate model (cyan line) for limits computation have been $k^{\min} = 0.5$ and $k^{\max} = 2.0$. The model is composed by 50 rectangles of 100 m width and constant density contrast $\Delta\rho = -150 \text{ kg/m}^3$, whose upper sides are situated at height 0 m. The observations set is composed by 40 equally spaced points (green dots) at height 0 m and without noise.

The portion of the search model corresponding to the interface heights will be built starting at an initial depth model \mathbf{z}^0 (which will be not used in the inversion process). This \mathbf{z}^0 model is computed using the initial residual anomalies and the Bouguer plate formula, in the same way as the initial model was computed in section 1.4.4. Figure 2.3 shows the context in a 2D synthetic case, taken from article 4. The cyan line is the initial model \mathbf{z}^0 from which, and by applying two scaling factors $k^{\min} < 1$ and $k^{\max} > 1$, the minimum \mathbf{z}_b^{\min} and maximum \mathbf{z}_b^{\max} heights for the search space are computed (red lines). This example is a noise-free case and represents a shallow basin with constant density contrast, so the initial model is close to the true one (blue bars), which is not common in real environments. It is necessary to stress that the initial model is only used for the limits computation, having no influence on the rest of the inversion process.

Once the search limits are defined, the inversion starts. The functional to minimize in the 2D and in the 3D cases is the data prediction error

$$f(\mathbf{z}) = \sum_{i=1}^N \left[\Delta g_{P_i}^{obs} - \Delta g_{P_i}^{com} \right]^2 = \|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2, \quad (2.39)$$

where $\Delta\mathbf{g}^{obs}$ is the observed anomaly, and $\Delta\mathbf{g}^{com}$ the computed one. From these values, the relative value (as a percentage) of the cost function for each generated model k can be computed as

$$c(\mathbf{z})_k^r = \frac{\|\Delta\mathbf{g}^{obs} - \Delta\mathbf{g}_k^{com}\|_2^2}{\|\Delta\mathbf{g}^{obs}\|_2^2} \cdot 100, \quad (2.40)$$

where the regional trend must be subtracted from $\Delta\mathbf{g}^{obs}$ and $\Delta\mathbf{g}_k^{com}$ in order to take only the influence of the anomalies produced by the basin itself.

Figure 2.4 shows the inversion results for the model shown in figure 2.3. The problem has been solved via PP-PSO, with a swarm size of 250 individuals and 150 iterations. It can be shown as blue lines all the models (27324 of the 37500 generated) that fit the data with a relative error level (equation (2.40)) less than 10 %. The best model (shown in red) has a relative error of 0.43 %. As it can be seen, although the models have an admissible error, all of them presents irregularities that are not geologically admissible in real environments. The best model also presents incompatible irregularities with the synthetic real model, although this inversion was performed with noise-free data.

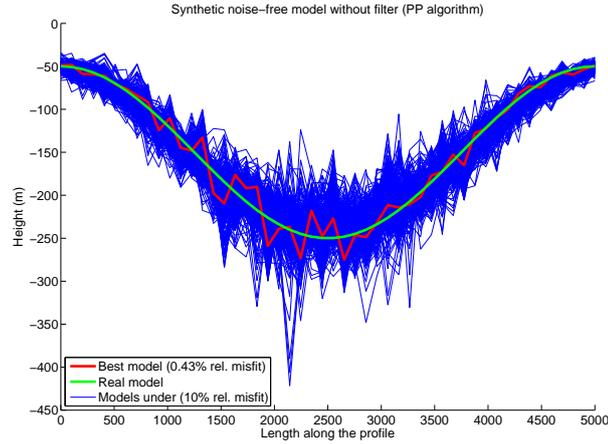


Figure 2.4: Results of the 2D gravity inversion without constraints presented in figure 2.3, using the PP-PSO algorithm. Green, real model; blue, models (27324) below 10 % relative error; red, best model (0.43 % relative error).

The results shown in figure 2.4 are very similar to those obtained by [Boschetti et al. \(1997\)](#), who pointed out the existence of exaggerated and arbitrary variations between contiguous and close rectangles, which produces unrealistic models from the geological point of view. This effect was also highlighted in a similar environment by [Parker \(1999\)](#), saying that *no smoothness constraint or regularization of any type were applied to the objective function, and the inversion appears somewhat unstable in regions with sparse data coverage* (the author does not present any figure nor goes into more details). While [Boschetti et al. \(1997\)](#) have used all the models below an error limit for the optimum model computation and uncertainty assessment, in this PhD Thesis the use of constraints in order to mitigate the high variability is proposed.

Firstly, *absolute constraints* ([Barbosa et al., 1997](#)) are directly applicable in PSO. If the basement depth is known at one or more points, the affected rectangles or prisms are identified and their search limits values in \mathbf{z}_b^{\min} and \mathbf{z}_b^{\max} can be modified according to the a priori data (fixing both limits to the same value of leaving a small variation range). The number of prisms to modify for each position with known depth will depend on factors such as the data point centering

level in the rectangle or prism, or other kind of information that suggests the application of the same depth in the near area.

Table 2.2: 5×5 filtering window example for 3D problems, whose coefficients are based on a bidimensional normal distribution with standard deviation $\sigma = 2$.

$$\begin{bmatrix} 0.7669 & 0.8471 & 0.8757 & 0.8471 & 0.7669 \\ 0.8471 & 0.9358 & 0.9673 & 0.9358 & 0.8471 \\ 0.8757 & 0.9673 & 1.0000 & 0.9673 & 0.8757 \\ 0.8471 & 0.9358 & 0.9673 & 0.9358 & 0.8471 \\ 0.7669 & 0.8471 & 0.8757 & 0.8471 & 0.7669 \end{bmatrix}$$

Secondly, the application of *relative constraints* is proposed in a different way as suggested by Barbosa et al. (1997). Relative constraints, as it was shown in section 1.4.3 for local optimization methods, introduces a smoothing effect in the solution, whose degree can be controlled at local scale via the values stored in the rows of the matrix \mathbf{R} , or at global scale with the μ^r parameter (equations (1.22)) and (1.23). In this PhD Thesis, the use of an average moving filtering is proposed as relative constraints application in both 2D and 3D. The filter is applied to all generated models in the PSO process which, once filtered, will be evaluated by equations (2.39) and (2.40). In 2D cases the filtering window will be a vector storing the appropriate coefficients, while in 3D cases it will be a square matrix (it could be also rectangular). In article 5, the weighted average filters are successfully tested, whose coefficients come from bidimensional normal distributions, which give more importance to central prisms and provide more flexibility when irregular sediments-basement surfaces are present. Table 2.2 shows a 5×5 filtering window example for weighted average, based on a bidimensional normal distribution of standard deviation $\sigma = 2$.

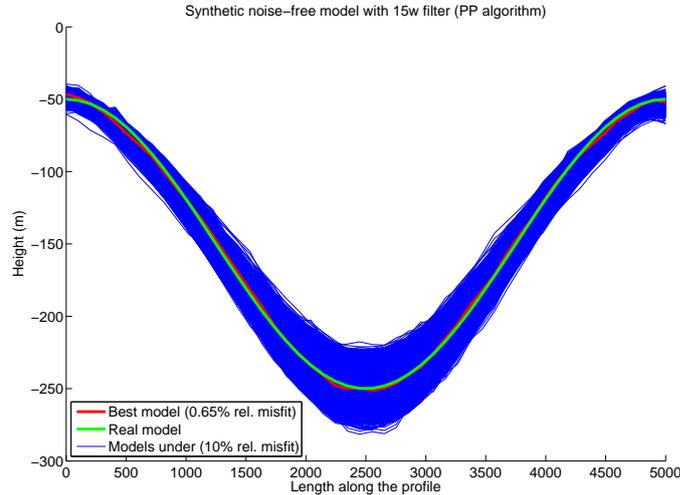


Figure 2.5: Results of the 2D gravity inversion with relative constraints presented in figure 2.3, using the PP-PSO algorithm. Green, real model; blue, models (24393) below 10 % relative error; red, best model (0.65 % relative error).

Figure 2.5 shows the inversion results of the model presented in figure 2.3 using a filtering windows with 15 elements. The smooth shape of the models, similar to the real model, can also be shown. Filtering window size and its numerical values are parameters whose evaluation

is part of the interpreter decision, which should be based on prior knowledge and external information about the sediments-basement general behavior. Small size windows will only take into account the immediate area to working prisms and generate models with more freedom to vary, which may be helpful if the sediments-basement interface presents important irregularities. As negative effect of small windows, irregularities as shown in figure 2.4 can appear. Otherwise, a similar effect can be obtained with bigger windows whose coefficients decrease in value as they move away from the center position. The optimum size of the filtering window could also be considered as an unknown adding its search limits to the PSO configuration.

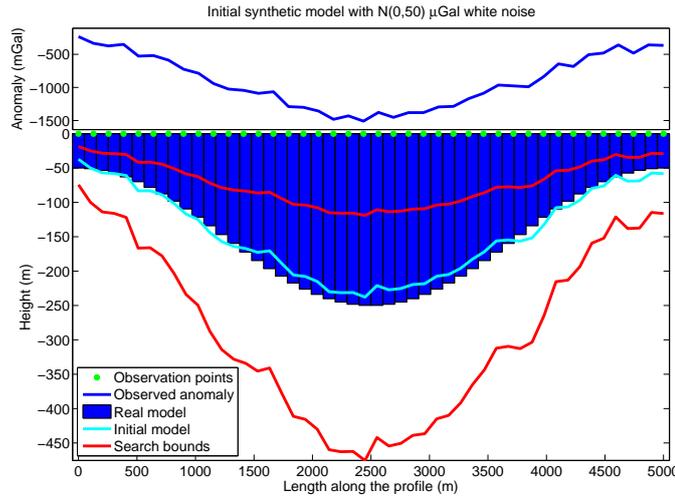


Figure 2.6: Search space limits (red lines) for a synthetic 2D inversion problem with white noise. The parameters are the same as for figure 2.3, plus the addition of $N(0, 50)$ μGal white noise.

Figure 2.6 shows the search limits for the example shown in figure 2.3, where white noise of distribution $N(0, 50)$ μGal has been added to the observed data, producing a relative error of 5.81% for the real model. The search limits have been computed as explained before, and the PSO parameters are: PP-PSO algorithm, swarm of 250 individuals, 150 iterations and filtering windows of size 15 and 1.0 values in all its positions. Figure 2.7 (left) shows all the models (as blue lines, 22612 from the total 37500 generated) which fit the data with a relative error below 10% for the noisy problem presented in figure 2.6. The best model has a relative error of 4.96% and, as in the noise-free case, filtering has excluded high variability models without geological meaning. In article 3, it has been demonstrated that noise in nonlinear problems has two main effects: the *displacement* of the solution in comparison to the true solution, and it makes the equivalence regions of low misfit to decrease in size. Figure 2.7 (right) shows the cost function topography for the two principal components, PCA (see article 4 for computation details), for the noise-free (upper) and noisy (lower) problems, whose solution has been shown in figures 2.5 and 2.7 (left). The described effects can be shown. Firstly, the cost function's minimum corresponding to the noisy problem does not coincide with the noise-free problem one, confirming this theoretical result. Also, the equivalent region decreasing is confirmed (in the noisy example the 5% relative error region has almost disappeared), which makes it difficult to detect by global search methods.

An important aspect to control in PSO is the swarm dispersion evolution along the iterations. The different PSO families are designed to collapse around their oscillation center (a dispersion below 5% of the initial one can be interpreted as the swarm has converged in a single point),

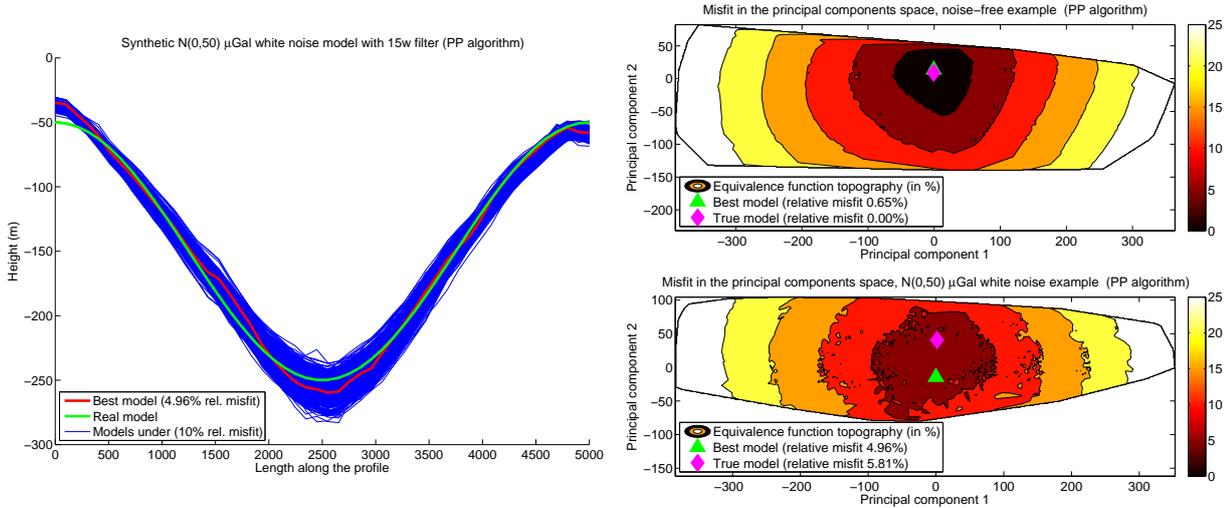


Figure 2.7: Left: 2D inversion results for the model presented in figure 2.6 using PP-PSO and filtering window of size 15. Green, real model; blue, models (22612) below 10 % relative error; red, best model (4.96 % relative error). Right: synthetic examples cost function topography for the first two principal components. Upper, noise-free example presented in figure 2.5; lower, noisy example presented in figure 2.7 (left).

but it is important for this collapse not to take place too early so that the search space can be properly explored. Figure 2.8 (right) shows the swarm dispersion along iterations for the problem stated in figure 2.6, for the GPSO, CC-PSO, CP-PSO, RR-PSO and PP-PSO family members. It can be seen in all cases how the swarm (the best solution in each iteration) keeps a dispersion value above 5% (except in exceptional cases in the CP-PSO member), which expresses a good exploratory behavior. GPSO, RR-PSO and PP-PSO members show in the bidimensional case a relatively high dispersion, which implies a good sampling of the search space. On the other hand, CC-PSO and CP-PSO members do not exhibit such a strong exploratory character, but they are above the 5% relative dispersion. Finally, figure 2.8 (left) shows the best model convergence evolution along iterations. All cases, except CC-PSO, show error decreasing with time which, in addition to the explorative character of the members, points out a good balance between the space sampling and the solution optimization.

2.2.4 Inverse problem solution and uncertainty assessment

Inverse problem solution does not only consist of providing the model that minimizes the functional (2.39), but it must also include its uncertainty assessment. While in linear inverse problems this task is not a difficult problem (see Aster et al. (2012), for example), the case of nonlinear problems is very different. As it was demonstrated in article 3, the application of linear techniques (after linearization of the original problem around an initial model) to nonlinear problems presents two main problems:

1. If the initial model is not good enough, local optimization methods can be trapped in local minima instead of in the cost function global minimum, providing a solution far from the real optimum.

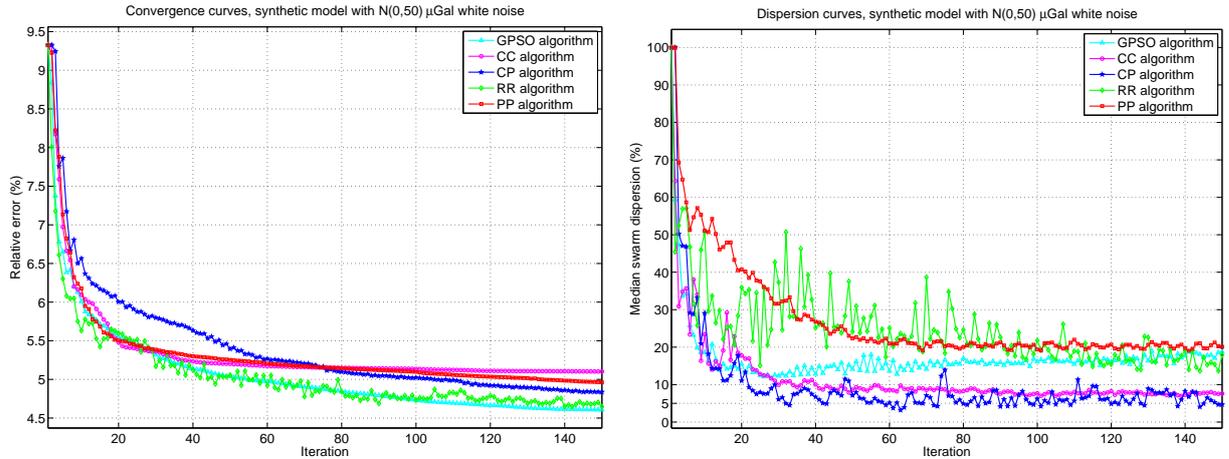


Figure 2.8: Convergence (left) and dispersion (right) curves corresponding to the solution of the problem presented in figure 2.6 with different PSO family members.

2. The equivalence region for nonlinear problems has a bended shape (Fernández-Martínez et al., 2012a), so its local approximation via linear techniques, whose equivalence regions are hyper-ellipsoids, involves an erroneous modeling, producing wrong uncertainty assessment.

In addition to the deterministic techniques, the Bayesian approach provides a rigorous way to determine the uncertainty assessment for the adopted solution. For a rigorous approximation using this way, a priori information is needed in the form of probability distribution of the observations and the model itself, which are combined in order to obtain the best a posteriori best model probability distribution (Tarantola and Valette, 1982a,b; Tarantola, 2005). Nevertheless, reliable a priori probability distribution for the model is in most cases impossible to obtain (Scales and Snieder, 1997). But this approximation does not only consist of applying the Bayesian criteria, but it must also include the sampling of all the low error equivalence regions, where the models fit the data below a given tolerance (see article 1). Usually, the sampling is carried out via Monte Carlo methods (Mosegaard and Tarantola, 1995; Gouveia and Scales, 1998) and without considering the structure of the cost function topography, which lack efficiency in high dimensional problems, an even make the task unfeasible (Scales and Tenorio, 2001). In this PhD Thesis, uncertainty assessment will be carried out using the generated models in the PSO process.

According to Scales and Snieder (1997, 2000); Tarantola (2006), a simple way to give an inverse problem solution consist of providing the set of models that fit the data below a given tolerance. In this way, the sets represented in figures 2.5 and 2.7 (left) would be solutions for the problems stated in figures 2.3 and 2.6, respectively. In this PhD Thesis, for both for 2D and 3D problems, the uncertainty region for a given relative error level (10%, for example) is computed as:

1. From all generated models, all the ones that have an error close to the working one, for example in the range 9.9% – 10.2%, are selected. This procedure is used because it is difficult that models with exactly a 10% error level have been generated. In figure 2.9 (left) the selected models, which come from the collection shown in figure 2.7 (left), can be observed. As it can be seen, there is a set of models above and below the *best* solution.

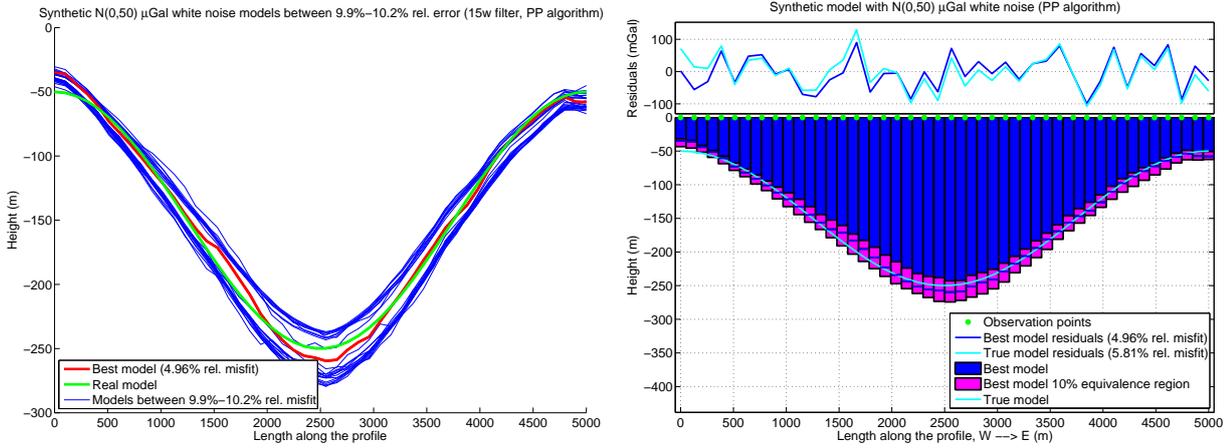


Figure 2.9: Noisy 2D inverse problem solution using relative constraints and PP-PSO algorithm, corresponding to the problem presented in figure 2.6. Left: models inside the 9.9%–10.2% relative error region. Right: final solution (best model and 10% equivalence region).

2. The average model of each set is computed, being the results the lower and upper limits for the equivalence region. Figure 2.9 (right) shows the solution with the 10% equivalence region for the models shown in figure 2.7 (left). It can be seen how the real model is not inside the 10% equivalence region in the profile left extreme, which is due to side effects in filtering. This method highlights the importance of working with a PSO member with good explorative behavior. Swarm dispersion is not only important for the global optimum detection, but also for a correct depiction of the a posteriori cost function equivalence regions.

Finally, the solution of the nonlinear 2D (for 3D the process is analogous) gravity inversion problem in sedimentary basins, i. e., the determination of the sediments-basement interface is presented in figure 2.9 (right). Although the *best* model is highlighted, the correct way to express the depth of any rectangle consist of giving its upper and lower limits to the working relative error, as any model between them will fit the data below the specified tolerance and, in absence of other criteria, it will be as valid as the above mentioned *best* model.

Articles published in peer-reviewed journals included in the *Journal Citation Reports*

Introduction

In this chapter, the articles published in peer-reviewed journals included in the *Journal Citation Reports* (JCR) during the PhD Thesis development are presented. Such works meet the requirements of the doctoral program for the modality of Thesis as Compendium of Publications.

All the articles were published in the *Journal of Applied Geophysics*, whose quality indicators, that can be seen in detailed form in appendix A, are the following:

- Categories in JCR: *Mining & mineral processing* and *Geosciences, multidisciplinary*.
- Impact factors: 1.301 (2013), 1.500 (2014), and 1.355 (2015).
- Positions in categories:
 - *Mining & mineral processing*: 7/21, Q2 (2013), 6/20, Q2 (2014), and 10/21, Q2 (2015).
 - *Geosciences, multidisciplinary*: 106/174, Q3 (2013), 96/175, Q3 (2014), and 116/184, Q3 (2015).

Due to the topics covered in the published works and the JCR categories in which the *Journal of Applied Geophysics* is included, we consider it as an adequate journal for presenting the results of the PhD Thesis to the scientific community.

ARTICLE 1

**From Bayes to Tarantola: New
insights to understand uncertainty in
inverse problems**

Authors:

Juan Luis Fernández Martínez
María Zulima Fernández Muñiz
José Luis García Pallero
Luis Mariano Pedruelo González

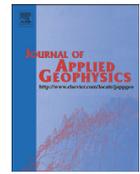
JOURNAL OF APPLIED GEOPHYSICS
2013, VOL. 98, PP. 62–72
[doi:10.1016/j.jappgeo.2013.07.005](https://doi.org/10.1016/j.jappgeo.2013.07.005)

Summary

In this first article, the importance of the uncertainty analysis in inverse problems, specially in Geophysics, is stressed. Three common problems, which are usually accepted by practitioners, are highlighted:

1. The uncertainty of an inverse problem solution has a random structure which is impossible to know.
2. Regularization techniques make the equivalent solutions to disappear.
3. Linearization techniques show accurately the uncertainty region for the adopted solution.

All these suppositions are not correct and they can lead to errors in decision making. In this work, new ways to understand the concept of uncertainty from a deterministic point of view are introduced, and a «geometric» definition of the ill-conditioned concept is proposed. Local optimization techniques are not capable to perform a correct uncertainty analysis for an inverse problem solution due to when the nonlinear equivalence region is approximated by the linearized hyper-quadric, the structure of the equivalent solutions is only locally depicted. On the other hand, Bayesian approaches and global optimization techniques are inefficient because they ignore the algebraic (or geometric) structure of the equivalence regions and are highly inefficient in high dimensional inverse problems. The understanding of the uncertainty from a deterministic point of view is a necessary step to design more efficient methods to sample the uncertainty space in high dimensional problems.



From Bayes to Tarantola: New insights to understand uncertainty in inverse problems[☆]



J.L. Fernández-Martínez^{a,*}, Z. Fernández-Muñiz^a, J.L.G. Pallero^b, L.M. Pedruelo-González^a

^a Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain

^b ETSI en Topografía, Geodésia y Cartografía, Universidad Politécnica de Madrid, Madrid, Spain

ARTICLE INFO

Article history:

Received 20 March 2013

Accepted 27 July 2013

Available online 6 August 2013

Keywords:

Inverse problems
Linear and nonlinear inversion
Uncertainty analysis
Decision making

ABSTRACT

Anyone working on inverse problems is aware of their ill-posed character. In the case of inverse problems, this concept (ill-posed) proposed by J. Hadamard in 1902, admits revision since it is somehow related to their ill-conditioning and the use of local optimization methods to find their solution. A more general and interesting approach regarding risk analysis and epistemological decision making would consist in analyzing the existence of families of equivalent model parameters that are compatible with the prior information and predict the observed data within the same error bounds. Otherwise said, the ill-posed character of discrete inverse problems (ill-conditioning) originates that their solution is uncertain. Traditionally nonlinear inverse problems in discrete form have been solved via local optimization methods with regularization, but linear analysis techniques failed to account for the uncertainty in the solution that it is adopted. As a result of this fact uncertainty analysis in nonlinear inverse problems has been approached in a probabilistic framework (Bayesian approach), but these methods are hindered by the curse of dimensionality and by the high computational cost needed to solve the corresponding forward problems. Global optimization techniques are very attractive, but most of the times are heuristic and have the same limitations than Monte Carlo methods. New research is needed to provide uncertainty estimates, especially in the case of high dimensional nonlinear inverse problems with very costly forward problems. After the discredit of deterministic methods and some initial years of Bayesian fever, now the pendulum seems to return back, because practitioners are aware that the uncertainty analysis in high dimensional nonlinear inverse problems cannot (and should not be) solved via random sampling methodologies. The main reason is that the uncertainty "space" of nonlinear inverse problems has a mathematical structure that is embedded in the forward physics and also in the observed data. Thus, problems with structure should be approached via linear algebra and optimization techniques. This paper provides new insights to understand uncertainty from a deterministic point of view, which is a necessary step to design more efficient methods to sample the uncertainty region(s) of equivalent solutions.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction: from J. Hadamard to A. Tarantola, several milestones in our story

Most of the inverse problems in geosciences can be written in discrete form as follows:

$$\mathbf{F}(\mathbf{m}) \approx \mathbf{d}. \quad (1)$$

In this relationship $\mathbf{m} = (m_1, m_2, \dots, m_n) \in M \subset \mathbb{R}^n$ is the estimated model (or model parameters) that belongs to a set of admissible models M defined in terms of some prior geological knowledge, $\mathbf{d} \in \mathbb{R}^s$ are the

observed data, and $\mathbf{F}(\mathbf{m}) = (f_1(\mathbf{m}), f_2(\mathbf{m}), \dots, f_s(\mathbf{m}))$ is the vector field representing the forward model, being $f_j(\mathbf{m})$ the j -scalar field component function of \mathbf{F} , accounting for the j -th data.

The symbol \approx in Eq. (1) means that this relation might not be exact (in which case the symbol $=$ would be the correct one), and the intersection problem stated in Eq. (1), that consists in finding the model or models \mathbf{m} whose predictions $\mathbf{F}(\mathbf{m})$ accurately match the observed data \mathbf{d} , might not have a solution. This fact is related to the ill-posed character of inverse problems, that is, either the inverse problem does not admit solution, either the solution exists and it is not unique, or finally the solution exists and is unique but it is unstable, that is, the solution does not depend continuously on the observed data. This last situation is usually referred as ill-conditioning of the inverse problem, and it is at this point where our story begins.

By writing this paper we try to make practitioners aware that, as Tarantola (2006) wrote, the idea of solving an inverse problem by obtaining only one 'best' model requires revision. It is important to

[☆] This paper is dedicated to the memory of Albert Tarantola.

* Corresponding author. Tel.: +34 985103199.

E-mail address: jlfm@uniovi.es (J.L. Fernández-Martínez).

understand that due to the effect of noise in data, we can affirm that the global optimum of the inverse problem cost function is the only model that has not generated by the observed data.

Although Albert Tarantola was the founder together with Bernard Valette of the probabilistic approach to inverse problems (Tarantola and Valette, 1982a,b), Tarantola (2006) finally proposed to use Karl Popper's logic of the scientific discovery (Popper, 1959) to approach and solve inverse problems. Interestingly, he compared solving inverse problems in high dimensions to finding a needle in a haystack, and proposed to use all available a priori information to sequentially create models ("potentially an infinite number of them"), to evaluate the forward problem for this ensemble, and finally to compare the predictions to the actual observations and use some criterion to decide if the fit is acceptable or unacceptable, given the uncertainties in the observations and physical constraints. The unacceptable models must be dropped. Conversely, collections of all models that have not been "falsified" represent the inverse problem solution. Although this methodology is what any modeler would like to accomplish, this "brute force approach" is limited by the inverse problem dimensionality (Curtis and Lomax, 2001; Tarantola, 2005), and by the computer requirements needed to evaluate a huge amount of forward solves in case of very costly nonlinear physics.

Unfortunately, uncertainty quantification in most of the publications dealing with inverse problems in geophysics (even in low dimensions) is not correctly performed, since uncertainty in the solution is usually badly understood. Modelers tend to adopt a non-robust (good or bad) minimum misfit solution, instead of trying to adopt a decision with its corresponding uncertainty quantification. This might be due to the fact that modelers are usually experts in their respective forward problems but not really in inverse problem theory. Also, uncertainty analysis in inverse problems is still a changing and dynamic subject of research where very different methodologies are proposed, and no clear indications are given to show how and why uncertainty appraisal should be performed, especially in high dimensional problems.

In our opinion nowadays no serious research involving inverse problems should be published without performing proper uncertainty quantification. The word proper does not mean in this case to provide the whole parameter posterior distribution, but a conditional posterior, that is, the region (or regions) of model parameters that fit the data within the same error bounds and are compatible with the prior information at disposal. These models will be called equivalent, and the region(s) of the model space containing these models will be called the region(s) of equivalence (linear, linearized or nonlinear). In the extreme case of very high dimensional inverse problems with very costly forward problems uncertainty quantification should just be able to unravel the existence of very different plausible scenarios.

1.1. Hadamard's ill-posed problems

The concept of ill-posed problems was first proposed by Jacques Hadamard (1902). He stated that only well-posed problems do correspond to physical phenomena. Obviously J. Hadamard was referring to physical models that in our case are the forward mathematical models used to provide the data predictions. He certainly could not imagine at that time the scientific and technological importance that inverse problems will have nowadays. Although the forward problems used in inverse methodologies are well-posed, the concept of ill-posed problem was inherited to the case of inverse problems.

As will be shown, one of the main causes of ill-posedness in discrete inverse problems is ill-conditioning that is related to the lack of continuity of the inverse operator over the entire data space. Instability is intimately related to the fact that inverse problems are solved as optimization problems and the small singular values of the linearized forward operator amplify (through the generalized inverse operator) the noise in data back into the model parameters providing different equivalent solutions. These solutions eventually might be spurious if some physical

constraints are violated, for instance, the model parameters should be positive and the inversion provides negative values. An excellent introduction about the inverse problems particularities is provided by Parker (1977).

The analysis of linear inverse problems serves to illustrate and understand the concept of ill-posedness. The first cause of ill-posedness in linear inverse problems is non-existence of the solution (incompatibility), that is, the observed data \mathbf{d} do not belong to the column space of the forward operator,

$$\text{Im}(\mathbf{F}) = \{\mathbf{y} \in \mathbb{R}^s : \mathbf{y} = \mathbf{F}\mathbf{x}, \mathbf{x} \in \mathbb{R}^n\}.$$

These incompatible linear problems are solved via generalized least-squares methods.

For compatible linear inverse problems, the second cause of ill-posedness is non-uniqueness which is related to the lack of injectivity of the forward linear operator \mathbf{F} . In this case \mathbf{F} has a non-trivial null space (or kernel), $\ker(\mathbf{F}) = \{\mathbf{m} \in M : \mathbf{F}\mathbf{m} = \mathbf{0}\}$, which is the linear subspace of the model space containing the models that have not influence on the data predictions. In the compatible and underdetermined case there exist an infinite number of solutions of the type:

$$\mathbf{m}_{sol} = \mathbf{m}_p + \mathbf{w}, \mathbf{w} \in \ker(\mathbf{F}),$$

where \mathbf{m}_p is any particular solution of the linear inverse problem.

Finally, in the case where $\ker(\mathbf{F})$ is trivial and the solution exists and is unique, ill-conditioning refers to the fact that finding the exact solution becomes very inaccurate. To explore the reason of this numerical inaccuracy, let us define the "numerical kernel" of \mathbf{F} as the linear subspace of the model space formed by the right singular vectors of \mathbf{F} associated to the singular values that are almost zero. Intuitively the numerical kernel of \mathbf{F} is formed by the models that have "little" influence on the data predictions. Otherwise said, although the inverse operator is continuous, in practice it is almost "discontinuous". To prove that, we only need to understand that in the case of linear discrete inverse problems, the continuity of any linear operator $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^s$ is a consequence of its bounded character, that is:

$$\exists K > 0 : \|\mathbf{F}\mathbf{x}\|_2 \leq K \|\mathbf{x}\|_2. \quad (2)$$

Thus, all the linear operators between finite dimensional spaces are continuous. Then, to define the inverse operator \mathbf{F}^{-1} , let us first restrict the linear application \mathbf{F} between \mathbb{R}^n and the column space of \mathbf{F} ($\text{Im}(\mathbf{F})$). Then the condition

$$\dim(\text{Im}(\mathbf{F})) = n \Leftrightarrow \dim(\text{Ker}(\mathbf{F})) = 0,$$

is needed for \mathbf{F}^{-1} to exist, and therefore being continuous.

In conclusion, the inverse operator \mathbf{F}^{-1} is continuous if and only if the forward linear operator \mathbf{F} is an injection (or a bijection if \mathbf{F} is restricted to $\text{Im}(\mathbf{F})$), and the null space (or kernel) of \mathbf{F} is trivial. Finally, the ill-conditioning in the least squares sense (for purely over-determined linear systems with no solution) is also related to accurately finding the minimum misfit solution of the corresponding least squares problem $\min_{\mathbf{m} \in \mathbb{R}^n} \|\mathbf{F}\mathbf{m} - \mathbf{d}\|_2^2$. In this case the matrix that is involved is $\mathbf{F}^T\mathbf{F}$ whose kernel coincides with the kernel of \mathbf{F} . In conclusion the existence of a kernel (numerical or not) is the cause for the lack of uniqueness and stability.

Apparently this definition seems not to be related to the existence of other equivalent solutions, those that provide similar data misfits. Nevertheless, the lack of injectivity of the forward operator \mathbf{F} provokes: 1) the existence of a family of equivalent solutions, and 2) the determination of the minimum misfit solution to be numerically inaccurate. The concept of ill-conditioning in linear inverse problems will be revised in Section 2.

Interesting to remark that sampling the family of equivalent solutions is not affected by numerical instabilities introduced by the optimization procedure, since the sampling only uses the forward operator \mathbf{F} which corresponds to a well-posed problem.

1.2. Backus–Gilbert theory and Tikhonov and Arsenin's regularization

Once the ill-posed character of inverse problems was admitted the question resided in finding stable solutions. The second milestone in this story is due to Backus and Gilbert (1970, 1968, 1967) and Tikhonov and Arsenin (1977).

In the field of continuous linear inverse problems, the work of Backus and Gilbert was of paramount importance to understand the mathematical structure of inverse problems, introducing concepts such as model resolution and showing the existing trade-off between resolution and error, which is an essential ingredient in inverse theory, regardless of the inversion method that is used (Pujol, 2013).

Backus and Gilbert method tried to construct the mapping between the solution of the inverse problem and the true solution as close to the identity as possible. For that purpose they introduced the resolution or averaging kernel and tried to minimize its spread, that is, to maximize the resolving power of the technique. Due to this fact Backus and Gilbert method is also called Optimally Localized Average (OLA) method. Instead of imposing smoothness constraints on the solution such as Tikhonov's regularization the OLA method imposes stability constraints, so that the solution would vary as little as possible as a function of the observed data noise. Details can be consulted in Pujol (2013). Aside from its historical interest, Backus and Gilbert formulation of the inverse problem constitutes a regularization technique that provides similar results to the Tikhonov and Arsenin's regularization method since stable solutions are smooth. In general the computation burden makes the Backus and Gilbert method unsuitable for other than one dimensional inverse problems. Recently Snieder (1991) has extended the Backus–Gilbert theory to nonlinear inverse problems.

The second interesting achievement was done by Tikhonov and Arsenin (1977), who introduced the concept of conditionally well-posed inverse problem that does not require solvability over the entire model space. They also introduced the correctness set where the inverse operator becomes continuous and the inverse problem conditionally well-posed. They proved that if the correctness set is compact (closed and bounded), it is possible to achieve a stable pseudo-solution (or quasi-solution) of the inverse problems by minimization of the misfit functional in some appropriate class of suitable models belonging to the correctness set. Nevertheless, in many situations it is difficult to describe the correctness set. To overcome this problem they introduced the regularization methods to approximate the original ill-posed inverse problem as a family of well-posed regularized problems, $\mathbf{m}_\varepsilon = \mathbf{F}_\varepsilon^{-1}(\mathbf{d})$, where $\varepsilon > 0$ is a regularization parameter. Obviously, in this methodology it is expected that the regularized solution will approach the true solution as the regularization parameter tends to zero, $\mathbf{m}_\varepsilon \rightarrow_{\varepsilon \rightarrow 0} \mathbf{m}^{true}$. This is not always the case in real practice. Also, in real inverse problems there is not a way to know the true model that has generated the data, if this model really exists, or it is a mathematical abstraction used to explain the observed data.

Although other regularization techniques exist, Tikhonov and Arsenin have also shown that the simplest way to stabilize the inversion was by introducing in the misfit function a penalization term that takes into account the distance to a prior reference model, \mathbf{m}_{ref} :

$$\min_{\mathbf{m} \in \mathbb{R}^n} \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}_{ref}\|_2^2, \quad (3)$$

the so-called zero order regularization. In statistics this method is called ridge regression. Tikhonov's regularization introduces a

Pareto's front, typical from multi-criteria optimization: to minimize the prediction error, $\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2$, and at the same time the distance to the prior reference model, $\|\mathbf{m} - \mathbf{m}_{ref}\|_2^2$.

The main role of the regularization is to select from the set of possible solutions those that depend continuously on the observed data and possess a specific property that depends on the regularization that has been chosen.

The effect of the regularization techniques in linear inverse problems will be analyzed in detail in Section 2.3. This analysis is also valid in the case of the linearized version of the nonlinear inverse problems.

This methodology has been applied for years, but modelers begin to realize that other equivalent solutions exist and the uncertainty on the regularized solution was not analyzed. Otherwise said, the fact that the solution has been stabilized through the Tikhonov's regularization, unfortunately it does not provoke the disappearance of the uncertainty in the solution. Similar considerations can be done for other kind of regularization approaches. Also, as we will show, in the case of nonlinear inverse problems, depending on how the reference model is chosen, it is not guaranteed that the regularized and the true solution (that has generated the observed data) will be located on the same valley of the cost function landscape, since the region of equivalence can be composed of several disconnected valleys of low misfits (Fernández-Martínez et al., 2012b).

1.3. Bayes, Tarantola and the uncertainty analysis of inverse problems

The third milestone concerns the uncertainty analysis of the solution in order to quantify risk and adopting probabilistic decision making approaches. In an optimization framework, uncertainty estimation in inverse problems involves finding the family M of models \mathbf{m} that fit the observed data $\mathbf{d} \in \mathbb{R}^s$ (comprising all the observables) within the same tolerance (tol) and are consistent with our prior knowledge for a certain norm p :

$$\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_p < tol. \quad (4)$$

Eventually the regularized functional (3) can be also adopted instead of Eq. (4). It is important to note that the definition in Eq. (4) only implies the use of a norm in the data space, and eventually in the model space if a regularization term is used, that is, the data and model spaces should be at least metric spaces. Furthermore if the Euclidean norm is used, the data space is also an Euclidean space with a dot product and orthogonal projections can be additionally performed, as in the least squares solution. Particularly no probabilistic interpretation of the uncertainty is given in Eq. (4). Nevertheless, as will be shown in the next section, both approaches (probabilistic and deterministic) are complementary.

The region (or regions) containing the models fulfilling condition (4) for different error tolerances, will be named in this paper as uncertainty "space" of the inverse problem $\mathbf{F}(\mathbf{m}) = \mathbf{d}$. In this case the word "space" has to be correctly understood as the region or regions containing the equivalent models, because this set of models is not a linear vector subspace of the model space, since

$$\forall \alpha, \beta \in \mathbb{R}, \forall \mathbf{m}_1, \mathbf{m}_2 : \begin{cases} \|\mathbf{F}(\mathbf{m}_1) - \mathbf{d}\|_p \leq tol, \\ \|\mathbf{F}(\mathbf{m}_2) - \mathbf{d}\|_p \leq tol, \end{cases} \quad (5)$$

does not imply

$$\|\mathbf{F}(\alpha \mathbf{m}_1 + \beta \mathbf{m}_2) - \mathbf{d}\|_p \leq tol. \quad (6)$$

There are many reasons for uncertainty in inversion, the most important of which are:

1. Measurement errors and noise in data that shift the solution of the inverse problem deforming the topography of the cost function

(Fernández-Martínez et al., 2012b). The solution with minimum misfit will never be the one that has generated the observed data. An example of this fact is shown in Section 3.2 for a simple two dimensional nonlinear regression problem.

2. Incomplete and discrete coverage of the data space that generates ambiguity and non-uniqueness in the model reconstruction. The data coverage is always discrete due to technical limitations and economic restrictions.
3. Physical assumptions and numerical approximations in the forward problem that originate inconsistencies and ill-conditioning in the inversion.
4. The presence of several scales of heterogeneity in the model space and the bandwidth limitations of the imaging techniques that are used. Typically fine scale heterogeneities are not informed by the observed data, and thus, belong to the uncertainty space of the inverse problem.

For instance, in geophysics, the sensitivity of a geophysical method can be defined as the ratio of the norm of the perturbation of the observed data to the norm of the perturbation of the model parameters. It can be proved that the sensitivity of a method is the Euclidean norm of the Jacobian (see for instance Zhdanov, 2002):

$$S(\mathbf{m}_0) = \sup \left\{ \frac{\|\Delta \mathbf{d}\|_2}{\|\Delta \mathbf{m}\|_2} \right\} = \|\mathbf{J}\mathbf{F}_{\mathbf{m}_0}\|_2, \quad (7)$$

where $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}$ is the Jacobian of the forward operator \mathbf{F} calculated in the model \mathbf{m}_0 . The geophysical model will be only sensitive to those perturbations of the model parameters that exceed the threshold $\frac{\delta d}{S(\mathbf{m}_0)}$, where δd is the level of noise in the observations. Any other variations of the model cannot be inferred from the data.

In an analogous way the resolution of the geophysical method is related to the norm of the inverse of the Jacobian operator $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^\dagger$ (Zhdanov, 2002):

$$R(\mathbf{m}_0) = \frac{1}{\|\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^\dagger\|_2}.$$

Two models in the neighborhood of \mathbf{m}_0 can be resolved if:

$$\|\Delta \mathbf{m}\|_2 \geq \frac{\delta d}{R(\mathbf{m}_0)} = \delta d \cdot \|\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^\dagger\|_2. \quad (8)$$

Then, if the inverse operator $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^\dagger$ is not bounded, the resolution of the geophysical method goes to zero. This is the case in linear inverse problems for models that lie in the direction of the kernel of the forward operator \mathbf{F} , if the forward operator \mathbf{F} is not injective (Fernández-Martínez et al., 2012b). Also, in the case of nonlinear problems $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^\dagger$ might not be bounded, nevertheless the nonlinear inverse problem might be “better posed” than the corresponding linearized inverse problem. Geometrically this corresponds to the fact that the linearized equivalence region around \mathbf{m}_0 is an elliptical cylinder, while the nonlinear equivalent curvilinear region has a limited extent.

The uncertainty analysis in geosciences is closely related to the work of Bayes (1763) and Tarantola and Valette (1982a,b). The fact that deterministic methods failed to account for the uncertainty in nonlinear inverse problems (they were not designed to accomplish this task) was the impetus for the development of probabilistic approaches to solve inverse problems influenced by Bayes theory. Furthermore, the problem of uncertainty has a natural and beautiful interpretation in a Bayesian framework. Bayes' rule allows one to factorize the conditional probability of a hypothesis H after a certain evidence E is observed (or posterior probability of H given E) in terms of the prior probability of H , the prior probability of E , and the conditional probability of E given H . In the context of nonlinear inversion, Bayes' rule states that a set of model

parameters is more probable if it explains the observed data with a higher probability:

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}. \quad (9)$$

The term $P(\mathbf{d}|\mathbf{m})$ is called the likelihood and typically depends exponentially of the data misfit $\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_p$ in a certain norm p (this assumes that the data misfit has a probability distribution in the family of Gaussians).

The interpretation of inverse problems in probabilistic terms has been initially proposed by Tarantola and Valette (1982a,b). This preliminary work introduced the probabilistic approach to inverse problems and established a very useful link to Bayes's theory. Nevertheless, the historic origin of Bayes' rule should not be forgotten (Bayes, 1763): “*An Essay towards solving a Problem in the Doctrine of Chance*”. This fact does not strictly apply here, because the inverse problem is based on physical laws, denoted by \mathbf{F} . Thus, the uncertainty of the inverse problem has to be somehow related to \mathbf{F} , \mathbf{d} , and the space of priors (the correctness set in the terminology of Tikhonov and Arsenin). Bayes's rule clearly explains that the posterior distribution of the model parameters depends on the prior (or reference model), and also on the data (noise included) through the likelihood term. Adopting the evidence as a normalization constant the posterior distribution could be written in the family exponentials as:

$$\begin{aligned} P(\mathbf{m}|\mathbf{d}) &= ke^{-\left(\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 + \epsilon^2 \|\mathbf{m}-\mathbf{m}_{ref}\|_2^2\right)} \\ &= ke^{-\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2} e^{-\epsilon^2 \|\mathbf{m}-\mathbf{m}_{ref}\|_2^2}. \end{aligned} \quad (10)$$

Thus, the regularization term induces the model prior probability distribution.

Taking into account relationship (10) it is possible to conclude that in the case of linear problems and Gaussian or uniform priors imply Gaussian posteriors that correspond to the linear hyper-quadratic, and in the case of nonlinear inverse problems the posterior will be non-Gaussian (corresponding to the nonlinear equivalence region) independent of the type of prior that is adopted.

Eventually other p norms for the likelihood and also for the prior can be used, such as in Lasso (Tibshirani, 1996), and induced different probabilistic distributions, both, in the likelihood term and in the prior. In this case the posterior will be also non-Gaussian.

Bayesian approaches do not just consist in the application of the maximum likelihood criterion to find the mode of the posterior distribution, which in fact is equivalent to deterministic inversion, but also sampling all the regions of models of low misfit (high likelihood) that are in accord with the prior model distribution. In conclusion, posterior sampling and the analysis of the topography of the cost function landscape are intimately related.

Although this connection between the deterministic and the stochastic worlds existed, Bayesian approaches popularized the use of random sampling methodologies, such as, Monte Carlo methods (see for instance Mosegaard and Tarantola, 1995) to compute the inverse model uncertainties. This was maybe due to the fact that the topography of the cost function was not correctly understood and modelers thought that the uncertainty space had a random structure with local minima located anywhere in the model space.

It will be shown in Section 3.1 that local optimization methods in nonlinear inverse problems usually use the linearized Hessian which is a symmetric and semi-definite positive matrix, allowing for optimization. Therefore, these methods are unable to catch the real uncertainty in the nonlinear sense, because the topography of the cost function induced by the linearized Hessian corresponds to a straight flat elongated valley, while the real nonlinear cost function topography consists of one or more curvilinear valleys (or basins) of low misfits, eventually connected by saddle points (Fernández-Martínez et al., 2012b). Also,

as it has already been pointed out, the linearized region of equivalence in the neighborhood of a model \mathbf{m}_0 usually has locally a larger extent than the corresponding nonlinear region of equivalence. This situation is related to the ill-conditioning of the Jacobian of the forward operator calculated in \mathbf{m}_0 . This ill-conditioning of the linearized problem always occurs in the case of highly underdetermined linear systems because the rank of the Jacobian matrix is not higher than the dimension of the observed data vector which is much lower than the number of unknowns.

Due to these drawbacks, Monte Carlo methods use random sampling techniques and do not approach the inverse problem as an optimization problem. Nevertheless, random sampling methodologies depend too strongly on the dimension of the parameter space and require innumerable forward solves. The same situation occurs for global optimization methods which are not perfect importance samplers, that is, sampling is not performed according to the true posterior distribution, since the stochastic search is biased towards the low misfit regions by using some fitness or misfit criteria. Nevertheless these methods can provide under exploratory conditions a very good proxy of the model posterior distribution (see for instance Fernández Álvarez et al., 2008; Fernández-Martínez et al., 2010a,b; Sen and Stoffa, 1995). As it happens for the Monte Carlo methods, the global optimization depends also strongly on the dimension of the parameter space (Curtis and Lomax, 2001) and on the computational cost to solve the forward problem.

For medium or large sized parameterizations and fast forward solvers, the above mentioned methods can be used combined with model reduction techniques and subspace methods (see for instance Fernández-Martínez et al., 2012a). Finally, the third type of inverse problems includes those that are highly nonlinear, with a very high number of parameters (hundreds of thousands or even millions) and very costly forward evaluations. This kind of inverse problems is highly underdetermined due to the fine discretization in the model space needed to perform very accurate forward predictions. Some examples of this type are the seismic history matching problem at the field scale, the full wave inversion problem, or the controlled source electromagnetic inverse problem. In these cases, model reduction techniques can be used to take into account the correlation between parameters, but the use of Markov Chain Monte Carlo methods, global optimization techniques or other kind of random sampling and stochastic methodologies is ill-advised or even precluded due to the high computational cost needed to obtain the data predictions. These methods might benefit from the use of distributed computing facilities; nevertheless the question about the convenience of the stochastic character of the sampling procedure still remains.

2. Ill-conditioning revisited

In this section we introduce a more general definition of the ill-condition character of inverse problems that is related to the structure of the regions of equivalence. Otherwise said, the ill-conditioned character of discrete inverse problems does not only originate the numerical determination of the model of lower misfit to be unstable, more important, ill-conditioning is at the origin of the uncertainty problem. Thus, we show that uncertainty is inherent to the solution of linear and nonlinear inverse problems and cannot be neglected. The research work of Jackson (1972, 1979) is very important in this field.

2.1. Linear inverse problems: case of a full rank endomorphism

The simplest case corresponds to a linear inverse problem, $\mathbf{Fm} = \mathbf{d}$, where $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a full rank endomorphism (bijective linear application). Linear algebra states that this problem admits a unique solution, $\mathbf{m}_c = \mathbf{F}^{-1}\mathbf{d}$, where \mathbf{F}^{-1} is the inverse of matrix \mathbf{F} . In this case \mathbf{F}^{-1} exists and, thus, is continuous.

The region containing the models that fit the observed the data \mathbf{d} within a given tolerance tol , $\|\mathbf{Fm} - \mathbf{d}\|_2 \leq tol$, is a hyper-quadic centered at \mathbf{m}_0 , having its main axes oriented following the right singular vectors (\mathbf{V} base) of the forward operator \mathbf{F} , provided by its singular value decomposition. The semi axes of this linear hyper-quadic are the inverse of the singular values of \mathbf{F} , named μ_k .

To prove this fact, let us write $\mathbf{m} = \mathbf{m}_c + \Delta\mathbf{m}$, where $\mathbf{m}_c = \mathbf{F}^{-1}\mathbf{d}$, is the unique solution of the linear system. Then, the equivalent models with prediction error tolerance, tol , satisfy:

$$\frac{\|\mathbf{Fm} - \mathbf{d}\|_2}{\|\mathbf{d}\|_2} \leq tol \Rightarrow (\Delta\mathbf{m})^T \mathbf{F}^T \mathbf{F} (\Delta\mathbf{m}) \leq tol^2 \|\mathbf{d}\|_2^2. \quad (11)$$

Referring the model increment $\Delta\mathbf{m}$ to the \mathbf{V} base, we arrive at:

$$(\Delta\mathbf{m}_V)^T \Sigma^T \Sigma (\Delta\mathbf{m}_V) \leq tol^2 \|\mathbf{d}\|_2^2 \Leftrightarrow \sum_{k=1}^{n=\text{rank}(\mathbf{F})} \left(\frac{\Delta m_{V_k}}{1/\mu_k} \right)^2 = tol^2 \|\mathbf{d}\|_2^2. \quad (12)$$

Fig. 1A shows an example in two dimensions for the linear operator $\mathbf{F} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. In this case the hyper quadric is a conic (ellipse). The longer semi axis is oriented in the direction of the \mathbf{V} base corresponding to

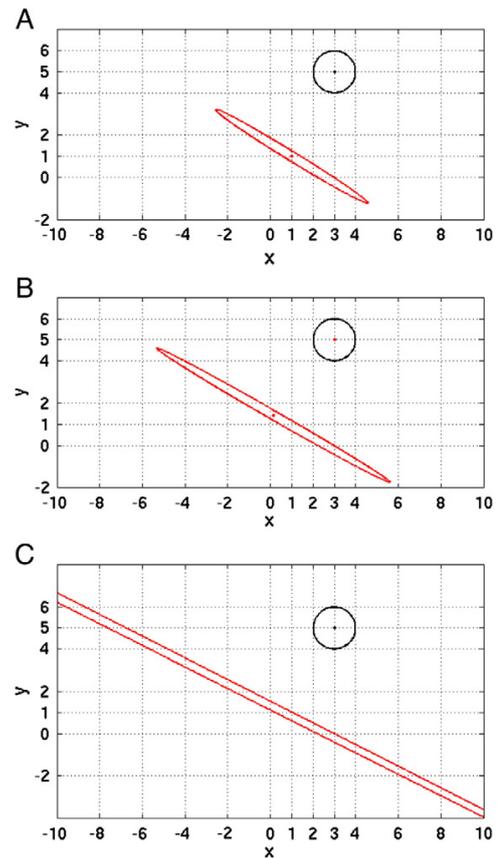


Fig. 1. (A) Linear equivalent region for $\mathbf{F} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$, $tol = 1$. The center of the hyper-quadic is $\mathbf{m}_c = \mathbf{F}^{-1}\mathbf{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. (B) Linear equivalent region for $\mathbf{F} = \begin{pmatrix} 1 & 2 \\ 2 & 3.3 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$, $tol = 1$. The center of the hyper-quadic is $\mathbf{m}_c = \mathbf{F}^{-1}\mathbf{b} = \begin{pmatrix} 0.14 \\ 1.43 \end{pmatrix}$. (C) Linear equivalent region for $\mathbf{F} = \begin{pmatrix} 1 & 2 \\ 2 & 3.9 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$, and $tol = 1$. The system becomes almost rank deficient. The center of the hyper-quadic is now $\mathbf{m}_c = \mathbf{F}^{-1}\mathbf{b} = \begin{pmatrix} -17.0 \\ 10.0 \end{pmatrix}$ and the region of equivalence almost degenerates to an elliptical cylinder (two straight lines in this case).

the lowest singular value. This ellipse will become a circle only if the matrix of the forward operator is orthogonal, because in this case $\mathbf{F}\mathbf{F}^T = \mathbf{F}^T\mathbf{F} = \mathbf{I}_n$, that is, the singular values of \mathbf{F} are equal to one. This is the case of perfect conditioning.

Fig. 1B shows how this ellipse elongates in the direction of the \mathbf{V} vector associated to the minimum singular value when the conditioning of the matrix \mathbf{F} worsens, becoming an elongated valley in this direction, and two parallel lines in the limit case when the matrix becomes singular (rank $(\mathbf{F}) = 1$). In this case the inverse operator does not exist anymore and the solution does not depend continuously on the data. The right singular vectors of \mathbf{F} (\mathbf{V} base) associated to the zero singular value expanding the null space of \mathbf{F} , do not have any influence on the data predictions, and they belong to the uncertainty space of our simple linear inverse problem. Taking into account definition (11) uncertainty does not only occur in the directions spanning the null-space of \mathbf{F} , but also in the other directions of the \mathbf{V} base, according to the values of $\frac{1}{\mu_k}$.

The condition number of the matrix \mathbf{F} is defined as the ratio of the maximum and minimum singular values of \mathbf{F} :

$$\kappa(\mathbf{F}) = \frac{\mu_{\max}}{\mu_{\min}} = \frac{1/\mu_{\min}}{1/\mu_{\max}}. \quad (13)$$

$\kappa(\mathbf{F})$ is obviously related to the eccentricity of the linear region of equivalence, which is an ellipse in this simple 2D case:

$$\varepsilon = \sqrt{1 - \left(\frac{\mu_{\min}}{\mu_{\max}}\right)^2} = \sqrt{1 - \left(\frac{1}{\kappa(\mathbf{F})}\right)^2}. \quad (14)$$

Ill-conditioned linear problems are those whose region of equivalence eccentricity tends to 1 and the equivalent models are located in oblong and elongated rectilinear valleys of the cost function topography. Thus, ill-conditioning should be viewed as a geometric property of the linear hyper-quadratic.

Nevertheless in inverse problems ill-conditioning is always related to the fact that the determination of the solution of the linear system, which is the center of this ellipse (linear equivalent region in the general case), is very inaccurate. The determination of the center ($\mathbf{m}_c = \mathbf{F}^{-1}\mathbf{d}$) is obvious if the conic is a circle. This is the case of a well-conditioned linear problem. In two dimensions this situation happens if the associate vectors of the hyper planes forming the linear system are orthogonal to each other. Conversely, if these vectors are almost parallel, the determination of the center of the ellipse is very difficult, and its eccentricity is almost 1. Thus, the following alternative definition might be proposed: "A linear inverse problem $\mathbf{F}\mathbf{m} = \mathbf{d}$ is ill-conditioned if the eccentricity of its region of equivalence is close to 1".

This definition is more geometric than the usual definition in terms of the condition number, and it is also more general than its implication about the difficult determination of the center of the hyper-quadratic due to effect of the noise in data.

Noise in data perturbs the location of the center of the hyper-quadratic since:

$$\mathbf{m}_{sol} = \mathbf{F}^{-1}(\mathbf{d} + \delta\mathbf{d}) = \mathbf{m}_c + \mathbf{F}^{-1}\delta\mathbf{d} = \mathbf{m}_c + \Delta\mathbf{m}_{noise}. \quad (15)$$

In the case of ill conditioned linear systems, the computation of \mathbf{m}_{sol} is very inaccurate due to the effect of the very small singular values of \mathbf{F} :

$$\mathbf{m}_{sol} = \mathbf{F}^{-1}(\mathbf{d} + \delta\mathbf{d}) = \mathbf{V}\Sigma^{-1}\mathbf{U}^T(\mathbf{d} + \delta\mathbf{d}) = \mathbf{V}\Sigma^{-1}(\mathbf{d} + \delta\mathbf{d})_{\mathbf{U}} = \sum_{k=1}^{r_l} \frac{(\mathbf{d} + \delta\mathbf{d})_{\mathbf{U}k}\mathbf{v}_k}{\mu_k} + \sum_{k=r_l+1}^n \frac{(\mathbf{d} + \delta\mathbf{d})_{\mathbf{U}k}\mathbf{v}_k}{\mu_k}, \quad (16)$$

where r_l indicates the number of singular values that generate the stable part of the solution, and n is in this case the rank of \mathbf{F} . This relationship can be generalized to ill-conditioned rectangular systems, replacing \mathbf{F}^{-1} by

the Moore–Penrose pseudo-inverse, \mathbf{F}^\dagger . It can be observed that noise in data $\delta\mathbf{d}$ gets amplified back to the solution \mathbf{m}_{sol} through the non-stable part of the forward pseudo-inverse $\left(\sum_{k=r_l+1}^n \frac{(\mathbf{d} + \delta\mathbf{d})_{\mathbf{U}k}\mathbf{v}_k}{\mu_k}\right)$ formed by the right singular vectors spanning the "numerical kernel" of \mathbf{F} .

2.2. The case of rectangular linear systems

The above mentioned analysis can be generalized to incompatible linear systems with rectangular matrix. In this case the region of equivalence, M_{tol} , is the pre-image (\mathbf{F}^{-1}) of a compact set in the data space, the ball center in \mathbf{d} of radius tol :

$$M_{tol} = \mathbf{F}^{-1}(\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2 \leq tol). \quad (17)$$

The tol value in formula (17) has to be interpreted as relative error tolerance, such as in formula (11).

The linear region of equivalence is now the hyper-quadratic (Fernández-Martínez et al., 2012b):

$$\mathbf{m}^T \mathbf{F}^T \mathbf{F} \mathbf{m} - 2\mathbf{m}^T \mathbf{F}^T \mathbf{d} + \mathbf{d}^T \mathbf{d} = tol^2, \quad (18)$$

or

$$\mathbf{m}_V^T \Sigma^T \Sigma \mathbf{m}_V - 2\mathbf{d}_U^T \Sigma \mathbf{m}_V + \mathbf{d}_U^T \mathbf{d}_U = tol^2, \quad (19)$$

if the hyper-quadratic is referred to the \mathbf{U} and \mathbf{V} orthonormal bases provided by the singular value decomposition of \mathbf{F} .

This linear hyper-quadratic is a hyper-ellipsoid for purely (or full-rank) over-determined linear systems, coinciding its center, $\mathbf{m}_c = \mathbf{F}^\dagger \mathbf{d}$, with the least squares solution of the linear inverse problem $\mathbf{F}\mathbf{m} \approx \mathbf{d}$. If the linear system is ill-conditioned the axes of the hyper-quadratic along the directions of the \mathbf{V} base corresponding to the smallest singular values μ_k are very long, and the determination of the center of the hyper-quadratic becomes very inaccurate in presence of noise. In the case of purely underdetermined or rank-deficient (either underdetermined or over-determined) linear systems, the hyper-quadratic degenerates to an elliptical cylinder with infinitely long axes (corresponding to the null singular values) in the directions spanning the kernel of the forward operator. In this case it is impossible to determine its center; therefore the least squares solution is not unique. The minimum norm solution (generalized least squares solution) is positioned along the valley and is the closest solution to the origin of the model space (null model vector). This solution has not components on the kernel of \mathbf{F} and it is named the natural or simplest solution. This solution is also sparse.

2.3. The effect of the regularization in linear inverse problems revisited

There are two main ways of stabilizing the inversion: truncation and damping. The Moore–Penrose pseudo-inverse truncates the spectrum of \mathbf{F} and provides minimum norm solutions, since no components in the null space of \mathbf{F} are added to the solution. For purely underdetermined linear systems, the solution is found in the orthogonal space to $\ker \mathbf{F}$, that is, in the column space of \mathbf{F}^T , $\text{Im } \mathbf{F}^T$. Thus, the solution lies in the intersection between $\text{Im } \mathbf{F}^T$ and the linear variety containing the solutions of the linear system. In the case of rank deficient linear systems (mixed type), the Moore–Penrose pseudo-inverse provides the minimum norm solution of the least squares normal equations that have infinite solutions. Thus, the pseudo-inverse solution lies in the intersection of the linear variety containing the solutions of the normal equations, $\mathbf{F}^T \mathbf{F} \mathbf{m}_C = \mathbf{F}^T \mathbf{d}$, and the null space of $\mathbf{F}^T \mathbf{F}$, that coincides with the null space of \mathbf{F} ($\ker \mathbf{F}^T \mathbf{F} \equiv \ker \mathbf{F}$). Either, in purely underdetermined systems or in mixed type systems, the minimum norm solution has not components of the null space of \mathbf{F} and for that reason is called the least squares natural solution.

By the other hand, zero-order Tikhonov regularization (Tikhonov and Arsenin, 1977) is the most commonly method used for regularization for ill-conditioned inverse problems, stabilizing their solution by adding an extra-term to the cost functional that penalizes the size of the solution in some sense:

$$\begin{aligned} \|\mathbf{Fm} - \mathbf{d}\|_2^2 + \varepsilon^2 \|\mathbf{m}\|_2^2 &= (\mathbf{Fm} - \mathbf{d})^T (\mathbf{Fm} - \mathbf{d}) + \varepsilon^2 \mathbf{m}^T \mathbf{m} \\ &= \|\mathbf{d}\|_2^2 + \mathbf{m}^T (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n) \mathbf{m} - 2\mathbf{m}^T \mathbf{F}^T \mathbf{d}, \end{aligned} \quad (20)$$

where ε^2 is the damping parameter that is close to zero.

In the case of linear inverse problems it is easy to understand why expression (20) serves to stabilize the inversion. Now, the normal equations are:

$$(\mathbf{F}\mathbf{F}^T + \varepsilon^2 \mathbf{I}_n) \mathbf{m} = \mathbf{F}^T \mathbf{d}, \quad (21)$$

and the operator $\mathbf{F}\mathbf{F}^T + \varepsilon^2 \mathbf{I}_n$ becomes full-rank, that is, a bijection. Thus, the inverse operator $(\mathbf{F}\mathbf{F}^T + \varepsilon^2 \mathbf{I}_n)^{-1}$ exists and it is continuous.

Tikhonov's regularization has two main effects:

- (1) The linear hyper-quadric becomes a hyper-ellipsoid whose longer axes are $1/\varepsilon$ in the directions of the \mathbf{V} vectors spanning the null space of \mathbf{F} ($\ker \mathbf{F}$).
- (2) The center of the hyper-quadric is shifted to $\mathbf{m}_\varepsilon^c = (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n)^{-1} \mathbf{F}^T \mathbf{d}$, and its main axes are slightly rotated.

Obviously, these two effects are almost negligible when the regularization parameter tends to zero ($\varepsilon^2 \rightarrow 0$). The main effect of the regularization is to limit the length of the misfit valley, causing the linear region of equivalence to become bounded. Thus, the zero-order Tikhonov's regularization is a practical way of stabilizing the determination of the center of the linear hyper-quadric, but its location might be perturbed by the presence of noise in data.

In terms of the optimal solution found the zero-order Tikhonov's regularization with no model of reference, does not add to the solution any effect from the base terms (the set of vectors $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ defined by the singular value decomposition of \mathbf{F}) spanning the null space of \mathbf{F} , since referred to the \mathbf{V} and \mathbf{U} bases the solution is:

$$\mathbf{m}_{\varepsilon V}^c = (\boldsymbol{\Sigma}^T \boldsymbol{\Sigma} + \varepsilon^2 \mathbf{I}_n)^{-1} \boldsymbol{\Sigma}^T \mathbf{d}_U = \sum_{k=1}^r \frac{\mu_k d_{Uk}}{\mu_k^2 + \varepsilon^2} \mathbf{v}_k. \quad (22)$$

Nevertheless, when a prior model is used in the regularization term (as in expression (3)), the solution becomes:

$$\mathbf{m}_\varepsilon = \mathbf{m}_{ref} + (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n)^{-1} \mathbf{F}^T (\mathbf{d} - \mathbf{Fm}_{ref}), \quad (23)$$

or referred to the \mathbf{V} base: \mathbf{m}^{ref}

$$\mathbf{m}_\varepsilon^V = \mathbf{m}_V^{ref} + (\boldsymbol{\Sigma}^T \boldsymbol{\Sigma} + \varepsilon^2 \mathbf{I}_n)^{-1} \boldsymbol{\Sigma}^T \Delta \mathbf{d}_U^{ref}. \quad (24)$$

$\Delta \mathbf{d}_U^{ref}$ is the reference model data misfit referred to the \mathbf{U} base, that is, $\Delta \mathbf{d}_U^{ref} = \mathbf{U}^T (\mathbf{d} - \mathbf{Fm}_{ref})$. The solution in Eq. (24) can finally be expressed as:

$$\mathbf{m}_\varepsilon^V = \sum_{k=1}^r \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{Uk}}{\mu_k^2 + \varepsilon^2} \mathbf{v}_k + \sum_{j=r+1}^n m_{V_j}^{ref} \mathbf{v}_j. \quad (25)$$

Expression (25) shows that now all the vectors in the \mathbf{V} base play a role in the regularized solution, \mathbf{m}_ε^c , since the coordinates that originally resided in the kernel of \mathbf{F} are now informed by the reference model \mathbf{m}_V^{ref} . Also, the coordinates in the r first vectors of the \mathbf{V} base are linear combinations of the reference model coordinates $m_{V_k}^{ref}$ and the data d_{Uk} . The weights of the linear combination are respectively ε^2 and μ_k . Fig. 2 shows a geometric interpretation of this fact: without model of

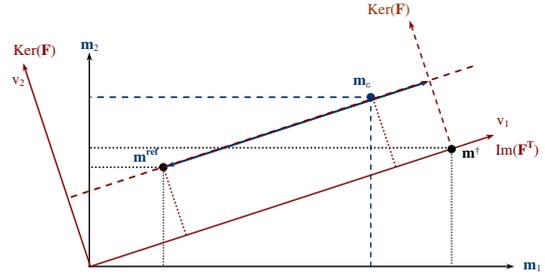


Fig. 2. Geometric interpretation of the regularization in two dimensions. Without model of reference, the zero order regularized solution has no components on the null space of \mathbf{F} . The reference model serves to incorporate information from the kernel.

reference, the Moore Penrose pseudo-inverse and the zero-order regularized solutions have no components on the null space of \mathbf{F} . The reference model serves to incorporate information from the kernel.

In other cases, high pass discrete difference operators Γ may be used to enforce smoothness in the solution (first or second order Tikhonov regularization). In geophysics, Occam's inversion is widely used (Constable et al., 1987) to generate these types of smooth models in linear and nonlinear inverse problems:

$$\|\mathbf{Fm} - \mathbf{d}\|_2^2 + \varepsilon^2 \|\Gamma(\mathbf{m} - \mathbf{m}_{ref})\|_2^2. \quad (26)$$

First and second order Tikhonov's regularizations penalize the roughness of the solution found, and also have the effect of incorporating components from $\ker \mathbf{F}$ into the solution. This is one of the reasons why this kind of regularization is widely used in industry. Nevertheless modelers do not realize that this incorporation is somehow artificial, because originally the inverse problem does not provide any information about these components.

Finally, it is very important to remark that the regularization does not cause the disappearance of the equivalent solutions. Fig. 3 illustrates the effect of the regularization on the nonlinear region of equivalence, M_{tol} . The regularization acts by restricting the optimization to the models that are on the intersection of M_{tol} and the ball $B(\mathbf{m}_{ref}; \rho) = \{\mathbf{m} : \|\mathbf{m} - \mathbf{m}_{ref}\| \leq \rho\}$, centered in \mathbf{m}_{ref} and a radius ρ , which depends on the damping parameter ε^2 that is used. The solution of the regularized inverse problem is the model $\mathbf{m}_\varepsilon \in B(\mathbf{m}_{ref}; \rho)$ that gives the closest prediction to the observed data \mathbf{d} . For small

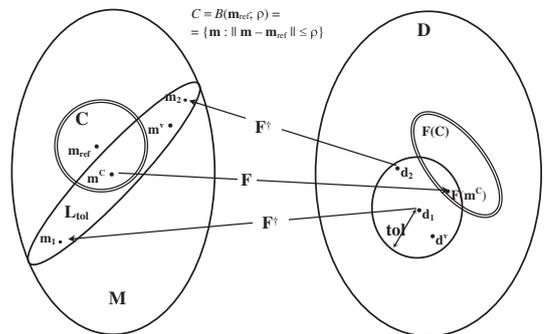


Fig. 3. Region of equivalence for a linear inverse problem. Region M_{tol} is defined as the pre-image of the ball centered in the observed data and having a radius equal to tol . The region is simple connected (hyper-quadric). The lack of injectivity of the forward operator provokes this region to become unbounded in the directions of the \mathbf{V} base spanning the kernel. The reference model serves to restrict the equivalent models to the intersection of M_{tol} and the ball centered in the model of reference (correctness set).

levels of noise, the regularized solution tends to the “true” solution when the damping parameter approaches to zero.

3. Nonlinear inverse problems

For nonlinear inverse problems we can apply the same kind of reasoning as we did for the linear inverse problems using their first order approximation in order to characterize the local regions of equivalence around a certain model \mathbf{m}_0 :

$$\mathbf{F}(\mathbf{m}) \approx \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0). \quad (27)$$

Then we have:

$$C(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 = \|\Delta\mathbf{d}^*\|_2^2 + \mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \mathbf{m} + 2\mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \Delta\mathbf{d}^*, \quad (28)$$

with

$$\Delta\mathbf{d}^* = \mathbf{F}(\mathbf{m}_0) - \mathbf{d} - \mathbf{JF}_{\mathbf{m}_0} \mathbf{m}_0. \quad (29)$$

The linearized region of equivalence around the model \mathbf{m}_0

$$\|\Delta\mathbf{d}^*\|_2^2 + \mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \mathbf{m} + 2\mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \Delta\mathbf{d}^* \leq \text{tol}^2, \quad (30)$$

or equivalently

$$(\mathbf{m} - \mathbf{m}_0)^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) + 2(\mathbf{m} - \mathbf{m}_0)^T \mathbf{JF}_{\mathbf{m}_0}^T \Delta\mathbf{d} + \|\Delta\mathbf{d}\|_2^2 \leq \text{tol}^2, \quad (31)$$

where $\Delta\mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}$ is the data misfit corresponding to the model \mathbf{m}_0 .

Thus, the linear region of equivalence in the nonlinear case involves the Jacobian of \mathbf{F} (i.e., the sensitivity matrix) calculated in the model \mathbf{m}_0 . The linear region of equivalence expressed in (30) and (31) approximates only locally the region of nonlinear equivalence, that is, models in the neighborhood of \mathbf{m}_0 fulfilling the condition $\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2 \leq \text{tol}$. Due to the dependency of the Jacobian of \mathbf{F} on the model \mathbf{m}_0 and also to higher order terms in $\mathbf{F}(\mathbf{m})$ (for instance the Hessian) the nonlinear region of equivalence has to bend forming a croissant-shape or boomerang-like regions (Fernández-Martínez et al., 2012b). Infinite size meandering regions where the forward predictions do not change could also theoretically exist depending on the nature of the forward operator \mathbf{F} .

The center of the linear hyper-quadratic in the nonlinear case coincides with the Gauss–Newton solution of the original nonlinear inverse problem (Fernández-Martínez et al., 2012b). To prove this fact, we just need to minimize the following second order approximation of the cost functional:

$$\min_{\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0} \|\Delta\mathbf{d}\|_2^2 + 2\Delta\mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \Delta\mathbf{d} + \Delta\mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \Delta\mathbf{m}. \quad (32)$$

The Gauss–Newton solution of the non-linear inverse problem solves the following linear system:

$$\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \Delta\mathbf{m} = -\mathbf{JF}_{\mathbf{m}_0}^T \Delta\mathbf{d}, \quad (33)$$

that is,

$$\mathbf{m}_{GN} = \mathbf{m} - \left(\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger} \mathbf{JF}_{\mathbf{m}_0}^T \Delta\mathbf{d}, \quad (34)$$

and coincides with the center of the linear hyper-quadratic stated in Eq. (30). Thus, the Gauss–Newton is also affected by the ill-conditioning of the linear system in Eq. (33) and the presence of noise in data.

Creeping and jumping solutions to the nonlinear inverse problem have been also proposed in the literature (Parker, 1994). These solutions are deduced considering by first order Taylor Series expansion of $\mathbf{F}(\mathbf{m})$ (as in expression (27)), and solving the linear system:

$$\mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) \approx \mathbf{d}, \quad (35)$$

in the least squares sense. The formulation using increments is called the creeping solution \mathbf{m}_{CP} to the nonlinear inverse problem:

$$\begin{aligned} \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m}_{CP} - \mathbf{m}_0) &\approx \mathbf{d} - \mathbf{F}(\mathbf{m}_0) \Rightarrow \\ \mathbf{m}_{CP} &= \mathbf{m}_0 + \left(\mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger} (\mathbf{d} - \mathbf{F}(\mathbf{m}_0)). \end{aligned} \quad (36)$$

Creeping solution coincides with the Gauss–Newton solution when:

$$\left(\mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger} = \left(\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger} \mathbf{JF}_{\mathbf{m}_0}^T \quad (37)$$

that is, in the case of purely over-determined linearized systems where the number of observed data is higher than the number of parameters and $\mathbf{JF}_{\mathbf{m}_0}$ is full rank. In other cases both solutions might be different depending on how the generalized inverse $\left(\mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger}$ is computed.

Jumping solution \mathbf{m}_J refers to the least squares solution of the linear system

$$\mathbf{JF}_{\mathbf{m}_0} \mathbf{m}_J \approx \mathbf{d} - \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0} \mathbf{m}_0, \quad (38)$$

that is:

$$\mathbf{m}_J = \left(\mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger} (\mathbf{d} - \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0} \mathbf{m}_0) = \mathbf{m}_{CP} + \left[\left(\mathbf{JF}_{\mathbf{m}_0}\right)^{\dagger} \mathbf{JF}_{\mathbf{m}_0} - \mathbf{I}_n\right] \mathbf{m}_0. \quad (39)$$

Jumping and creeping solutions coincide for purely over-determined linearized systems, but differ in the case of purely underdetermined and mixed type systems (Parker, 1994). This fact also shows the dependency between the solution and the numerical method that has been adopted to solve iteratively the nonlinear inverse problem. Obviously, all these possible solutions belong to the nonlinear region of equivalence.

As other local optimization methods, jumping and creeping solutions are not designed to analyze the nonlinear uncertainty in inverse problems. In the next section we explain why these methods fail in approaching nonlinear uncertainty.

3.1. Why local optimization algorithms fail in approaching uncertainty?

The main reason is that these methods approach the cost function in the neighborhood of a model \mathbf{m}_0 with the corresponding local hyper-quadratic. To see that, let us consider a second order Taylor series expansion centered in the model \mathbf{m}_0 :

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + \mathbf{ST}_{\mathbf{m}_0} + \mathbf{o}\left(\|\mathbf{m} - \mathbf{m}_0\|_2^2\right), \quad (40)$$

where $\mathbf{JF}_{\mathbf{m}_0} = \left(\frac{\partial \mathbf{F}}{\partial m_1}(\mathbf{m}_0), \frac{\partial \mathbf{F}}{\partial m_2}(\mathbf{m}_0), \dots, \frac{\partial \mathbf{F}}{\partial m_n}(\mathbf{m}_0)\right)$ is the Jacobian matrix of \mathbf{F} in \mathbf{m}_0 ; $\mathbf{ST}_{\mathbf{m}_0}$ is the second order term including the Hessians of the function components ($f_i(\mathbf{m})$) of the forward operator \mathbf{F} ,

$$\mathbf{ST}_{\mathbf{m}_0} = \frac{1}{2} \begin{pmatrix} (\mathbf{m} - \mathbf{m}_0)^T Hf_1(\mathbf{m}_0)(\mathbf{m} - \mathbf{m}_0) \\ (\mathbf{m} - \mathbf{m}_0)^T Hf_2(\mathbf{m}_0)(\mathbf{m} - \mathbf{m}_0) \\ \vdots \\ (\mathbf{m} - \mathbf{m}_0)^T Hf_s(\mathbf{m}_0)(\mathbf{m} - \mathbf{m}_0) \end{pmatrix}, \quad (41)$$

and $\mathbf{o}\left(\|\mathbf{m} - \mathbf{m}_0\|_2^2\right)$ is a vector function that vanishes faster than $\|\mathbf{m} - \mathbf{m}_0\|_2^2$. Considering the second order approximation written

in Eq. (40), the cost function can be approximated the following second order polynomial:

$$C(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \approx \|\Delta\mathbf{d} + \mathbf{J}\mathbf{F}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + \mathbf{S}\mathbf{T}_{\mathbf{m}_0}\|_2^2 \approx \|\Delta\mathbf{d}\|_2^2 + 2(\mathbf{m} - \mathbf{m}_0)^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \Delta\mathbf{d} + (\mathbf{m} - \mathbf{m}_0)^T \mathbf{H}\mathbf{C}(\mathbf{m}_0)(\mathbf{m} - \mathbf{m}_0), \quad (42)$$

where $\mathbf{H}\mathbf{C}(\mathbf{m}_0)$ is the exact Hessian of the cost function $C(\mathbf{m})$:

$$\mathbf{H}\mathbf{C}(\mathbf{m}_0) = \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} + \sum_{k=1}^s (f_k(\mathbf{m}_0) - d_k) \mathbf{H}f_k(\mathbf{m}_0). \quad (43)$$

A linear combination of positive definite matrices is guaranteed to be also definite positive if the weights are positive. This does not hold in the general case here, since each of the Hessian matrices $\mathbf{H}f_k$ might not even have a positive semi-definite positive character. Also the weights $(f_k(\mathbf{m}_0) - d_k)$ might not be positive. Thus, the term $\sum_{k=1}^s (f_k(\mathbf{m}_0) - d_k) \mathbf{H}f_k(\mathbf{m}_0)$ in Eq. (43) usually causes the exact Hessian of $C(\mathbf{m})$ to lose its positive semi-definite character. In this case, the exact second order equivalent region can have sill points connecting different basins of equivalent solutions.

Most of the local optimization methods approximate the curvature of the cost function by the linearized Hessian, $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}$, which is semi-definite positive. Therefore, the saddle points of the cost function cannot be found in these analysis, and the uncertainty is reduced to the linearized hyper-quadric. Otherwise said, local optimization methods are not designed to explore and locate different basins of solutions, due to the numerical approximations that are used in their design. It is also worth mentioning that the computation of $\mathbf{H}\mathbf{C}(\mathbf{m}_0)$ is computationally very expensive, even impossible, in high dimensional inverse problems.

Finally, Fig. 4 shows the effect of the regularization searching the solution in the nonlinear region of equivalence. As in the linear case the regularization acts by restricting the optimization to the models that are on the intersection of M_{tol} and the ball $B(\mathbf{m}_{ref}; \rho) = \{\mathbf{m} : \|\mathbf{m} - \mathbf{m}_{ref}\| \leq \rho\}$. The solution of the regularized nonlinear inverse problem is the model $\mathbf{m}_c \in B(\mathbf{m}_{ref}; \rho)$ that gives the closest prediction to the observed data \mathbf{d} . Nevertheless, in this case several unconnected basins of low misfits can coexist in the cost function landscape if \mathbf{F} is no injective. The “true” solution would not be found if the ball $B(\mathbf{m}_{ref}; \rho)$ does not intersect the basin where the true solution lies. Also, the linearized region of

equivalence calculated in a model that belongs to the nonlinear equivalent region only captures a local part of the nonlinear uncertainty space, when it does. This figure also serves to explain the dangers of the regularization and to highlight that the regularization does not cause the uncertainty to vanish.

3.2. The effect of model parameterization

We finally analyze the effect of the model re-parameterization in the uncertainty analysis of the solution. In some cases model re-parameterization is used to linearize some kind of inverse problems. Nevertheless, this procedure has also some dangers, since it is important to know that the equivalent region in the linearized case and the nonlinear equivalent region are different. Also, both problems might behave very different with respect to noise in data.

To show these facts numerically, let us consider for instance a simple example of nonlinear regression $z = Kx^\alpha y^\beta, K \in \mathbb{R}$. This problem consists in finding the real constants α, β that provides minimum fitting error in the least squares sense of the experimental data:

$$\{(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_s, y_s, z_s)\}.$$

The nonlinear equivalent region M_{tol} of value tol is defined as follows:

$$M_{tol} = \left\{ \mathbf{m} = (\alpha, \beta) \in \mathbb{R}^2 : \sqrt{\sum_{k=1}^m (z_k - Kx_k^\alpha y_k^\beta)^2} < \|\mathbf{z}\|_2 \cdot tol \right\}. \quad (44)$$

The above mentioned inverse problem can be linearized as follows:

$$\ln z_j - \ln K = \alpha \ln x_j + \beta \ln y_j, \quad j = 1 \dots s. \quad (45)$$

This provides an over-determined linear system of the type:

$$\begin{pmatrix} \ln x_1 & \ln y_1 \\ \ln x_2 & \ln y_2 \\ \vdots & \vdots \\ \ln x_s & \ln y_s \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \ln z_1 - \ln K \\ \ln z_2 - \ln K \\ \vdots \\ \ln z_s - \ln K \end{pmatrix}. \quad (46)$$

Calling \mathbf{A} the coefficient matrix, and \mathbf{d} the right hand term of the linear system (46), the linearized equivalent region is:

$$L_{tol} = \left\{ \mathbf{m} = (\alpha, \beta) \in \mathbb{R}^2 : \|\mathbf{A}\mathbf{m} - \mathbf{d}\|_2 < \|\mathbf{d}\|_2 \cdot tol \right\}. \quad (47)$$

Relative misfits are used in this case for comparative purposes between both regions of equivalence, M_{tol} and L_{tol} .

We have designed a synthetic experiment where the observed data (100 points) was generated using $K = 3; \alpha = 2; \beta = 1$. Fig. 5A and B shows the different contours of M_{tol} and L_{tol} . We also show the region corresponding to $tol = 5\%$ in both cases. It can be observed that both regions differ in size and also in orientation. Typically L_{tol} is broader than M_{tol} and also the misfits are lower for L_{tol} (data over-fitting). The first fact serves to explain that the linearized problem usually has a worse conditioning than the original nonlinear problem. Also if the problem was highly nonlinear the croissant shape will be more pronounced. Fig. 5C and D shows the effect of a Gaussian noise of zero mean and the standard deviation equal to 10% of the standard deviation of the observed data. It is possible to observe that noise deforms the topography of the cost function in both cases and shifts the minimum misfit solutions. Also, the nonlinear inverse problem seems to be more robust than the linearized inverse problem, but this fact also depends on the type of noise and its projection (correlation) onto the observed data. Obviously this simple numerical experiment serves to

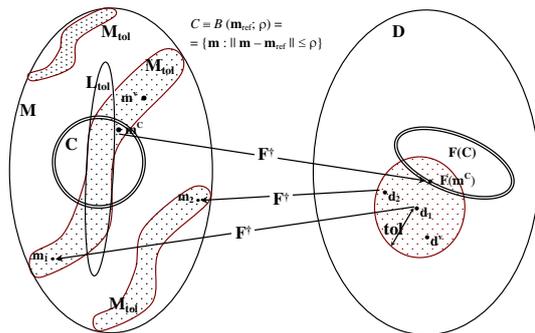


Fig. 4. Region of equivalence for a nonlinear inverse problem. In this case the region M_{tol} has a meandering shape. Different unconnected regions might exist depending on the injectivity of the forward nonlinear operator \mathbf{F} . As in the linear case the reference model serves to restrict the equivalent models to the intersection of M_{tol} and the ball centered in the model of reference 2 (correctness set); nevertheless the solution of the regularized problem might reside in a different basin than the true solution, depending on the correctness set induced by the regularization and the reference model. L_{tol} stands for the linear hyper quadric computed in a model of the nonlinear equivalent region.

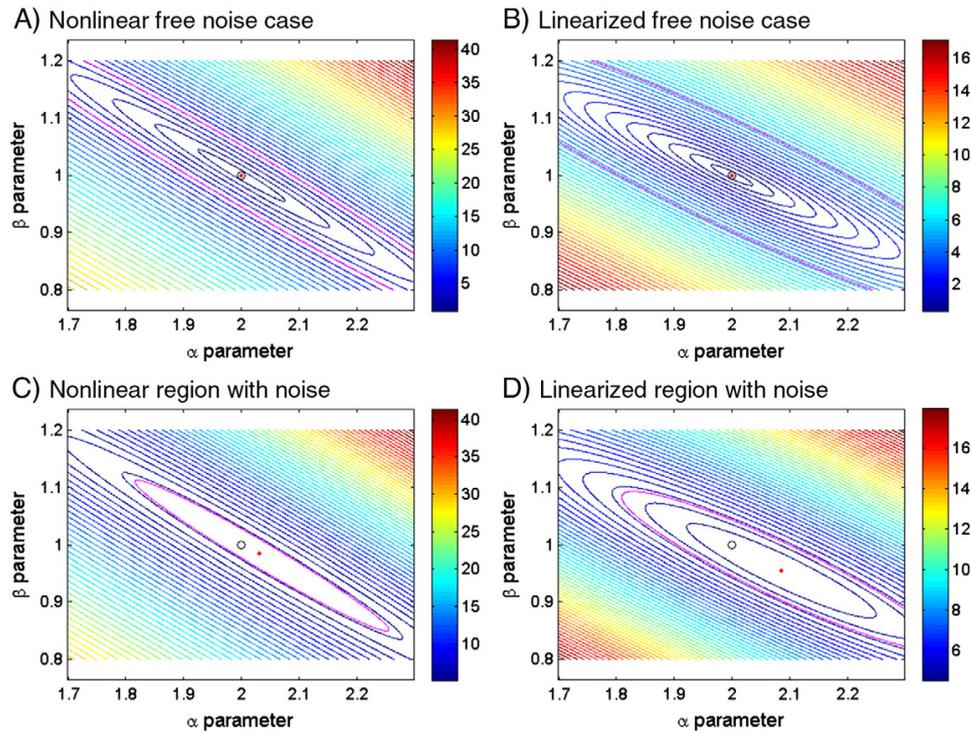


Fig. 5. Effect of the model re-parameterization in a simple nonlinear regression problem. (A) Nonlinear equivalent region; (B) Linearized equivalent region; (C) Nonlinear equivalent region affected by noise; (D) Linearized equivalent region affected by noise. In magenta we show the limits of the region of 5% misfit.

exemplify the pitfalls that the modeler might find when model re-parameterizations are used.

4. Implications in applied geophysics, conclusions and future solutions

Inverse problems in applied geophysics are always discrete and usually ill-conditioned. Also, the geophysical fields of interest are often quite weak and must be observed in an environment that contains extraneous fields that contribute noise to the observational data. Thus, the typical situation in geophysical inversion is to deal with data that has finite dimension, is insufficient in number and also inaccurate due to the effect of noise (Barhen et al., 2000). These features together with the modeling and numerical errors cause the inverse solution to be uncertain, and the geophysical inverse problem has to be solved in two different stages: solution optimization/search and appraisal. The appraisal stage (or uncertainty analysis of the solution) is motivated by the fact that the observed data does not contain enough information to determine a geophysical plausible unique solution.

Practitioners tend to underestimate the uncertainty of inverse problems solutions relying on the fact that: 1) Uncertainty has in any case a random structure impossible to be known. 2) Regularization techniques are the panacea and cause the disappearance of the equivalent solutions. These equivalent solutions are sometimes annoying, thus modelers prefer to ignore its existence, calling them spurious. 3) Linearization techniques can provide a precise idea about the solution that has been adopted.

All these three assumptions are incorrect, motivating in some cases that decisions that are taken based on these solutions are also

incorrect, and from time to time, oil exploration boreholes based on wrong solutions to be dry!

This paper provides new insights to understand uncertainty in the solution of inverse problem from a deterministic point of view. Linear algebra is the part of mathematics that deals with problems that have an underlying structure, such as the space of solutions for linear systems of equations. Although linear algebra techniques do not directly applies to nonlinear inverse problems, the use of iterative linearization techniques serves to understand the local properties of the space of solutions that are imbedded in the gradient and Hessian of the cost function.

Local optimization methods and regularization techniques fail to account for the uncertainty analysis of nonlinear inverse problems because they approximate the nonlinear region of equivalent models by the linearized hyper-quadric that only reflects locally (when it does) the structure of the space of equivalent solutions. Bayesian approaches and global optimization methods are also very inefficient because they tend to ignore the “algebraic (or geometric) structure” of the space of equivalent solutions, and they are hampered by dimensionality issues and the high computational cost to solve their respective forward problems.

Understanding uncertainty from a deterministic point of view is a necessary step to design more efficient methods to sample the uncertainty space in nonlinear inverse problems. As H. Poincaré has said, nothing is random, only our ignorance. In our opinion statistical methods should be restricted to problems where the underlying physical laws are unknown and data-driven reasoning is the only way to approach the prediction/learning/decision problem. In any case, given the explanations and proofs provided in this paper, we can affirm without any doubt that epistemological uncertainty in inverse problems has not a random structure.

Further research is needed to incorporate mathematical analysis into these sampling approaches. This paper proposes a paradigm shift from random sampling approaches to a redesign of old-fashioned, forgotten and despised deterministic approaches. In essence, uncertainty in nonlinear inverse problems needs not be treated as a random process, since the equivalencies have a natural algebraic interpretation (Fernández-Martínez et al., 2012b). Local optimization methods can be redesigned to perform sampling while optimizing. By extending the topological relationships from linear to nonlinear problems, it is possible to develop more efficient algorithms to perform the posterior sampling and significantly improve the efficiency of existing methodologies for uncertainty estimation for high dimensional nonlinear inverse problems. This subject will be addressed in future research papers.

Model reduction techniques and very sparse sampling techniques also play a very important role in the design of these algorithms able to approach uncertainty in very high dimensions and with very costly forward problems. An example is the geometric sampling algorithm (Tompkins et al., 2011a,b, 2013).

Coming back to Tarantola's research work (2005, 2006), he pointed out that the model space is generally quite empty of equivalent models that fit the observed data, and that we should focus the uncertainty analysis on these "interesting" regions in the model space. In his own words (page 46, Tarantola, 2005): "*large-dimensional spaces tend to be terribly empty. Hitting by chance the circle inscribed in a square is easy. Hitting by chance the sphere inscribed in a cube is a little bit more difficult. When the dimension n of the space grows, the probability of hitting the hyper-sphere inscribed in a hyper-cube rapidly tends to zero (for $n > 10$)*". This fact, known as the dimensionality curse (Curtis and Lomax, 2001), serves to explain why the random exploration of large-dimensional spaces is very difficult. Furthermore, the uncertainty is not the same in all the directions of the model space, that is, the "volumetric" region of equivalence in some directions is longer than larger. This result also serves to explain why the use of model reduction techniques and subspace methods is successful in quantifying uncertainty in high dimensions, since the nonlinear equivalence region can be approximated as a nonlinear variety in a suitable linear subspace of \mathbb{R}^n . The intuitive geometric idea illustrating this fact is to locate a curve contained in a hyper-plane. This result also suggests that the dimension of this subspace will not be greater than 10. Although the explanation given in this paper is intuitive, this fact has also been found in real high dimensional inverse problems (see for instance Tompkins et al., 2011a,b). As anecdote, 10 are the dimensions used in string theory to explain the universe.

Future approaches to quantify uncertainty should include the use of model reduction techniques (as we have justified above), if possible the use of simplified forward physics (forward surrogates) to accelerate the forward predictions, and distributed and parallel computing facilities. Taking into account the no-free lunch theorem in search and optimization (Wolpert and Macready, 1997), this paper should not be interpreted as an allegation against global optimization methods and/or Bayesian approaches.

Acknowledgments

We would like to thank Sven Treitel for his suggestions and improvements of the original manuscript. Sven is always present when a scientific selfless assistance is needed.

This article is dedicated to the memory of Albert Tarantola, who made remarkable contributions in the progress of inverse problems. Albert, all those who have learned from your work, although we did not fully agree about everything, we will be eternally grateful to you. In fact, this should be always the way of scientific discovery.

References

- Backus, G.E., Gilbert, F., 1967. Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society* 13 (1–3), 247–276.
- Backus, G.E., Gilbert, F., 1968. The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society* 16, 169–205.
- Backus, G.E., Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data. *Philosophical Transactions of the Royal Society of London, Series A* 266, 123–192.
- Barhen, J., Berryman, J.G., Borcea, L., Dennis, J., Groot-Redlin, C. de, Gilbert, F., Gill, P., Reinkenschloss, M., Johnson, L., McEvilly, T., More, J., Newman, G., Oldenburg, D., Parker, P., Porto, B., Sen, M., Torczon, V., Vasco, D., Woodward, N.B., 2000. *Optimization and Inverse Problems: Earth Sciences Division, Report LBNL-46959*. Ernest Orlando Lawrence Berkeley National Laboratory.
- Bayes, T., 1763. An essay towards solving a problem in the doctrine of chance. *Philosophical Transactions of the Royal Society of London* 53, 370–418 (communicated by Mr. Price in a letter to John Canton).
- Constable, S.C., Parker, K.L., Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from EM sounding data. *Geophysics* 52 (3), 289–300. <http://dx.doi.org/10.1190/1.1442303>.
- Curtis, A., Lomax, A., 2001. Prior information, sampling distributions and the curse of dimensionality. *Geophysics* 66 (2), 372–378. <http://dx.doi.org/10.1190/1.1444928>.
- Fernández Álvarez, J.P., Fernández-Martínez, J.L., Menéndez Pérez, C.O., 2008. Feasibility analysis of the use of binary genetic algorithms as importance samplers application to a geoelectrical VES inverse problem. *Mathematical Geosciences* 40, 375–408. <http://dx.doi.org/10.1007/s11004-008-9151y>.
- Fernández-Martínez, J.L., García Gonzalo, E., Fernández Álvarez, J.P., Kuzma, H.A., Menéndez Pérez, C.O., 2010a. PSO: a powerful algorithm to solve geophysical inverse problems. Application to a 1D-DC resistivity case. *Journal of Applied Geophysics* 71 (1), 13–25. <http://dx.doi.org/10.1016/j.jappgeo.2010.02.001>.
- Fernández-Martínez, J.L., García Gonzalo, E., Naudet, V., 2010b. Particle swarm optimization applied to the solving and appraisal of the streaming potential inverse problem. *Geophysics* 75 (4), WA3–WA15. <http://dx.doi.org/10.1190/1.3460842>.
- Fernández-Martínez, J.L., Mukerji, T., García Gonzalo, E., Suman, A., 2012a. Reservoir characterization and inversion uncertainty via a family of particle swarm optimizers. *Geophysics* 77 (1), M1–M16.
- Fernández-Martínez, J.L., Fernández Muñiz, Z., Tompkins, M.J., 2012b. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* 77 (1), W1–W15. <http://dx.doi.org/10.1190/geo2011-0341.1>.
- Hadamard, J., 1902. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin* 13, 49–52.
- Jackson, D.D., 1972. Interpretation of inaccurate, insufficient and inconsistent data. *Geophysical Journal of the Royal Astronomical Society* 28, 97–109.
- Jackson, D.D., 1979. The use of a priori data to resolve non-uniqueness in linear inversion. *Geophysical Journal of the Royal Astronomical Society* 35 (1–3), 121–136.
- Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research* 100 (B7), 12431–12447.
- Parker, R.L., 1977. Understanding inverse theory. *Annual Review of Earth and Planetary Sciences* 5, 35–64. <http://dx.doi.org/10.1146/annurev.ea.05.050177.000343>.
- Parker, R.L., 1994. *Geophysical Inverse Theory*. Princeton University Press 9780691036342.
- Popper, K., 1959. *The Logic of Scientific Discovery*. Routledge Classics 0415278449.
- Pujol, J.M., 2013. The Backus–Gilbert method and their minimum-norm solution. *Geophysics* 78 (3), 1–22.
- Sen, M., Stoffa, P.L., 1995. *Global Optimization Methods in Geophysical Inversion*. Elsevier Press, New York 0-444-81767-0.
- Snieder, R., 1991. An extension of Backus–Gilbert theory to nonlinear inverse problems. *Inverse Problems* 7, 409–433.
- Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*. SIAM, 342 pages (ISBN 0-89871-572-5).
- Tarantola, A., 2006. Popper, Bayes and the inverse problem. *Nature Physics* 2, 492–494.
- Tarantola, A., Valette, B., 1982a. Inverse problems = quest for information. *Journal of Geophysics* 50 (3), 159–170.
- Tarantola, A., Valette, B., 1982b. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics and Space Physics* 20, 219–232.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58 (1), 267–288.
- Tikhonov, A.N., Arsenin, V.Y., 1977. *Solution of Ill-posed Problems*. Wiley, New York 9780470991244.
- Tompkins, M.J., Fernández-Martínez, J.L., Alumbaugh, D.L., Mukerji, T., 2011a. Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: marine CSEM examples. *Geophysics* F263. <http://dx.doi.org/10.1190/1.3581355>.
- Tompkins, M.J., Fernández-Martínez, J.L., Fernández-Muñiz, Z., 2011b. Marine electromagnetic inverse solution appraisal and uncertainty using model-derived basis functions and sparse geometric sampling. *Geophysical Prospecting* 1–19. <http://dx.doi.org/10.1111/j.1365-2478.2011.00955.x>.
- Tompkins, M.J., Fernández-Martínez, J.L., Fernández-Muñiz, Z., 2013. Comparison of sparse grid geometric and random sampling methods in nonlinear inverse solution uncertainty estimation. *Geophysical Prospecting* 61 (1), 28–41.
- Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 67–82.
- Zhdanov, M., 2002. *Geophysical inverse theory and regularization problems*. Methods in Geochemistry and Geophysics, 36. Elsevier. ISBN: 0-444-51089-3 (36).

ARTICLE 2

**The effect of noise and Tikhonov's
regularization in inverse problems.
Part I: The linear case**

Authors:

Juan Luis Fernández Martínez
José Luis García Pallero
María Zulima Fernández Muñiz
Luis Mariano Pedruelo González

JOURNAL OF APPLIED GEOPHYSICS
2014, VOL. 108, PP. 176–185
[doi:10.1016/j.jappgeo.2014.05.006](https://doi.org/10.1016/j.jappgeo.2014.05.006)

Summary

In this article, the effect of noise and Tikhonov regularization in discrete linear inverse problems is analyzed. The aim of this paper is the understanding from an analytic point of view, and regardless of the dimension of the model space, how the noise and the regularization deform the cost function topography for linear inverse problems. The main conclusions of this work are:

- Noise in data perturbs the position of the least squares solution with respect to the real model. In the case of compatible systems this displacement is the only influence of the noise.
- Noise deforms homogeneously the cost function topography. For white and positive correlated noise, the equivalence regions always decrease in size with respect to the ones from the noise-free problem. When dealing with negative correlated noise, equivalence regions might increase in size. This fact could be used to optimize the efficiency of global search methods.
- Regularization stabilizes the solution by limiting the axes of the equivalence hyper-quadric in the directions of the null space of the forward operator, transforming the hyper-quadric from an elliptic hyper-cylinder to a hyper-ellipsoid.
- Regularization deforms the equivalence regions in a non homogeneous way, acting differently in each model component.
- Regularization does not make equivalence regions to disappear, i. e., there will still be multiple models, which will fit data below a given tolerance error.

All enumerated conclusions have been reached analytically, showing synthetic examples which confirm the theoretical results.



The effect of noise and Tikhonov's regularization in inverse problems. Part I: The linear case



J.L. Fernández-Martínez ^{a,*}, J.L.G. Pallero ^b, Z. Fernández-Muñiz ^a, L.M. Pedruelo-González ^a

^a Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain

^b ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Madrid, Spain

ARTICLE INFO

Article history:

Received 25 February 2014

Accepted 7 May 2014

Available online 20 May 2014

Keywords:

Noise

Regularization

Uncertainty analysis

Linear inversion

ABSTRACT

Inverse problems are a special kind of optimization problems because the cost function involves data, always affected by noise, that in the case of ill-conditioning is amplified back into the model parameters through the generalized inverse operator. Then, the inversion might provide spurious solutions if no regularization techniques are used. For a given misfit tolerance the models that fit the observed data are called equivalent, and are located in a region of the model space that is bounded by the linear hyper-quadratic surface. This paper analyzes in detail the role of noise in data in linear inverse problems, providing a geometrical interpretation for the role of the regularization. The noise shifts the solution found by least squares methods and deforms homogeneously the topography of the cost function, while Tikhonov's regularization transforms the linear hyper-quadratic from an elliptical cylinder to a very oblong hyper-ellipsoid in the directions that originally spanned the kernel of the linear forward operator. Furthermore, in the case of the regularization, this deformation is anisotropic affecting differently the axes of the linear hyper-quadratic. The model of reference informs the coordinates of the solution that originally resided in the kernel of the forward operator. The differences with nonlinear inversion are highlighted in the second accompanying paper. This knowledge, although theoretical at this stage, might impact how the uncertainty analysis is performed in geophysical inversion, since noise in data is always present, and good prior models are not always available.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Discrete inverse problems $\mathbf{F}(\mathbf{m}) = \mathbf{d}$ are a special kind of optimization problems because the cost function $c(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_p$ involves the observed data \mathbf{d} , that are always affected by noise. Therefore, noise will deform the topography of the cost function corresponding to the noise-free case, both, in linear and nonlinear inversions. In the previous equation \mathbf{F} stands for the forward problem, and \mathbf{m} for the set of discrete model parameters to be found.

In the case of linear inverse problems, the forward operator \mathbf{F} corresponds to a linear application between two linear spaces $\mathbf{F}: \mathbb{R}^n \rightarrow \mathbb{R}^s$, and the cost function is usually defined as the squared Euclidean norm of the residual vector

$$c(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2. \quad (1)$$

It is well-known that for ill-conditioned linear systems the noise in data is amplified back into the model parameters through the Moore–

Penrose's pseudo-inverse, providing very different solutions (called equivalent) if no regularization techniques are used (see for instance Aster et al. (2012); Menke (1989)). This amplification depends on the right singular vectors $\mathbf{v}_k \in \mathbf{V}$ of the forward operator \mathbf{F} , corresponding to the singular values $\mu_k \in \Sigma$ (considering the singular value decomposition $\mathbf{F} = \mathbf{U}\Sigma\mathbf{V}^T$) that are close to zero and “almost” contribute to span the null space of \mathbf{F} , $\ker(\mathbf{F}) = \{\mathbf{m} \in \mathbb{R}^n: \mathbf{F}\mathbf{m} = \mathbf{0}\}$, that is, the linear subspace of \mathbb{R}^n containing the models that have no influence on the data predictions. These equivalent solutions might be considered spurious if some physical assumptions are violated, such as providing negative parameter values for positive definite magnitudes. Otherwise, when prior information is at disposal, it can be used to judge their quality.

The existence of equivalent solutions has traditionally generated mistrust about the use of inverse methods in applied geophysics. Most of the times equivalent models are related to resolution issues of the geophysical techniques involved, and also to the effect of noise in data. Nevertheless, to the best of our knowledge, a complete analysis of the role of noise in linear and nonlinear inversions has not been yet performed. Most of the research has been focused in designing regularization techniques to stabilize the inversion by decreasing the impact of noise in the solution found by least squares and local optimization methods. Also, although the SVD and the Moore–Penrose's pseudo-inverse are linear algebra techniques, the effect of noise $\delta\mathbf{d}$ in data has

* Corresponding author.

E-mail addresses: jlfm@uniovi.es (J.L. Fernández-Martínez), jlg.pallero@upm.es (J.L.G. Pallero), zulima@uniovi.es (Z. Fernández-Muñiz), marianopg@educastur.princast.es (L.M. Pedruelo-González).

been in most of the cases studied through statistical analysis and numerical experiments (see for instance Banks and Grosej (1998)), that is, no analytical theory has been developed for this particular purpose.

In fact, noise δd in data can be considered as the outcome of a random discrete stochastic vector \mathbf{N} , whose statistical distribution is a priori unknown. But it is also true, that once the data have been measured, the noise in the current data set will be one of the multiple realizations of this unknown stochastic process \mathbf{N} . The problem treated in this paper does not include analyzing the statistical distribution of the observational noise, but understanding the effect of noise in the inversion for a given realization of the stochastic process \mathbf{N} . It is also important to distinguish between random noise in the data and the modeling errors, namely the error resulting from the choice of a wrong mathematical model, that is, the wrong physics when inverting a given data set, whether this physics is based on linear or nonlinear formulations. This important subject will be addressed in future research work.

The reasons that have motivated this research are the following:

1. Understanding, from an analytical point of view, how noise deforms the topography of the cost function in linear inversion. Particularly, the effect of noise is very important in geophysical inversion, since the geophysical fields of interest are often quite weak and must be observed in an environment that contains extraneous fields that contribute with noise to the observed data (Barhen et al., 2000).
2. Understanding the role of Tikhonov's regularization, comparing its effect in the topography of the cost function (linear hyper-quadratic) to the effect of the noise.
3. To expand the interest of this research beyond the need of regularization for least squares solutions, since this knowledge might be very important in the uncertainty assessment of inverse problem solutions.

The structure of this paper is as follows: Section 2 analyzes the importance of linear inversion in applied geophysics; Section 3 analyzes how noise affects the linear hyper-quadratic of equivalence; Section 4 analyzes the effect of the Tikhonov's regularization in the topography of the cost function, explaining the importance of the reference model, and outlining the main differences/similarities compared to the effect of noise; Section 5 performs a similar analysis for other regularization techniques, such as Lasso. Finally, in Section 6 we present the conclusions.

This analysis is generalized to nonlinear inverse problems in the second accompanying paper (Fernández-Martínez et al., 2014), in order to understand the main differences induced by the nonlinearities. Although theoretical at this stage, this knowledge might be very important to properly approach the uncertainty analysis in linear and nonlinear inverse problems (Fernández-Martínez et al., 2012, 2013a). Some preliminary results have been already presented at international congresses (Fernández-Martínez et al., 2013b).

2. The importance of linear inverse problems in geophysics

Discrete linear inverse problems are very important in geophysics, and come from the numerical approximation of continuous linear inverse problems of the form

$$d(s_k) = \int_a^b K(x, s_k) m(x) dx, \quad k = 1, \dots, n, \quad (2)$$

where $d(s_k)$ is the observed data at discrete locations s_k , $m(x)$ is the model function that is the unknown of the inverse problem, and $K(x, s_k)$ is the integral kernel function that relates the model and the data measured at a certain location, being a, b the theoretical limits of integration.

Different interesting problems in geophysics correspond to linear inverse problems, as for example:

1. The inverse gravity problem when the density contrast is chosen as the unknown (Blakely, 1996)

$$\Delta g_{s_k} = -G \int_a^b \frac{D}{[D^2 + (s_k - x)^2]^{3/2}} \Delta \rho(x) dx, \quad (3)$$

where G is the gravity constant, D is the constant depth of the linear body that generates the anomaly Δg at any location s_k , $\Delta \rho(x)$ is the unknown density contrast function, and $[a, b]$ denotes the limits of the body under study.

2. In electrical methods, the charge distribution q on a thin wire, of length l and radius a , maintained at a constant voltage V

$$V(z) = \frac{1}{4\pi\epsilon_0} \int_0^l \frac{q(z')}{\sqrt{a^2 + (z - z')^2}} dz'. \quad (4)$$

3. The transmission tomography inverse problem, consisting in finding the slowness distribution $s(x)$ of a geological medium such as

$$t_k = \int_{\Gamma_k} \frac{1}{s(x)} d\Gamma_k, \quad k = 1, \dots, m, \quad (5)$$

where t_k are the observed traveltimes of m pairs of sources and receivers, and Γ_k are the ray paths that are a priori unknown. This problem is in fact nonlinear due to the dependency of Γ_k on $s(x)$, but is solved iteratively through an ensemble of linear problems of the type (5), where the ray paths are known and updated through a ray tracing forward model.

4. Geophysical problems that are described by a convolution integral (Linear Time-Invariant or LTI – systems) of the kind

$$y(t) = x(t) * h(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x(t - \tau) h(\tau) d\tau. \quad (6)$$

The identification of the impulse function $h(t)$ of a LTI system originates the solution of a linear inverse problem. Continuous linear inverse problems originate discrete linear inverse problems of the kind $\mathbf{Fm} = \mathbf{d}$, due to the expansion of the solution of the continuous inverse problem in a suitable set of basis functions of a linear subspace with finite dimension.

5. Finally, the iterative solution of nonlinear geophysical inverse problems through linearization

$$\mathbf{F}(\mathbf{m}) \approx \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0), \quad (7)$$

also originates a sequence of linear inverse problems to be solved.

Therefore, the analysis of discrete linear inverse problems has a capital importance in applied geophysics.

3. Effect of noise in linear discrete inverse problems

In this section we analyze the effect of noise in linear inverse problems through linear algebra techniques, since its understanding is crucial to highlight the main differences with respect to nonlinear inversion.

Inverse problems in discrete form can be written as

$$\mathbf{F}(\mathbf{m}) \approx \mathbf{d}^{obs}, \quad (8)$$

where $\mathbf{m} = (m_1, m_2, \dots, m_n) \in \mathbf{M} \subset \mathbb{R}^n$ is the estimated model that belongs to a set of admissible models \mathbf{M} defined in terms of some prior geological knowledge, $\mathbf{d}^{obs} \in \mathbb{R}^s$ is the observed data, $\mathbf{F}(\mathbf{m}) = (f_1(\mathbf{m}), f_2(\mathbf{m}), \dots, f_s(\mathbf{m}))$ is the vector field representing the forward model, and $f_j(\mathbf{m})$ is the scalar field accounting for the j -th data.

Linear inverse problems are those where the component functions of the vector field \mathbf{F} depend linearly on the model \mathbf{m} . In this case, solving $\mathbf{F}\mathbf{m} = \mathbf{d}^{obs}$, $\mathbf{F} \in M_{(s,n)}(\mathbb{R})$ is equivalent to find the intersection of the set of s hyper-planes in \mathbb{R}^n : $f_j(\mathbf{m}) - d_j = 0$, $j = 1, \dots, s$. The presence of noise $\delta\mathbf{d}$ in the observed data \mathbf{d}^{obs} causes this intersection problem to have no solution. This incompatibility is also caused by the fact that the linear operator \mathbf{F} is only an approximation of the reality (modeling errors). Predicting the reality using \mathbf{F} is called solving the forward problem. Due to the incompatibility of the linear system of equations, the inverse problem is solved via linear least squares (LLS) methods, finding

$$\mathbf{m}^\dagger : c(\mathbf{m}^\dagger) = \min_{\mathbf{m} \in \mathbb{R}^n} \|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2, \quad (9)$$

where the generalized solution of this optimization problem is provided by the Moore–Penrose’s pseudo-inverse (see for instance Aster et al. (2012)).

Let us call $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta\mathbf{d}$, where \mathbf{d}^{true} is the observed data that would be measured in the absence of noise (noise-free data), and \mathbf{m}^{true} is the hypothetical model that has generated these data, that is, $\mathbf{F}\mathbf{m}^{true} = \mathbf{d}^{true}$. Then, we have

$$\mathbf{m}^\dagger = \mathbf{F}^\dagger \mathbf{d}^{obs} = \mathbf{F}^\dagger (\mathbf{d}^{true} + \delta\mathbf{d}) = \mathbf{m}^{true} + \Delta\mathbf{m}, \quad (10)$$

with

$$\Delta\mathbf{m} = \mathbf{F}^\dagger \delta\mathbf{d} = \mathbf{V}\Sigma^\dagger \mathbf{U}^T \delta\mathbf{d} = \sum_{k=1}^{r=\text{rank}(\mathbf{F})} \frac{\delta d_{Uk}}{\mu_k} \mathbf{v}_k, \quad (11)$$

where r is the rank of \mathbf{F} , and $\delta\mathbf{d}_U = \mathbf{U}^T \delta\mathbf{d}$ is the noise term $\delta\mathbf{d}$ referred to the \mathbf{U} orthonormal base, provided by the singular value decomposition of the forward operator \mathbf{F} . Eq. (11) can be split into two parts, the one that provides the stable part, and the one that causes noise in data to be amplified back into the model parameters, providing different equivalent solutions. This second unstable part is associated to the \mathbf{v}_k vectors of the \mathbf{V} base corresponding to the vanishing singular values $\mu_k \rightarrow 0$. This effect is called ill-conditioning of the discrete linear inverse problem. Also, in the presence of noise the solution found \mathbf{m}^\dagger will never coincide with true solution \mathbf{m}^{true} , that has generated the observed data.

Some authors include in the noise term not only the random components in the observational data, but also the modeling errors. Nevertheless, it is important to understand that the observational noise $\delta\mathbf{d}$ and the modeling errors $\Delta\mathbf{d}$ are two different concepts. The following relationships apply: $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta\mathbf{d}$, and $\mathbf{d}^{obs} = \mathbf{F}\mathbf{m}^\dagger + \Delta\mathbf{d}$. Thus, data modeling is correct only if $\delta\mathbf{d} = \mathbf{d}^{obs} - \mathbf{F}\mathbf{m}^\dagger = \Delta\mathbf{d}$, that is, if $\delta\mathbf{d}$ lies in the orthogonal space to $\text{Im}(\mathbf{F})$, which is the null space of \mathbf{F}^T ($\ker(\mathbf{F}^T)$). If $\delta\mathbf{d} \notin \ker(\mathbf{F}^T)$, the modeling errors and noise in data are different, and noise and signal are mixed. In this paper we will only talk about $\delta\mathbf{d}$. Future research will be devoted to the distinction of both contributions.

Let us examine more in detail how noise affects the region of equivalent solutions $L_{tol} = \{\mathbf{m} \in \mathbf{M} : \|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2 \leq tol^2\}$. We have

$$\|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2 = \|\mathbf{F}\mathbf{m}\|_2^2 - 2(\mathbf{F}\mathbf{m})^T \mathbf{d}^{obs} + \|\mathbf{d}^{obs}\|_2^2. \quad (12)$$

The boundary of the L_{tol} region is the hyper-quadric containing the linear equivalent models predicting the observed data with error tolerance tol (Fernández-Martínez et al., 2012)

$$\|\mathbf{F}\mathbf{m}\|_2^2 - 2(\mathbf{F}\mathbf{m})^T \mathbf{d}^{obs} + \|\mathbf{d}^{obs}\|_2^2 - tol^2 = 0, \quad (13)$$

or

$$\|\Sigma\mathbf{m}_V\|_2^2 - 2(\Sigma\mathbf{m}_V)^T \mathbf{d}_U^{obs} + \|\mathbf{d}_U^{obs}\|_2^2 - tol^2 = 0, \quad (14)$$

with $\mathbf{d}_U^{obs} = \mathbf{U}^T \mathbf{d}^{obs}$, and $\mathbf{m}_V = \mathbf{V}^T \mathbf{m}$. In this last case the hyper-quadric is referred to the \mathbf{U} and \mathbf{V} (principal axes) orthogonal bases.

Taking into account Eqs. (13) and (14), the following can be observed:

- (i). Noise in data does not affect the matrix of the hyper-quadric, $\mathbf{F}^T \mathbf{F}$. Thus, the orientation of the linear hyper-quadric (\mathbf{V} principal axes) remains the same.
- (ii). Noise perturbs the center of the hyper-quadric, which coincides with the least squares solution $\mathbf{m}_c = \mathbf{m}^\dagger$ of the linear inverse problem stated in Eq. (9).
- (iii). Finally, noise in data deforms homogeneously the topography of the unperturbed cost function in the region of equivalent models, bounded by L_{tol} (the hyper-quadric that inspires our analysis).

Assertion (iii) can also be proved analytically. Eq. (14) can be written in the form

$$\mathbf{m}_V^T \Sigma^T \Sigma \mathbf{m}_V - 2\mathbf{m}_V^T \Sigma^T \mathbf{d}_U^{obs} = tol^2 - (\mathbf{d}_U^{obs})^T \mathbf{d}_U^{obs}. \quad (15)$$

Taking into account that $r = \text{rank}(\mathbf{F}) = \text{rank}(\Sigma)$, the left and right hand sides of Eq. (15) are respectively

$$\sum_{k=1}^r \mu_k^2 m_{V_k}^2 - 2 \sum_{k=1}^r \mu_k m_{V_k} d_{Uk}^{obs} = \sum_{k=1}^r \left[(\mu_k m_{V_k} - d_{Uk}^{obs})^2 - (d_{Uk}^{obs})^2 \right], \quad (16)$$

and

$$tol^2 - (\mathbf{d}_U^{obs})^T \mathbf{d}_U^{obs} = tol^2 - \sum_{k=1}^r (d_{Uk}^{obs})^2 - \sum_{k=r+1}^s (d_{Uk}^{obs})^2. \quad (17)$$

Using the relationships of Eqs. (16) and (17), Eq. (14) can be written as

$$\sum_{k=1}^r (\mu_k m_{V_k} - d_{Uk}^{obs})^2 = tol^2 - \sum_{k=r+1}^s (d_{Uk}^{obs})^2. \quad (18)$$

If $r = \text{rank}(\mathbf{F}) < n$, the components m_{V_k} with $k = r + 1, \dots, n$ are not determined by the linear system. Geometrically, the hyper-quadric becomes an elliptical cylinder in the directions of \mathbf{v}_k with $k = r + 1, \dots, n$ (Fernández-Martínez et al., 2012). The center of the hyper-quadric, written in the \mathbf{V} base in Eq. (18), for the model components that are informed by the data is

$$m_{ck} = \frac{d_{Uk}^{obs}}{\mu_k}, \quad k = 1, \dots, r, \quad (19)$$

while for the rest of the components, residing in the kernel of \mathbf{F} , this center does not exist: there exist several lines of centers corresponding to the valley lines of the elliptical cylinder. The lengths of the axes of the hyper-quadratic are

$$\left[\frac{\sqrt{\text{tol}^2 - \sum_{k=r+1}^s (d_{Uk}^{\text{obs}})^2}}{\mu_k}, \infty \right]_{\substack{k=1, \dots, r \\ j=r+1, \dots, n}} \quad (20)$$

If $r = \text{rank}(\mathbf{F}) = n$ (purely over-determined and even-determined linear systems) all the lengths of the axes of the hyper-quadratic are finite (hyper-ellipsoid).

Expression (20) serves to explain the contribution to the axes lengths of the data (noise included) that has not been modeled by \mathbf{F} (term $\sum_{k=r+1}^s (d_{Uk}^{\text{obs}})^2$), and the conditioning of the forward operator \mathbf{F} , depending on the decreasing behavior of its spectrum (term $1/\mu_k$). The third contribution in Eq. (20) is the error tolerance value tol , that fixes the extent of the hyper-quadratic.

Now considering $d_{Uk}^{\text{obs}} = d_{Uk}^{\text{true}} + \delta d_{Uk}$, the equation of the hyper-quadratic is

$$\sum_{k=1}^r [\mu_k m_{V_k} - (d_{Uk}^{\text{true}} + \delta d_{Uk})]^2 = \text{tol}^2 - \sum_{k=r+1}^s (d_{Uk}^{\text{true}} + \delta d_{Uk})^2, \quad (21)$$

where the model and the data (noise included) are referred respectively to \mathbf{V} and \mathbf{U} bases.

The term $\sum_{k=r+1}^s (d_{Uk}^{\text{true}} + \delta d_{Uk})^2$ in Eq. (21) is non-null for incompatible over-determined and rank-deficient linear systems. For compatible systems we have $\sum_{k=r+1}^s (d_{Uk}^{\text{true}} + \delta d_{Uk})^2 = 0$ independently of the linear system dimensions, and the noise only shifts the solution of the linear system but does not deform the topography of the equivalence region. This fact can be observed in Fig. 1, where noise has been added to a \mathbb{R}^2 squared full-rank linear system.

The hyper-quadratic for the noise-free case is

$$\sum_{k=1}^r (\mu_k m_{V_k} - d_{Uk}^{\text{true}})^2 = \text{tol}^2 - \sum_{k=r+1}^s (d_{Uk}^{\text{true}})^2. \quad (22)$$

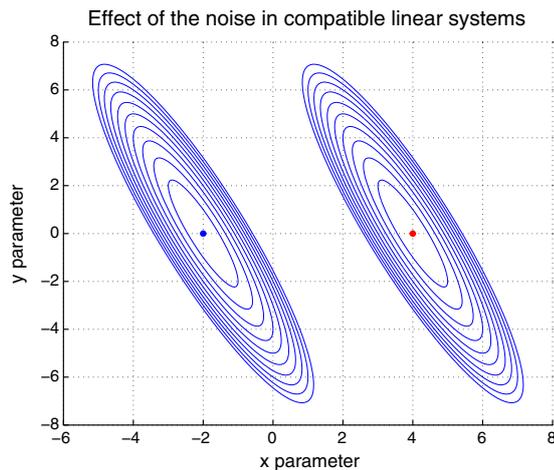


Fig. 1. Case of compatible systems. The addition of noise provokes the shifting of the equivalence region when the perturbed linear system is also compatible. No deformation of this region is induced by the noise.

The following condition has to be fulfilled for both hyper-quadratics (Eqs. (21) and (22)) to have the same axes lengths

$$\begin{aligned} \text{tol}^* &= \sqrt{\text{tol}^2 + \sum_{k=r+1}^s [(d_{Uk}^{\text{true}} + \delta d_{Uk})^2 - (d_{Uk}^{\text{true}})^2]} \\ &= \sqrt{\text{tol}^2 + \sum_{k=r+1}^s (\delta d_{Uk}^2 + 2\delta d_{Uk} d_{Uk}^{\text{true}})}. \end{aligned} \quad (23)$$

Eq. (23) could also be used to find the value of tol^* assuming that $\mathbf{d}^{\text{true}} \approx \mathbf{d}^{\text{fitted}}$ coincides with the data predicted by the least squares solution. In this case

$$\sum_{k=r+1}^s [(d_{Uk}^{\text{true}} + \delta d_{Uk})^2 - (d_{Uk}^{\text{true}})^2] \approx \sum_{k=r+1}^s [(d_{Uk}^{\text{obs}})^2 - (d_{Uk}^{\text{fitted}})^2]. \quad (24)$$

Calling $\Delta \text{tol} = \sum_{k=r+1}^s (\delta d_{Uk}^2 + 2\delta d_{Uk} d_{Uk}^{\text{true}})$, we have

- (i). If $\Delta \text{tol} > 0$, then $\text{tol}^* > \text{tol}$, that is, in the presence of noise the region with the same size corresponds to a region of higher misfit. Thus, the regions of equivalent models will decrease in size. Particularly this will happen if $\sum_{k=r+1}^s \delta d_{Uk} d_{Uk}^{\text{true}} = 0$ as, for instance, in presence of white noise. In this case locating the equivalence region of low misfits ($\text{tol} \leq 10\%$ for instance) by means of global optimization methods becomes a harder task due to its smaller size. Conversely, finding the global optimum of the perturbed cost function via local optimization methods inside this region is easier. Although the noise $\delta \mathbf{d}$ does not affect the conditioning of the forward operator \mathbf{F} , it could be considered as a kind of regularization since for the same value of the error tolerance the axes of the hyper-quadratic become smaller. Finally when using global optimization methods under the presence of noise, the posterior search is usually performed in a region of higher misfit than the one corresponding to the noise-free case. This fact contributes to give the impression that finding the regions of medium misfit becomes easier under the presence of noise, when in fact these regions do decrease in size.
- (ii). Conversely, if $\Delta \text{tol} < 0$, then $\text{tol}^* < \text{tol}$, and the region with the same size in presence of noise corresponds to a region of lower misfit. In this case, the regions of equivalent models will increase in size. This fact could be used to locate more easily the low misfit region by artificially adding an adequate type of colored noise to the observed data.

The question resides now in knowing the sign of Δtol . The following analysis might help. We have

$$\|\mathbf{d}_U^{\text{obs}}\|_2^2 = \|\mathbf{d}_U^{\text{true}} + \delta \mathbf{d}_U\|_2^2 = \|\mathbf{d}_U^{\text{true}}\|_2^2 + \|\delta \mathbf{d}_U\|_2^2 + 2\delta \mathbf{d}_U^T \mathbf{d}_U^{\text{true}}, \quad (25)$$

or equivalently

$$\|\mathbf{d}^{\text{obs}}\|_2^2 = \|\mathbf{d}^{\text{true}} + \delta \mathbf{d}\|_2^2 = \|\mathbf{d}^{\text{true}}\|_2^2 + \|\delta \mathbf{d}\|_2^2 + 2\delta \mathbf{d}^T \mathbf{d}^{\text{true}}. \quad (26)$$

Thus, in general

$$\|\mathbf{d}^{\text{obs}}\|_2^2 > \|\mathbf{d}^{\text{true}}\|_2^2 \Leftrightarrow \|\delta \mathbf{d}\|_2^2 + 2\delta \mathbf{d}^T \mathbf{d}^{\text{true}} \geq 0. \quad (27)$$

Condition (27) applies in the following cases

- (iii). Noise is uncorrelated to the true data: $\delta \mathbf{d}^T \mathbf{d}^{\text{true}} = 0$.
- (iv). Noise is positively correlated to the true data: $\delta \mathbf{d}^T \mathbf{d}^{\text{true}} > 0$.
- (v). Although noise is negatively correlated to the true data, we have that $\|\delta \mathbf{d}\|_2^2 \geq -2\delta \mathbf{d}^T \mathbf{d}^{\text{true}}$.

Taking into account this reasoning it can be concluded that in most of the cases $\Delta tol > 0$, although the other possibility might also happen in practice.

In conclusion, in linear inverse problems noise deforms homogeneously the topography of the cost function, shifting the center of the linear region of equivalence (hyper-quadric) that coincides with the least squares solution. The term homogeneously refers to the fact that the deformation does not depend on the model \mathbf{m}_0 that is considered in this analysis. Besides, in the case of compatible systems, the noise does not provoke deformation, only shifts the location of the noise-free solution.

3.1. Synthetic example

To show the previous conclusions numerically, we have designed a synthetic example for a linear regression problem.

For that purpose we had generated a data set $\{(x_1, y_1), (x_2, y_2), \dots, (x_s, y_s)\}$, according to the theoretical model $y_k = \alpha x_k + \beta + \varepsilon$, where $\alpha = 4$ and $\beta = 2$. We have added for the first test a Gaussian random variable $\varepsilon \rightarrow N(0, 0.05)$, trying to mimic the noise in the observed data during measurement. For the second test we have added to the previous one an extra noise term $\delta \rightarrow N(0, 0.025)$.

Fig. 2 shows the contour of 9% relative misfit corresponding to the topography cost function in the noise-free case, which is centered at (α, β) . We also show the same contour line corresponding to the topography of the cost function in the first case (noise $\varepsilon \rightarrow N(0, 0.05)$). This contour line is centered in the neighborhood of the least squares solution $\mathbf{m}^\dagger = (4.25, 1.68)$. It can be observed that it corresponds to the linear hyper-quadric stated in Eq. (13) and the perturbed solution is shifted from the noise-free hyper-quadric. Finally in the same figure we show the same contour line of 9% tolerance region in the second case (noise $\varepsilon \rightarrow N(0, 0.05) + \delta \rightarrow N(0, 0.025)$). Now, the least squares solution shifted to $\alpha_1 = 3.57, \beta_1 = 2.68$, located further away from the true solution and the region of equivalence shrinks further. The original 9% equivalent region corresponds in size approximately to the regions of 10.31% and 11.66% relative errors for both cases (5% and 7.5% relative noise). The analytical formula used to perform this computation was Eq. (18).

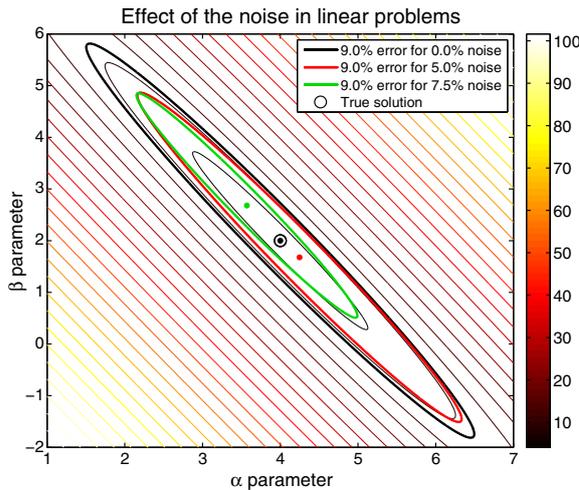


Fig. 2. Synthetic linear inverse problem. The 9% equivalence regions for noise-free problem, and also adding 5% and 7.5% levels of Gaussian noise to the original data are shown.

Although the numerical example treats the simpler regression problem, the conclusions from this analysis are valid to any kind of inverse problem in higher dimensions, such as the linear tomography problem or the gravimetric inverse problem.

In the next section we analyze the effect of the zero order Tikhonov's regularization in linear inverse problems combined with noise, showing that noise and regularization have similar effects.

4. How regularization acts in linear inverse problems

Regularization refers to the mathematical techniques that are used to stabilize the inversion in linear and nonlinear problems. In this section we analyze the role of the regularization and its effect on the topography of the cost function, showing geometrically its effect on the solution that has been found. The aim is to expand the interest of this analysis beyond the need of regularization for least squares solutions.

Regularization in linear inverse problems can be performed by truncation and by damping. Truncation consists in performing hard thresholding of the smallest singular values of \mathbf{F}

$$\mu_k^\dagger = \begin{cases} \mu_k & \text{if } \mu_k > \mu_{\min}, \\ 0 & \text{if } \mu_k \leq \mu_{\min}. \end{cases} \quad (28)$$

Then the solution of the linear system in the \mathbf{V} base is $\left(\frac{d_{jk}^{obs}}{\mu_k}, 0_j\right)_{k=1, \dots, t}$, where t is the number of singular values that $j = t + 1, \dots, n$

have not been threshold. The solution is now

$$\mathbf{m}_t^\dagger = \sum_{k=1}^{t \leq r} \frac{d_{jk}^{obs}}{\mu_k} \mathbf{v}_k = \mathbf{V} \Sigma_t^\dagger \mathbf{U}^T \mathbf{d}^{obs}, \quad (29)$$

where Σ_t^\dagger stands for the Moore–Penrose's pseudo-inverse after the thresholding is performed. The truncation serves to avoid the spurious effect of noise in data due to the smallest singular values of \mathbf{F} . This is done at expenses of increasing the dimension of the “numerical” kernel of \mathbf{F} that is now spanned by the vectors $(\mathbf{v}_{t+1}, \dots, \mathbf{v}_n)$. It is interesting to remark that it is not guaranteed that the minimum norm solution found in this case would have a physical meaning, since noise also affects the components d_{jk}^{obs} corresponding to the singular values that have not been threshold. This fact is usually observed in practice when the Moore–Penrose's pseudo-inverse is used.

The second strategy is performed via prior information and damping. Zero order Tikhonov's regularization consists in minimizing a linear combination of the data prediction misfit and the distance to a reference model, called \mathbf{m}^{ref}

$$\min_{\mathbf{m} \in \mathbb{R}^n} \left[\|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2 \right]. \quad (30)$$

In a Bayesian context, this is equivalent to consider an \mathbf{m}^{ref} -mean normally distributed prior on the parameter vector \mathbf{m} . Now, the regularized least squares solution is (see for instance (Aster et al., 2012))

$$\mathbf{m}^\varepsilon = \mathbf{m}^{ref} + (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n)^{-1} \mathbf{F}^T (\mathbf{d}^{obs} - \mathbf{F}\mathbf{m}^{ref}). \quad (31)$$

Introducing the singular value decomposition of \mathbf{F} we arrive at

$$\mathbf{m}^\varepsilon = \mathbf{m}^{ref} + \mathbf{V} (\Sigma^T \Sigma + \varepsilon^2 \mathbf{I}_n)^{-1} \Sigma^T \mathbf{U}^T (\mathbf{d}^{obs} - \mathbf{F}\mathbf{m}^{ref}), \quad (32)$$

that is

$$\mathbf{m}_V^\varepsilon = \mathbf{m}_V^{ref} + (\Sigma^T \Sigma + \varepsilon^2 \mathbf{I}_n)^{-1} \Sigma^T \Delta \mathbf{d}_U^{ref}, \quad (33)$$

where the models are referred to the \mathbf{V} base, and $\Delta \mathbf{d}_U^{ref}$ is the reference model data misfit referred to the \mathbf{U} base

$$\Delta \mathbf{d}_U^{ref} = \mathbf{U}^T (\mathbf{d}^{obs} - \mathbf{Fm}^{ref}) = \mathbf{d}_U^{obs} - \Sigma \mathbf{m}_V^{ref}. \tag{34}$$

Then, the coordinates of the regularized solution \mathbf{m}^ϵ in the \mathbf{V} base are

$$\begin{aligned} \left(m_{V_k}^{ref} + \frac{\mu_k}{\mu_k^2 + \epsilon^2} \Delta d_{Uk}^{ref}, m_{V_j}^{ref} \right)_{\substack{k=1, \dots, r \\ j=r+1, \dots, n}} &= \\ = \left(\frac{\epsilon^2 m_{V_k}^{ref} + \mu_k d_{Uk}^{obs}}{\mu_k^2 + \epsilon^2}, m_{V_j}^{ref} \right)_{\substack{k=1, \dots, r \\ j=r+1, \dots, n}}. \end{aligned} \tag{35}$$

Finally, the regularized solution \mathbf{m}^ϵ can be written as follows

$$\mathbf{m}^\epsilon = \sum_{k=1}^r \frac{\epsilon^2 m_{V_k}^{ref} + \mu_k d_{Uk}^{obs}}{\mu_k^2 + \epsilon^2} \mathbf{v}_k + \sum_{j=r+1}^n m_{V_j}^{ref} \mathbf{v}_j. \tag{36}$$

Now all the vectors in the \mathbf{V} base inform the regularized solution \mathbf{m}^ϵ , and

1. The coordinates in the r first vectors of the \mathbf{V} base are linear combinations between the reference model coordinates $m_{V_k}^{ref}$, and the observed data components d_{Uk}^{obs} . The weights of this linear combination are respectively $\frac{\epsilon^2}{\mu_k^2 + \epsilon^2}$ and $\frac{\mu_k}{\mu_k^2 + \epsilon^2}$.
2. The coordinates of \mathbf{m}^\dagger , that originally resided in the kernel of \mathbf{F} , are now informed by the reference model coordinates $m_{V_j}^{ref}$.

Fig. 3 illustrates this fact in two dimensions. As it can be observed the model \mathbf{m}^\dagger provided by the Moore–Penrose’s pseudo-inverse has no components on the kernel of the forward operator \mathbf{F} . For that reason this solution is called, the natural or simplest solution. The reference model serves to incorporate components of the kernel of \mathbf{F} into the inversion. It is also interesting to remark that the Levenberg–Marquardt’s solution of the linear inverse problem (obtained with $\mathbf{m}^{ref} = 0$) does not incorporate components from the kernel but has a greater norm than the Moore–Penrose’s pseudo-inverse solution.

The hyper-quadric of equivalence for the regularized solution is

$$\|\mathbf{Fm} - \mathbf{d}^{obs}\|_2^2 + \epsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2 = tol^2, \tag{37}$$

that is

$$\begin{aligned} \|\mathbf{Fm}\|_2^2 - 2(\mathbf{Fm})^T \mathbf{d}^{obs} + \|\mathbf{d}^{obs}\|_2^2 \\ + \epsilon^2 (\|\mathbf{m}\|_2^2 - 2\mathbf{m}^T \mathbf{m}^{ref} + \|\mathbf{m}^{ref}\|_2^2) - tol^2 = 0. \end{aligned} \tag{38}$$

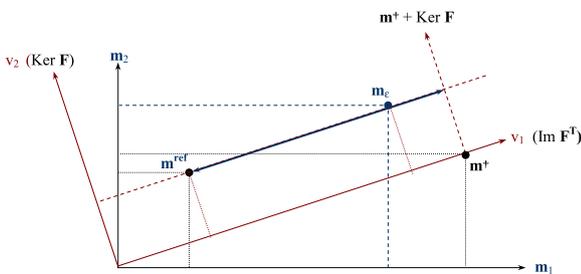


Fig. 3. Geometric interpretation for the regularization. The solution \mathbf{m}^\dagger has no components in kernel of \mathbf{F} , but \mathbf{m}^ϵ incorporates the reference model coordinates to the model components that lie in the kernel of \mathbf{F} .

Referred to the \mathbf{V} base the hyper-quadric can be written as follows

$$\begin{aligned} \mathbf{m}_V^T \Sigma^T \Sigma \mathbf{m}_V - 2\mathbf{m}_V^T \Sigma^T \mathbf{d}_U^{obs} + (\mathbf{d}_U^{obs})^T \mathbf{d}_U^{obs} \\ + \epsilon^2 [\mathbf{m}_V^T \mathbf{m}_V - 2\mathbf{m}_V^T \mathbf{m}_V^{ref} + (\mathbf{m}_V^{ref})^T \mathbf{m}_V^{ref}] = tol^2, \end{aligned} \tag{39}$$

where we have taken into account that orthogonal transformations conserve the norm of vectors. Eq. (39) can be written as

$$\begin{aligned} \mathbf{m}_V^T (\Sigma^T \Sigma + \epsilon^2 \mathbf{I}) \mathbf{m}_V - 2\mathbf{m}_V^T (\Sigma^T \mathbf{d}_U^{obs} + \epsilon^2 \mathbf{m}_V^{ref}) = \\ = tol^2 - (\mathbf{d}_U^{obs})^T \mathbf{d}_U^{obs} - \epsilon^2 (\mathbf{m}_V^{ref})^T \mathbf{m}_V^{ref}. \end{aligned} \tag{40}$$

Expanding Eq. (40) we arrive at the final analytical equation for the linear hyper-quadric in the regularized case

$$\begin{aligned} \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \epsilon^2} m_{V_k} - \frac{\epsilon^2 m_{V_k}^{ref} + \mu_k d_{Uk}^{obs}}{\sqrt{\mu_k^2 + \epsilon^2}} \right)^2 \\ + \epsilon^2 \sum_{k=r+1}^n (m_{V_k} - m_{V_k}^{ref})^2 = tol^2 + \sum_{k=1}^r \frac{(\epsilon^2 m_{V_k}^{ref} + \mu_k d_{Uk}^{obs})^2}{\mu_k^2 + \epsilon^2} \\ - \|\mathbf{d}_U^{obs}\|_2^2 - \epsilon^2 \sum_{k=1}^r (m_{V_k}^{ref})^2. \end{aligned} \tag{41}$$

Now, calling

$$\Delta_r = \sum_{k=1}^r \frac{(\epsilon^2 m_{V_k}^{ref} + \mu_k d_{Uk}^{obs})^2}{\mu_k^2 + \epsilon^2} - \|\mathbf{d}_U^{obs}\|_2^2 - \epsilon^2 \sum_{k=1}^r (m_{V_k}^{ref})^2, \tag{42}$$

the axes of the hyper-quadric are

$$\left[\frac{\sqrt{tol^2 + \Delta_r}}{\sqrt{\mu_i^2 + \epsilon^2}}, \left(\frac{\sqrt{tol^2 + \Delta_r}}{\epsilon} \right)_j \right]_{\substack{i=1, \dots, r \\ j=r+1, \dots, n}}, \tag{43}$$

instead of

$$\left(\frac{\sqrt{tol^2 - \|\mathbf{d}_U^{obs}\|_2^2 + \sum_{k=1}^r (d_{Uk}^{obs})^2}}{\mu_i}, \infty_j \right)_{\substack{i=1, \dots, r \\ j=r+1, \dots, n}} \tag{44}$$

with no regularization.

The regularization with a model of reference \mathbf{m}^{ref} prescribes the center of the hyper-quadric to $m_{V_j}^{ref}$ for the coordinates of the solution that originally resided on the kernel of \mathbf{F} , and bounds the axes in these \mathbf{v}_j directions to $\sqrt{tol^2 + \Delta_r}/\epsilon$, instead of infinity. Accordingly, the equivalent region changes from an elliptical cylinder (with no regularization) to a very oblong hyper-ellipsoid in the directions of the kernel of \mathbf{F} .

The following relationship applies for both equivalent regions of tolerance tol^* (with regularization) and tol (without regularization) to have the same size (or axes)

$$\frac{\sqrt{tol^{*2} + \Delta_r}}{\sqrt{\mu_i^2 + \epsilon^2}} = \frac{\sqrt{tol^2 - \sum_{k=r+1}^s (d_{Uk}^{obs})^2}}{\mu_i}, \quad i = 1, \dots, r, \tag{45}$$

182

J.L. Fernández-Martínez et al. / Journal of Applied Geophysics 108 (2014) 176–185

thus

$$tol^r = \sqrt{\frac{\mu_i^2 + \varepsilon^2}{\mu_i^2} \left[tol^2 - \sum_{k=r+1}^s (d_{Uk}^{obs})^2 \right]} - \Delta_r. \quad (46)$$

Finally, taking into account Eq. (42), we arrive at

$$tol^r = \sqrt{\frac{\mu_i^2 + \varepsilon^2}{\mu_i^2} tol^2 + \varepsilon^2 \sum_{k=1}^r \frac{(d_{Uk}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2} - \frac{\varepsilon^2}{\mu_i^2} \sum_{k=r+1}^s (d_{Uk}^{obs})^2}, \quad (47)$$

which can be rewritten as

$$tol^r(\mu_i) = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left[tol^2 - \sum_{k=r+1}^s (d_{Uk}^{obs})^2 \right] + \varepsilon^2 \sum_{k=1}^r \frac{(d_{Uk}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}}. \quad (48)$$

In this case $tol^*(\mu_i)$ depends on the axes index i , due to the effect of the regularization. Also, we have $tol^*(\mu_i) > tol$ if

$$tol^2 > \sum_{k=r+1}^s (d_{Uk}^{obs})^2 - \mu_i^2 \sum_{k=1}^r \frac{(d_{Uk}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}, \quad (49)$$

that is, for any region with tolerance $tol > tol_{ci}(\mu_i)$, where

$$tol_{ci}(\mu_i) = \sqrt{\sum_{k=r+1}^s (d_{Uk}^{obs})^2 - \mu_i^2 \sum_{k=1}^r \frac{(d_{Uk}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}}. \quad (50)$$

The regions with error tolerance $tol > tol_{ci}(\mu_i)$ decrease in size due to the effect of the regularization. In this case models have to be found in larger areas, deforming differently the axes of the region of equivalence according to Eq. (48). Conversely, $tol^*(\mu_i) < tol$ for any tolerance region with $tol < tol_{ci}(\mu_i)$. These equivalent regions increase in size with respect to the unregularized case. The value of $tol_{ci}(\mu_i)$ depends on each singular value μ_i and on the value of $d_m = \sum_{k=r+1}^s (d_{Uk}^{obs})^2$. In the case of purely under-determined systems we have $r = s \Rightarrow d_m = 0$, and $tol_{ci}(\mu_i)$ does not

exist, becoming a complex number. Taking into account Eq. (48), we arrive at $tol^*(\mu_i) > tol$, that is, the regularization decreases the size of all equivalent regions. This situation also happens for over-determined linear systems when $d_m = \sum_{k=r+1}^s (d_{Uk}^{obs})^2 \approx 0$, that is, the system is “almost” compatible, since compatibility implies $d_{Uk}^{obs} = 0, \forall k = r+1, \dots, s$. Also due to the fact that tol_{ci} depends on μ_i , it might happen that $tol_{ci}(\mu_i)$ does not exist for large values of μ_i , that correspond to the directions of the smallest uncertainty (model variability). Thus, the regularization only affects the axes of the linear equivalence region (hyper-quadric) corresponding to the smallest singular values, deforming the region of linear equivalence in an anisotropic manner. If the reference model predicts the

observed data, that is, $d_{Uk}^{obs} = \mu_k m_{V_k}^{ref}, 1, \dots, r$, then $tol_{ci}(\mu_i) = \sqrt{\sum_{k=r+1}^s (d_{Uk}^{obs})^2}$.

Fig. 4 shows the effect of the regularization in the previous synthetic case with noise of 5%, adopting a reference model $\mathbf{m}^{ref} = (3.5, 2.5)$, and $\varepsilon^2 = 0.0$ and $\varepsilon^2 = 0.05$. The 9% error contour line in the regularized case is shown, compared to the original case without regularization. For very small values of the regularization parameter, this effect is almost negligible. In this synthetic case the tol_{ci} values are complex numbers, that is, the regularization decreases the size of the equivalent region.

Following the same methodology, it is possible to analyze the combined effect of noise and regularization with respect to the noise-free case. Comparing the equivalent regions of tolerance tol^* (for the noisy case with regularization) and tol (for noise-free case without regularization), we have that

$$\frac{\sqrt{tol^{*2} + \Delta_r}}{\sqrt{\mu_i^2 + \varepsilon^2}} = \frac{\sqrt{tol^2 - \sum_{k=r+1}^s (d_{Uk}^{true})^2}}{\mu_i}, \quad (51)$$

where Δ_r is given by Eq. (42). Then, we arrive at

$$tol^*(\mu_i) = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left[tol^2 - \sum_{k=r+1}^s (d_{Uk}^{true})^2 \right] + \Lambda}, \quad (52)$$

where

$$\Lambda = \sum_{k=r+1}^s \left[(d_{Uk}^{obs})^2 - (d_{Uk}^{true})^2 \right] + \varepsilon^2 \sum_{k=1}^r \frac{(d_{Uk}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}. \quad (53)$$

Similar conclusions to the above mentioned case can be achieved. In this case we have:

$$tol^*(\mu_i) = \sqrt{\sum_{k=r+1}^s (d_{Uk}^{true})^2 - \frac{\mu_i^2}{\varepsilon^2} \sum_{k=r+1}^s \left[(d_{Uk}^{obs})^2 - (d_{Uk}^{true})^2 \right] - \mu_i^2 \sum_{k=1}^r \frac{(d_{Uk}^{obs} - \mu_k m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}}. \quad (54)$$

Finally, it is possible to compare the equivalent regions for the regularized solution, with and without noise. Imposing that the linear equivalence regions will have the same axes lengths, we arrive at

$$\frac{\sqrt{tol^{*2} + \Delta_r^{obs}}}{\sqrt{\mu_i^2 + \varepsilon^2}} = \frac{\sqrt{tol^2 + \Delta_r^{true}}}{\sqrt{\mu_i^2 + \varepsilon^2}}, \quad (55)$$

that is

$$tol^* = \sqrt{tol^2 + \Delta_r^{true} - \Delta_r^{obs}}, \quad (56)$$

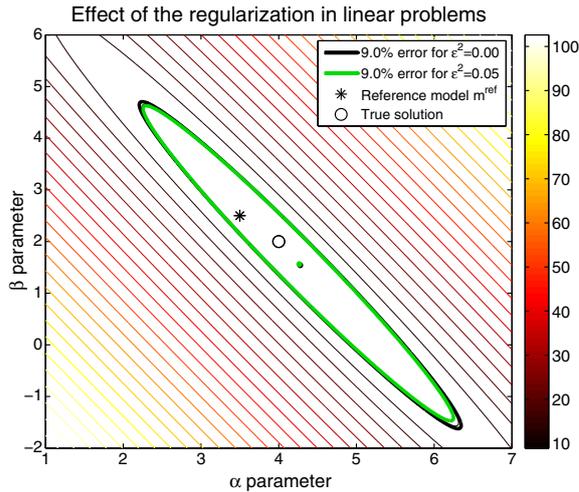


Fig. 4. Effect of the regularization in the previous synthetic case (Fig. 2) with a reference model $\mathbf{m}^{ref} = (3.5, 2.5)$, 5% of Gaussian noise, and damping parameters $\varepsilon^2 = 0.0$ and $\varepsilon^2 = 0.05$. The regions of 9% relative error for the non-regularized and regularized cases are shown.

with

$$\Delta^{true} = \sum_{k=1}^r \frac{(\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{true})^2}{\mu_k^2 + \varepsilon^2} - \|\mathbf{d}_U^{true}\|_2^2 - \varepsilon^2 \sum_{k=1}^r (m_{V_k}^{ref})^2, \quad (57)$$

and

$$\Delta^{obs} = \sum_{k=1}^r \frac{(\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs})^2}{\mu_k^2 + \varepsilon^2} - \|\mathbf{d}_U^{obs}\|_2^2 - \varepsilon^2 \sum_{k=1}^r (m_{V_k}^{ref})^2. \quad (58)$$

Performing some algebraic operations we finally have

$$tol^* = \sqrt{tol^2 + \sum_{k=r+1}^s \delta d_{U_k} (\delta d_{U_k} + 2d_{U_k}^{true}) + \varepsilon^2 \sum_{k=1}^r \frac{\delta d_{U_k} (\delta d_{U_k} - 2\mu_k m_{V_k}^{ref})}{\mu_k^2 + \varepsilon^2}}. \quad (59)$$

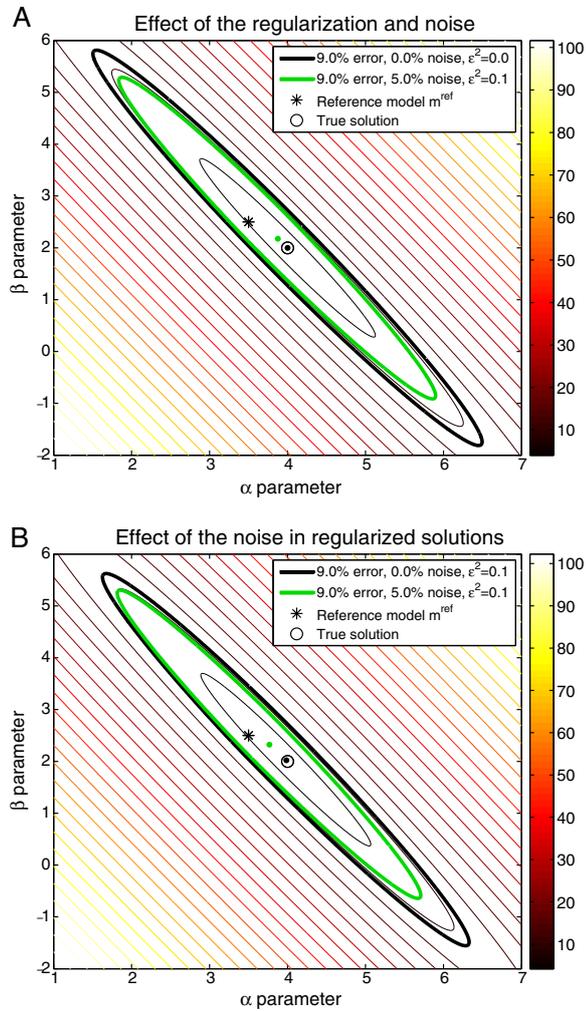


Fig. 5. A) Combined effect of the noise and the regularization with respect to the noise-free and non-regularized cases. B) Comparison of the equivalent regions for the noise-free regularized and the regularized noisy cases.

In this case, tol^* does not depend on the model component, or hyper-quadric axis index. Again, similar considerations to the former case can be done depending on the sign of $\Delta^{true} - \Delta^{obs}$.

Fig. 5 shows these two last comparisons for the same synthetic example mentioned above. Fig. 5-A shows the 9% tolerance regions for the noise-free non-regularized case, and adding a Gaussian noise $\delta \rightarrow N(0,0.05)$ to the regularized case, with a reference model \mathbf{m}^{ref} and a damping parameter $\varepsilon^2 = 0.01$. It can be observed that the region of equivalence in the presence of noise and with regularization decreases in size with respect to the noise-free non-regularized case. The values of tol_k^2 are negative in this particular case ($-4.72 \cdot 10^4$ and $-0.08 \cdot 10^4$). The values of $tol^*(\mu_k)$ are 10.12% and 11.27%, corresponding respectively to both singular values of the forward operator ($\mu_1 = 18.28, \mu_2 = 1.61$). Fig. 5-B compares the same regions of equivalence (9% error tolerance) for the regularized case ($\varepsilon^2 = 0.01$) with 5% Gaussian noise, with respect to the noise-free. In this case the region with noise decreases in size, since $\Delta^{true} - \Delta^{obs} = 15.67$, and $tol^* = 10.24\%$.

5. Other regularization techniques

First and second order Tikhonov's regularizations are also frequently used in practice (see for instance (Constable et al., 1987)) and can be written as

$$\min_{\mathbf{m} \in \mathbb{R}^n} \left[\|\mathbf{Fm} - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{\Gamma}(\mathbf{m} - \mathbf{m}^{ref})\|_2^2 \right], \quad (60)$$

where $\mathbf{\Gamma}$ is a high-pass difference operator (gradient/Laplacian) used to enforce smoothness in the solution (first and second order model derivatives). The region of equivalence is in this case

$$\|\mathbf{Fm}\|_2^2 - 2(\mathbf{Fm})^T \mathbf{d}^{obs} + \|\mathbf{d}^{obs}\|_2^2 + \varepsilon^2 (\mathbf{m} - \mathbf{m}^{ref})^T \mathbf{\Gamma}^T \mathbf{\Gamma} (\mathbf{m} - \mathbf{m}^{ref}) - tol^2 = 0. \quad (61)$$

This expression generalizes the case where $\mathbf{\Gamma} = \mathbf{I}_n$. Matrix $\mathbf{\Gamma}$ has the effect of introducing components from the kernel of \mathbf{F} into the solution, even if no model of reference is used.

Finally, Tibshirani (1996) introduced the Lasso (Least absolute shrinkage and selection operator) method that uses the L_1 norm on the parameter regularization

$$\min_{\mathbf{m} \in \mathbb{R}^n} \left[\|\mathbf{Fm} - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_1 \right]. \quad (62)$$

This problem is equivalent to an unconstrained minimization of the least-squares with a penalty $\varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_1$ added. In a Bayesian context this implies to use a \mathbf{m}^{ref} -mean Laplace prior distribution on the model parameters. The corresponding optimization problem is nonlinear and is usually solved through convex optimization methods.

One of the main differences between Lasso and zero-order Tikhonov's regularization (with no model of reference) is that as the penalty is increased, the L_1 regularization will cause more and more parameters of the solution to be driven to zero (or to \mathbf{m}^{ref} in the general case), while in the case of Tikhonov's regularization that is associated to the L_2 norm, although these parameters are reduced they still remain non-null. For that reason the L_1 -regularized formulation of the least squares is useful in contexts where sparse solutions (solutions with fewer nonzero parameters) are preferred in order to reduce the number of variables upon which the given solution depends.

Fig. 6 shows numerically for the above mentioned synthetic problem the difference between the regularized equivalent regions for the Lasso and Tikhonov's regularizations as the damping parameter increases. The reference model used was the same as that in the previous example ($\mathbf{m}^{ref} = (3.5, 2.5)$). It can be observed that the Lasso equivalence region is always smaller than the one corresponding to the Tikhonov's regularization. If the damping parameter ε^2 is increased from 0.25 to 5, the equivalence region loses its elliptical shape to adapt progressively to

184

J.L. Fernández-Martínez et al. / Journal of Applied Geophysics 108 (2014) 176–185

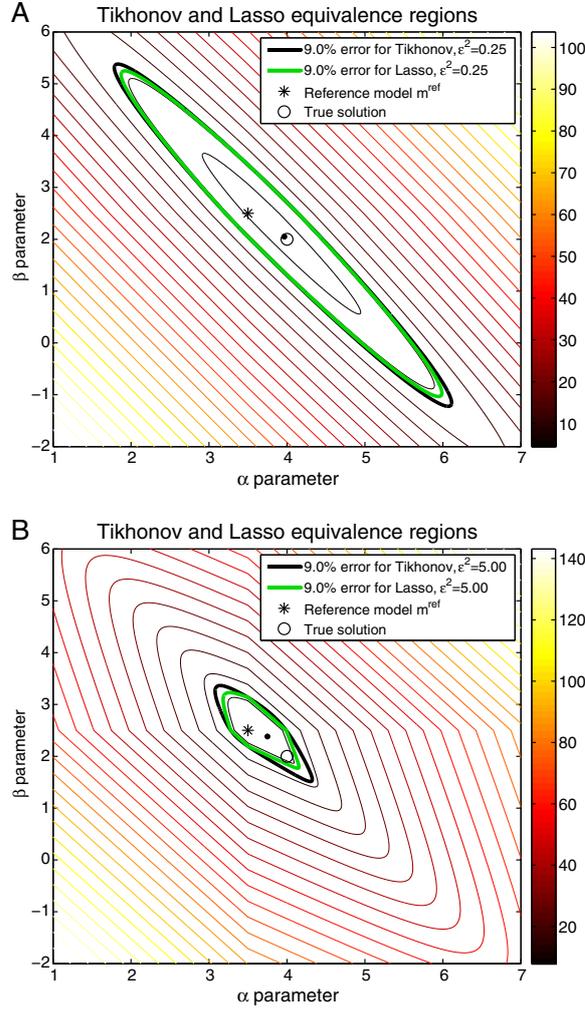


Fig. 6. Comparison between the L_2 (Tikhonov) and L_1 (Lasso) equivalent regions for increasing values of the damping parameter. A) $\varepsilon^2 = 0.25$, B) $\varepsilon^2 = 5$. In both cases the reference model used was $\mathbf{m}^{\text{ref}} = (3.5, 2.5)$, and the background corresponds to the topography of the Lasso cost function.

the L_1 ball. Also, it is interesting to note that the cost function topography will be less affected by noise in data if a L_1 norm is used in the prediction error, since it corresponds to more robust norms. This is not the Lasso case, where the L_1 norm is only used in the regularization term.

6. Conclusions

In this paper we have analyzed the effects of the noise and that of the regularization in the solution of linear discrete inverse problems. Noise is an important ingredient in inverse problems, since its effect enters into the cost function through the observed data. Noise in data shifts the solution found by least squares methods, which is the center of the hyper-quadric of equivalence, and deforms homogeneously the topography of the cost function, in the sense that all the regions of equivalence either increase or decrease in size for different values of the error tolerance. A particular case occurs for compatible linear systems, where noise only shifts the solution that has been found and no deformation takes place. In any case, noise does not affect the orientation of the linear hyper-quadric of equivalence.

We have also provided a geometrical interpretation of the role of the regularization. The regularization serves to limit the sizes of the axes of the hyper-quadric in the directions corresponding to the null space of the forward operator. Geometrically, the original hyper-quadric which is an elliptical cylinder becomes an ellipsoid with high eccentricity. Its center coincides with the regularized solution \mathbf{m}^{ε} of the linear inverse problem. The regularization also deforms the region of equivalence in an anisotropic manner, acting differently in each component of the model solution. Reference models serve to inform the solution in the model components that were initially located on the null space of the forward operator, and to interpolate the rest of the components of the model that resided in the column space of the linear adjoint operator $\text{Im}\mathbf{F}^T$. Although the regularization has a positive effect in decreasing the uncertainty on the solution that has been adopted, a family of equivalent models still exists in the regularized case, forming the regularized hyper-quadric, whose analytical expression is known. Similar considerations can be made in the case of Lasso regularization.

This analysis is generalized to nonlinear inversion in an accompanying paper, and although theoretical at this stage, it might impact in the future how the uncertainty analysis is performed in geophysical inversion, since noise in data is always present and good prior models are not always available.

Acknowledgment

The authors would like to acknowledge the editor of the Journal of Applied Geophysics, Professor Klaus Holliger, for his support in understanding the need of theoretical analysis in geophysical inversion, and also the feedback of anonymous referees, that helped us to improve the quality of these manuscripts.

Appendix A. Additional formulas

The left hand of Eq. (40) can be split into 3 parts

1. $\mathbf{m}_V^T (\Sigma^T \Sigma + \varepsilon^2 \mathbf{I}) \mathbf{m}_V = \sum_{k=1}^r (\mu_k^2 + \varepsilon^2) m_{V_k}^2 + \varepsilon^2 \sum_{k=r+1}^n m_{V_k}^2$,
2. $-2\mathbf{m}_V^T \Sigma^T \mathbf{d}_j^{\text{obs}} = -2 \sum_{k=1}^r \mu_k d_{Uk}^{\text{obs}} m_{V_k}$,
3. $-2\varepsilon^2 \mathbf{m}_V^T \mathbf{m}_V^{\text{ref}} = -2\varepsilon^2 \sum_{k=1}^r m_{V_k} m_{V_k}^{\text{ref}} - 2\varepsilon^2 \sum_{k=r+1}^n m_{V_k} m_{V_k}^{\text{ref}}$,

where we have taken into account the fact that $r = \text{rank}(\mathbf{F}) = \text{rank}(\Sigma)$.

It is possible to group these terms into two different kinds of sums

A) For $k = 1$ to r

$$\sum_{k=1}^r (\mu_k^2 + \varepsilon^2) m_{V_k}^2 - 2 \sum_{k=1}^r m_{V_k} (\mu_k d_{Uk}^{\text{obs}} + \varepsilon^2 m_{V_k}^{\text{ref}}), \quad (\text{A.1})$$

that is

$$\sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} m_{V_k} - \frac{\mu_k d_{Uk}^{\text{obs}} + \varepsilon^2 m_{V_k}^{\text{ref}}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 - \sum_{k=1}^r \left(\frac{\mu_k d_{Uk}^{\text{obs}} + \varepsilon^2 m_{V_k}^{\text{ref}}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2, \quad (\text{A.2})$$

B) For $k = r + 1$ to n

$$\varepsilon^2 \sum_{k=r+1}^n (m_{V_k} - m_{V_k}^{\text{ref}})^2 - \varepsilon^2 \sum_{k=r+1}^n (m_{V_k}^{\text{ref}})^2. \quad (\text{A.3})$$

Then, the right hand side of Eq. (40) can be expressed as

$$\text{tol}^2 = \|\mathbf{d}_U^{\text{obs}}\|_2^2 - \varepsilon^2 \sum_{k=1}^r (m_{V_k}^{\text{ref}})^2 - \varepsilon^2 \sum_{k=r+1}^n (m_{V_k}^{\text{ref}})^2. \quad (\text{A.4})$$

Considering the relationships of Eqs. (A.2), (A.3) and (A.4), Eq. (40) becomes

$$\begin{aligned} & \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} m_{V_k} - \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \varepsilon^2 \sum_{k=r+1}^n (m_{V_k} - m_{V_k}^{ref})^2 = \\ & = tol^2 + \sum_{k=1}^r \frac{(\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs})^2}{\mu_k^2 + \varepsilon^2} - \| \mathbf{d}_U^{obs} \|_2^2 - \varepsilon^2 \sum_{k=1}^r (m_{V_k}^{ref})^2, \end{aligned} \quad \text{A.5}$$

which is the final analytical equation for the linear hyper-quadratic in the regularized case.

Appendix B. Case of weighted least squares

Let us analyze the case

$$\min [(\mathbf{Fm} - \mathbf{d})^T \mathbf{W}(\mathbf{Fm} - \mathbf{d})], \quad \text{(B.1)}$$

where \mathbf{W} is symmetric and positive definite weighting matrix. The expression for the hyper-quadratic is:

$$\mathbf{m}^T \mathbf{F}^T \mathbf{W} \mathbf{F} \mathbf{m} - 2 \mathbf{m}^T \mathbf{F}^T \mathbf{W} \mathbf{d} + \mathbf{d}^T \mathbf{W} \mathbf{d} - tol^2 = 0 \quad \text{(B.2)}$$

Taking into account the Cholesky's decomposition of \mathbf{W}

$$\mathbf{W} = \mathbf{L} \mathbf{L}^T, \quad \text{(B.3)}$$

where \mathbf{L} is lower triangular, Eq. (B.2) can be written as follows:

$$\mathbf{m}^T \mathbf{F}_w^T \mathbf{F}_w \mathbf{m} - 2 \mathbf{m}^T \mathbf{F}_w^T \mathbf{d}_w + \mathbf{d}_w^T \mathbf{d}_w - tol^2 = 0, \quad \text{(B.4)}$$

where $\mathbf{F}_w = \mathbf{L}^T \mathbf{F}$ and $\mathbf{F}_w = \mathbf{L}^T \mathbf{d}$. This new expression for the hyper-

quadratic is equivalent to Eq. (13). Therefore the analysis for the weighted problem is similar to the un-weighted case. Finally, the same methodology is valid for the weighted regularized problem.

References

- Aster, R.C., Borchers, B., Thurber, C.H., 2012. *Parameter Estimation and Inverse Problems*, 2nd ed. Academic Press, New York.
- Banks, H.T., Grosse, I., 1998. *A Comparison of Noise Generation Techniques and the Effects on Inverse Problem Calculations*. Tech. rep. North Carolina State University, Raleigh, NC.
- Barhen, J., Berryman, J.G., Borcea, L., Dennis, J., de Groot-Hedlin, C., Gilbert, F., Gill, P., Heinkenschloss, M., Johnson, L., McEvilly, T., Moré, J., Newman, G., Oldenburg, D., Parker, P., Porto, B., Sen, M., Torczon, V., Vasco, D., Woodward, N.B., 2000. *Optimization and Geophysical Inverse Problems*. Tech. rep. Ernest Orlando Lawrence Berkeley National Laboratory, San Jose, California.
- Blakely, R.J., 1996. *Potential Theory in Gravity and Magnetic Applications*, 1st ed. Cambridge University Press, New York.
- Constable, S.C., Parker, R.L., Constable, C.G., March 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics* 52 (3), 289–300.
- Fernández-Martínez, J.L., Fernández Muñiz, Z., Tompkins, M.J., January 2012. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* 77 (1), W1–W15.
- Fernández-Martínez, J.L., Fernández Muñiz, Z., Pallero, J.L.G., Pedruelo-González, L.M., 2013a. From Bayes to Tarantola: new insights to understand uncertainty in inverse problems. *J. Appl. Geophys.* 98, 62–72.
- Fernández-Martínez, J.L., Fernández Muñiz, Z., Pallero, J.L.G., Pedruelo-González, L.M., 2013b. Noise, regularization and uncertainty: new insights for linear and nonlinear inverse problems. *IPDO 2013: 4th Inverse Problems, Design and Optimization Symposium*. Albi (France).
- Fernández-Martínez, J.L., Pallero, J.L.G., Fernández-Muñiz, Z., Pedruelo-González, L.M., 2014. The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case. *J. Appl. Geophys.* <http://dx.doi.org/10.1016/j.jappgeo.2014.05.005>.
- Menke, W., 1989. *Geophysical data analysis: discrete inverse theory*. No. 45 in International Geophysics Series, Academic Press, New York.
- Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58 (1), 267–288.

ARTICLE 3

**The effect of noise and Tikhonov's
regularization in inverse problems.
Part II: The nonlinear case**

Authors:

Juan Luis Fernández Martínez
José Luis García Pallero
María Zulima Fernández Muñiz
Luis Mariano Pedruelo González

JOURNAL OF APPLIED GEOPHYSICS
2014, VOL. 108, PP. 186–193
[doi:10.1016/j.jappgeo.2014.05.005](https://doi.org/10.1016/j.jappgeo.2014.05.005)

Summary

In this work, the effect of noise and Tikhonov regularization in discrete nonlinear inverse problems is analyzed. The effect of noise in nonlinear problems has been barely analyzed in scientific literature and, when it was done, the performed analysis does not differ from the one used in linear problems, which has serious limitations. In this article, as it was done in the work 2, an analytical approach is used in a way that the results could be applied to any particular problem. The main conclusions are:

- As in the linear case, noise in data «moves» the least squares solution with respect to the real model, so the *true* solution can not be found.
- Noise deforms the linearized cost function in a non homogeneous way, in the sense that the deformation depends on the initial model \mathbf{m}_0 selected to solve the problem.
- Under certain conditions, the noise might decrease the size of the regions of very low misfits and at the same time increase the size of the regions of medium misfit. The inverse problem solution becomes very difficult to find using search methods, but locating the region of medium misfit is easier.
- As in the linear case, regularization deform the linearized regions of equivalence anisotropically. The difference is that in the nonlinear case this deformation is non homogeneous and depends on the model \mathbf{m}_0 that is adopted.
- In the case of multimodal cost functions the regularization does not make the relative minima to disappear, so the possibility that the local optimization methods are trapped in local minima is present and still depending on the initial model \mathbf{m}_0 . The regularization only has the effect of improving locally the ill-conditioning of the direct operator Jacobian matrix.
- The linearization only approximates locally the equivalence regions of the original nonlinear problem. For a given error level, the linearized equivalence region contain models with a higher error level than the specified level, while other models which are in the tolerance level remain outside.

These are important conclusions that highlight the importance of a correct uncertainty analysis.



The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case



J.L. Fernández-Martínez ^{a,*}, J.L.G. Pallero ^b, Z. Fernández-Muñiz ^a, L.M. Pedruelo-González ^a

^a Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain

^b ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Madrid, Spain

ARTICLE INFO

Article history:

Received 25 February 2014

Accepted 7 May 2014

Available online 20 May 2014

Keywords:

Noise

Regularization

Uncertainty analysis

Nonlinear inversion

ABSTRACT

In this paper we analyze the roles of noise in data and that of the regularization for nonlinear inverse problems, generalizing the analysis shown in a first accompanying paper for linear inversion. As in the linear case, the noise shifts the solution found by nonlinear optimization, but in the nonlinear case it also deforms non-homogeneously the topography of the cost function. Under certain conditions, the noise might decrease the size of the regions of very low misfits, and at the same time increase the size of the regions of medium misfits. As a result of that the inverse problem solution becomes more difficult to be found by search methods, but locating the region (or regions) of medium misfits is easier. Also, with respect to the use of local optimization methods the noise might act as a certain kind of regularization, since it serves to decrease the size of the region of equivalence. The effect of the regularization in the linearized region of equivalence is similar to the linear case; nevertheless this region does not coincide with the nonlinear region of equivalence. Although the use of regularization techniques has a positive effect in decreasing the uncertainty of the solution that has been found, unfortunately it does not provoke the disappearance of the nonlinear equivalent models. This manuscript, and its accompanying paper, provides new insights about of the role of noise and that of the regularization in the linear and nonlinear uncertainty analysis of inverse problems. This knowledge is very important, especially in applied geophysics, because noise in data is always present, and the possibility of having at disposal good prior information is rare. Therefore, the nonlinear uncertainty analysis and appraisal of the solution is always needed.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Discrete inverse problems are ill-posed in the sense that the observed data do not contain enough information to uniquely constrain the inverse model parameters. Noise in data, incomplete data coverage/sampling, and numerical errors and modeling hypothesis cause the solution of any inverse problem to be uncertain. Due to this issue, solving an inverse problem must always include the uncertainty analysis on the solution that has been adopted. This step is also known as model appraisal.

In a first accompanying paper (Fernández-Martínez et al., 2014) we have analyzed the effects of the noise in data and that of the regularization in linear inverse problems, showing that the noise in data shifts the solution from the true model that has generated the data, and deforms homogeneously the topography of the cost function. The deformation is called homogeneous because it does not depend on the model that

is considered to perform this analysis. We also have shown how depending on the kind of noise, the region of linear uncertainty might (or not) decrease in size. We have also shown the role of the zero-order regularization in stabilizing the inversion, and in informing the components of the solution that lie on the kernel of the forward operator, when a reference model (prior) is used in the regularization term. Although the theory of linear analysis is well established, much of the practitioners in geophysical inversion seem not to be aware of these results.

As we have shown in the first accompanying paper, most of the linear inverse problems come from the numerical discretization of continuous inverse problems that are described by Fredholm integral equations of the first kind. Nevertheless, there exists in applied geophysics a wide class of problems where the dependence of the forward model \mathbf{F} and the geophysical model parameters \mathbf{m} to achieve the data predictions \mathbf{d} , at certain locations in the data space, is nonlinear. The forward model \mathbf{F} in some of these problems might be given by a set of partial differential equations. It is possible to find examples of these nonlinear inverse problems in electromagnetism, electrical methods, gravimetry, magnetics, magnetotellurics, etc.

The effect of noise in nonlinear inversion has been barely analyzed in the literature, not differentiating its analysis from the linear case, and

* Corresponding author.

E-mail addresses: jlfm@uniovi.es (J.L. Fernández-Martínez), jlg.pallero@upm.es (J.L.G. Pallero), zulima@uniovi.es (Z. Fernández-Muñiz), marianopg@educastur.princast.es (L.M. Pedruelo-González).

somehow implicitly assuming that this understanding will not bring any advance in real practice. Nevertheless, the concept of noise is intimately related to the quality of the solution that is found via the inversion of the forward model \mathbf{F} . Therefore, as it has been shown for the linear case, a poor understanding of the effect of noise, may cause noise and modeling errors to be confused, and the signal and the noise be mixed after inversion. The analysis of the effect of modeling errors in inversion will be performed in future research work. The effect of noise has been analyzed, for particular nonlinear inverse problems, via numerical experimentation. For instance, Vugrin, (2005) studied the effect of the noise in a groundwater identification problem via global search methods, concluding that: 1) the noise affects the test problems by increasing the minimum objective function value; 2) in most of the cases the noisy versions of smooth optimization problems require fewer iterations to find the solution than the smooth problem itself; and 3) a decrease in the perturbed objective function does not always result in a decrease in the *true* fit. The word *true* refers to the model parameters used to generate the synthetic observed data. Obviously, in real problems the *true* solution is always unknown, or it is a mathematical abstraction. In this paper we analytically prove that the noise might decrease the size of regions of low misfits, and at the same time increasing the size of regions of medium misfits. Therefore, with respect to the use of local optimization methods the noise might act as a certain kind of regularization. The germ of this idea has been known to electrical engineers for a long time, being initially proposed by Blackman and Tukey in 1958 (Blackman and Tukey, 1958a,b).

The second part of this paper is devoted to study the effect of the regularization in nonlinear inversion. Inverse problem solutions are usually referred to as unstable. In a nutshell, instability in inverse problems $\mathbf{F}(\mathbf{m}) = \mathbf{d}$ is due to the lack of continuity of the generalized inverse operator. In linear discrete inverse problems this lack of continuity is due to the fact that the linear forward operator \mathbf{F} is not injective, that is, \mathbf{F} has a non-trivial kernel. Accordingly, different regularization techniques have been introduced and extensively studied (Engl et al., 1996; Kaltenbacher et al., 2008; Tikhonov and Arsenin, 1977). Tikhonov and Arsenin introduced the concept of correctness set where the inverse operator becomes continuous and the inverse problem well-posed. They also proved that it is possible to achieve a stable pseudo solution (or quasi solution) of the inverse problem, by performing minimization in some appropriate class of suitable models belonging to the correctness set (see Zhdanov, (2002) for an excellent comprehensive review about this subject). For that purpose the correctness set should be compact (bounded and closed, that is, including its boundary). A way of inducing the correctness set is to approximate the original ill-posed inverse problem by a family of well-posed regularized problems of the type $\mathbf{m}_\varepsilon = \mathbf{F}_\varepsilon^{-1}(\mathbf{d})$, where $\varepsilon > 0$ is the regularization parameter, designed to approach to zero. The simplest way to define \mathbf{F}_ε is by introducing in the regularized misfit function $c_\varepsilon(\mathbf{m})$, a penalization term that takes into account the distance to the prior or reference model \mathbf{m}^{ref}

$$c_\varepsilon(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_p + \varepsilon \|\mathbf{m} - \mathbf{m}^{ref}\|_q, \quad (1)$$

where p and q , respectively represent the norms adopted in the data and model spaces (usually the Euclidean norms).

Nevertheless, most of all these complex and beautiful pieces of mathematical research, apply for inverse problems with a unique solution, noise levels tending to zero, and forward operators \mathbf{F} fulfilling certain regularity conditions. This is not the case in real applications, and especially in applied geophysics, where the fields of interest are often quite weak and must be observed in an environment that contains extraneous fields that contribute with noise to the observational data (Barhen et al., 2000). Also, it has been shown that the solution provided by local optimization methods is highly dependent on the initial guess and on the type of the regularization and prior information that have been adopted. Due to these limitations, the Bayesian approach of

inverse problems has been proposed in geosciences (Tarantola and Valette, 1982a,b) to account for inverse model uncertainties. However, Bayesian techniques are restricted to low dimensional inverse problems with fast forward computations (Tarantola, 2006). Extensive literature has been dedicated to this subject, whose analysis is not the aim of the present paper.

Recently Fernández-Martínez et al., (2012, 2013) studied the topography of the cost function in linear and nonlinear inverse problems, briefly analyzing the effect of the noise and the role of the regularization, showing that the region (or regions) containing the models that fit the observed data with a certain error tolerance, and are compatible with the prior information, have a flat curvilinear valley shape. In this paper we analyze more in detail the roles of noise in data, and that of the regularization, in nonlinear inverse problems. The case of linear inverse problems has been analyzed in a first accompanying paper (Fernández-Martínez et al., 2014), showing that the noise shifts the solution found by least squares, and deforms homogeneously the region of equivalent models (linear hyper-quadratic). In the case of nonlinear inverse problems the noise also shifts the solution found by nonlinear optimization methods, and deforms the topography of the cost function, this time, in a non-homogeneous way, that is, this deformation depends on the model that is considered to perform this analysis.

With respect to role of the regularization, similar results to the linear case, are shown for the linearized inverse problem. Nevertheless, the nonlinear case is very different since the low misfit region (nonlinear region of equivalence) might be composed of several disconnected flat curvilinear basins. Therefore, the use of an inadequate reference model (prior information) might lead to wrong conclusions, when no uncertainty assessment is performed. Although the use of regularization techniques has a positive effect in decreasing the uncertainty on the solution that has been found (by providing the correctness set of solutions), the regularization does not provoke the disappearance of the nonlinear equivalent models, as it would be highly desired. Linear analysis does not serve to account for the nonlinear uncertainty, since the nonlinear and linearized equivalent regions have different geometric shapes. Further research is needed to address the uncertainty analysis in high dimensional nonlinear inverse problems, especially in the case of inverse problems with very costly forward problems (Fernández-Martínez et al., 2012, 2013; Osypov et al., 2008; Tompkins et al., 2011a,b, 2013), such as the geophysical techniques that are currently used in oil and gas upstream activities (full wave seismic inversion, controlled source electromagnetics, reservoir history matching and production optimization, etc).

The final aim of this paper, by studying the effect of noise and the regularization in the inverse function landscape, is to make practitioners aware of all these complexities, and to contribute in boosting this very interesting research line, consisting in the epistemological uncertainty analysis of very complex systems, due to its economic relevance in applied geophysics, and more generally in science and technology.

2. The effect of the noise in nonlinear inverse problems

In this section we analytically analyze how noise in data perturbs the location of the global minimum found by nonlinear optimization, while in Section 3 we study how noise affects to the cost function landscape, particularly in the region of nonlinear equivalence.

In the case of nonlinear inverse problems, let us denote by

$$\begin{aligned} c_p(\mathbf{m}) &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 = \|\mathbf{F}(\mathbf{m}) - (\mathbf{d}^{true} + \delta\mathbf{d})\|_2^2 = \\ &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}\|_2^2 + \|\delta\mathbf{d}\|_2^2 - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] = \\ &= c(\mathbf{m}) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2, \end{aligned} \quad (2)$$

the perturbed cost function, due to the effect of noise $\delta\mathbf{d}$ in data. The necessary condition to find the model \mathbf{m}_0 , where the minimum of $c_p(\mathbf{m})$ is achieved, is $\nabla c_p(\mathbf{m}_0) = \mathbf{0}$. Taking into account that

$$\frac{\partial c_p(\mathbf{m})}{\partial m_i} = \frac{\partial c(\mathbf{m})}{\partial m_i} - 2 \sum_{k=1}^s \frac{\partial f_k(\mathbf{m})}{\partial m_i} \delta d_k, \quad (3)$$

we have

$$\nabla c_p(\mathbf{m}) = \nabla c(\mathbf{m}) - 2\mathbf{J}\mathbf{F}_m^T \delta\mathbf{d}. \quad (4)$$

Thus, the necessary stationary condition $\nabla c_p(\mathbf{m}_0) = \mathbf{0}$ implies

$$\nabla c(\mathbf{m}_0) = 2\mathbf{J}\mathbf{F}_{m_0}^T \delta\mathbf{d}, \quad (5)$$

instead of $\nabla c(\mathbf{m}_0) = \mathbf{0}$, for the noise-free case. This simple analysis shows that in presence of noise the minimum of the perturbed cost function will never coincide with the minimum in the noise-free case. Also it is easy to understand that the solution found with no regularization might be spurious, since the noise is directly involved in its determination (see Eq. (5)). Obviously, its effect might not be very important if the level of noise is close to zero, but this is not the usual case in geophysical practice.

Taking into account that

$$\nabla c(\mathbf{m}_0) = 2\mathbf{J}\mathbf{F}_{m_0}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}], \quad (6)$$

then, Eq. (5) implies that

$$2\mathbf{J}\mathbf{F}_{m_0}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}] = 2\mathbf{J}\mathbf{F}_{m_0}^T \delta\mathbf{d} \Leftrightarrow \mathbf{J}\mathbf{F}_{m_0}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{obs}] = \mathbf{0}, \quad (7)$$

that is, the stationary condition is equivalent to fulfill

$$\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{obs} \in \ker \mathbf{J}\mathbf{F}_{m_0}^T = (\text{Im} \mathbf{J}\mathbf{F}_{m_0}^T)^\perp. \quad (8)$$

Eq. (8) means that the prediction data error in the model \mathbf{m}_0 , candidate to be the solution of the inverse problem $\mathbf{F}(\mathbf{m}) = \mathbf{d}^{obs}$, has to lie in the kernel of $\mathbf{J}\mathbf{F}_{m_0}^T$. Therefore, it has to be orthogonal to the column space of the Jacobian of \mathbf{F} in \mathbf{m}_0 . This result generalizes what it is known for linear inverse problems and least squares methods, since in the linear case we have $\mathbf{F}(\mathbf{m}_0) = \mathbf{F}\mathbf{m}_0$, and $\mathbf{J}\mathbf{F}_{m_0}^T \equiv \mathbf{F}$. Also, for the Hessian we have

$$Hc_p(\mathbf{m}) = Hc(\mathbf{m}) - H\psi(\mathbf{m}), \quad (9)$$

with $\psi(\mathbf{m}) = 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}]$. Taking into account that

$$Hc(\mathbf{m}) = 2\mathbf{J}\mathbf{F}_m^T \mathbf{J}\mathbf{F}_m + 2 \sum_{k=1}^s [f_k(\mathbf{m}) - d_k^{true}] Hf_k(\mathbf{m}), \quad (10)$$

$$\frac{\partial^2 \psi(\mathbf{m})}{\partial m_i \partial m_j} = 2 \sum_{k=1}^s \frac{\partial^2 f_k(\mathbf{m})}{\partial m_i \partial m_j} \delta d_k, \quad (11)$$

and

$$H\psi(\mathbf{m}) = \begin{pmatrix} \frac{\partial^2 \psi(\mathbf{m})}{\partial m_1^2} & \cdots & \frac{\partial^2 \psi(\mathbf{m})}{\partial m_1 \partial m_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \psi(\mathbf{m})}{\partial m_n \partial m_1} & \cdots & \frac{\partial^2 \psi(\mathbf{m})}{\partial m_n^2} \end{pmatrix}, \quad (12)$$

then

$$\begin{aligned} Hc_p(\mathbf{m}) &= \\ &= 2\mathbf{J}\mathbf{F}_m^T \mathbf{J}\mathbf{F}_m + 2 \sum_{k=1}^s [f_k(\mathbf{m}) - (d_k^{true} + \delta d_k)] Hf_k(\mathbf{m}) = \\ &= 2\mathbf{J}\mathbf{F}_m^T \mathbf{J}\mathbf{F}_m + 2 \sum_{k=1}^s [f_k(\mathbf{m}) - d_k^{obs}] Hf_k(\mathbf{m}), \end{aligned} \quad (13)$$

as it should be (see for example Aster et al. (2012)).

In addition

$$c_p(\mathbf{m}_0) = c(\mathbf{m}_0) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2, \quad (14)$$

and

$$c_p(\mathbf{m}^{true}) = c(\mathbf{m}^{true}) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}^{true}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2. \quad (15)$$

Thus, we have

$$\begin{aligned} c_p(\mathbf{m}_0) - c_p(\mathbf{m}^{true}) &= \\ &= c(\mathbf{m}_0) - c(\mathbf{m}^{true}) + 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}^{true}) - \mathbf{F}(\mathbf{m}_0)] = \\ &= \delta c + 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}^{true}) - \mathbf{F}(\mathbf{m}_0)], \end{aligned} \quad (16)$$

where $\delta c = c(\mathbf{m}_0) - c(\mathbf{m}^{true}) > 0$, since \mathbf{m}^{true} corresponds to the minimum of the noise-free cost function $c(\mathbf{m})$. Also, $c_p(\mathbf{m}^{true}) > c_p(\mathbf{m}_0)$ if

$$\delta c + 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}^{true}) - \mathbf{F}(\mathbf{m}_0)] < 0. \quad (17)$$

In this case, the true model \mathbf{m}^{true} that has generated the data will have a greater misfit than the solution \mathbf{m}_0 of the nonlinear inverse problem that has been found in the presence of noise. Nevertheless the model \mathbf{m}^{true} will still be located in the region of equivalence containing the low-misfit models (nonlinear region of equivalence). This is one of the main reasons to explain why uncertainty analysis is always needed, and especially in geophysical inversion: due to the effect of noise it will be impossible to recover the true solution of the inverse problem, but it will be located in the nonlinear equivalence region in an isoline of higher misfit than the global minimum.

3. The effect of noise on the cost function landscape for nonlinear inverse problems

Three different subjects will be involved in this section:

1. To analytically show how noise affects the topography of the nonlinear inverse problem cost function;
2. To study how the noise affects the linearized inverse problem. The analysis shown in a first accompanying paper (Fernández-Martínez et al., 2014) will help to provide the answer. This analysis is important since most nonlinear least squared methods used linearizations, that are obviously affected by the noise;
3. To analyze if the linearization helps to understand the effect of noise in nonlinear inverse problems, and particularly concerning the nonlinear uncertainty analysis of the solution.

To explore more in detail how the noise affects the cost function landscape for nonlinear inverse problems, let us consider a model \mathbf{m}_0 that belongs to the nonlinear equivalent region of value tol

$$V_{tol} = \{ \mathbf{m} : c_p(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2 \leq tol \}, \quad (18)$$

and adopting a second order Taylor expansion of $\mathbf{F}(\mathbf{m})$ in the neighborhood of \mathbf{m}_0

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + \mathbf{ST}_{\mathbf{m}_0} + o(\|\mathbf{m} - \mathbf{m}_0\|_2^2). \quad (19)$$

In this notation, $\mathbf{ST}_{\mathbf{m}_0} \in \mathbb{R}^s$ is the vector that contains the Hessian terms of the component functions of \mathbf{F}

$$\mathbf{ST}_{\mathbf{m}_0}(j) = \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T \text{Hf}_j(\mathbf{m} - \mathbf{m}_0), \quad (20)$$

and $o(\|\mathbf{m} - \mathbf{m}_0\|_2^2)$ is a vector function that vanishes faster than $\|\mathbf{m} - \mathbf{m}_0\|_2^2$ when \mathbf{m} tends to \mathbf{m}_0 . Then, we have

$$\begin{aligned} c_p(\mathbf{m}) &= \left\| \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{obs} + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + \mathbf{ST}_{\mathbf{m}_0} + o(\|\mathbf{m} - \mathbf{m}_0\|_2^2) \right\|_2^2 = \\ &= \mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \mathbf{m} + 2(\Delta \mathbf{d}^* - \delta \mathbf{d})^T (\mathbf{JF}_{\mathbf{m}_0} \mathbf{m} + \mathbf{ST}_{\mathbf{m}_0}) \\ &\quad + 2\mathbf{ST}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \mathbf{m} + \mathbf{ST}_{\mathbf{m}_0}^T \mathbf{ST}_{\mathbf{m}_0} + (\Delta \mathbf{d}^* - \delta \mathbf{d})^T (\Delta \mathbf{d}^* - \delta \mathbf{d}) \\ &\quad + 2o(\|\mathbf{m} - \mathbf{m}_0\|_2^2)^T (\Delta \mathbf{d}^* - \delta \mathbf{d} + \mathbf{JF}_{\mathbf{m}_0} \mathbf{m} + \mathbf{ST}_{\mathbf{m}_0}) + \\ &\quad o(\|\mathbf{m} - \mathbf{m}_0\|_2^2)^T o(\|\mathbf{m} - \mathbf{m}_0\|_2^2), \end{aligned} \quad (21)$$

where $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta \mathbf{d}$ and $\Delta \mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true} - \mathbf{JF}_{\mathbf{m}_0} \mathbf{m}_0$.

Obviously, this second order Taylor expansion truncation introduces an error in the approximation of the cost function topography that might be very important in the uncertainty analysis of the inverse problem solution. Besides, the linearized equivalent region of value tol around \mathbf{m}_0 is deduced adopting the first Taylor order approximation of $\mathbf{F}(\mathbf{m}) \approx \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0)$, and is the hyper-quadric

$$L_{Rtol}(\mathbf{m}_0) = \mathbf{m} : \left\| \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) - \mathbf{d}^{obs} \right\|_2 < tol. \quad (22)$$

In this case we have neglected all the higher order terms and particularly those induced by the second order term $\mathbf{ST}_{\mathbf{m}_0}$.

Equivalently, this region can be written in terms of model increments $\Delta \mathbf{m} = \mathbf{m} - \mathbf{m}_0$ as follows

$$\Delta \mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \Delta \mathbf{m} + 2(\Delta \mathbf{d} - \delta \mathbf{d})^T \mathbf{JF}_{\mathbf{m}_0} \Delta \mathbf{m} + (\Delta \mathbf{d} - \delta \mathbf{d})^T (\Delta \mathbf{d} - \delta \mathbf{d}) < tol^2,$$

where $\Delta \mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}$.

The center of the linearized hyper-quadric coincides with the Gauss–Newton solution of the nonlinear inverse problem in \mathbf{m}_0 (Fernández-Martínez et al., 2012)

$$\Delta \mathbf{m}_{GN} = -(\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0})^\dagger \mathbf{JF}_{\mathbf{m}_0}^T (\Delta \mathbf{d} - \delta \mathbf{d}) = \Delta \mathbf{m}_{GN}^{true} + (\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0})^\dagger \mathbf{JF}_{\mathbf{m}_0}^T \delta \mathbf{d}. \quad (23)$$

Note that the solution found by the Gauss–Newton method for the noise free case, \mathbf{m}_{GN}^{true} , is perturbed by the presence of noise $\delta \mathbf{d}$ through the Gauss–Newton pseudo inverse operator $(\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0})^\dagger$. Now, considering the singular value decomposition of $\mathbf{JF}_{\mathbf{m}_0} = \mathbf{U} \Sigma_{\mathbf{m}_0} \mathbf{V}^T$ we have

$$\begin{aligned} (\mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0})^\dagger \mathbf{JF}_{\mathbf{m}_0}^T \delta \mathbf{d} &= (\mathbf{V} \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \mathbf{V}^T)^\dagger \mathbf{V} \Sigma_{\mathbf{m}_0}^T \mathbf{U}^T \delta \mathbf{d} = \\ &= \mathbf{V} (\Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0})^\dagger \Sigma_{\mathbf{m}_0}^T \mathbf{U}^T \delta \mathbf{d} = \sum_{k=1}^r \frac{\delta d_{Uk}}{\mu_k} \mathbf{v}_k, \end{aligned} \quad (24)$$

and the noise in data is amplified back into the model parameters due to the action of the smallest singular values μ_k of the Jacobian $\mathbf{JF}_{\mathbf{m}_0}$. This effect is similar to the ill-conditioning effect shown for linear inverse problems, but in this case the amplification depends on the model \mathbf{m}_0 that is considered to perform the linearization. It is important to note

that while the linearized problem can be locally very badly ill-conditioned, and the linearized equivalence region $L_{Rtol}(\mathbf{m}_0)$ will correspond to an elliptical cylinder, the true nonlinear equivalence region $V_{tol}(\mathbf{m}_0)$ might have a finite extent, being its meandering shape explained (or induced) by the continuous dependence of $L_{Rtol}(\mathbf{m}_0)$ on $\mathbf{JF}_{\mathbf{m}_0}$. It can be said in this case, that the nonlinearities in the forward and inverse problems, help to limit the extent of the nonlinear equivalence region in order to fit the observed data.

As in the linear case (Fernández-Martínez et al., 2014) the noise does not affect the conditioning of the matrix of the linearized hyper-quadric ($\mathbf{JF}_{\mathbf{m}_0}$), causing homogeneous deformation of the linearized region of equivalence. Referred to the \mathbf{U} and \mathbf{V} bases provided by the singular value decomposition of the Jacobian $\mathbf{JF}_{\mathbf{m}_0}$, the linearized region of equivalence of value tol^* is

$$\Delta \mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta \mathbf{m}_V + 2\Delta \mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta \mathbf{m}_V + \|\Delta \mathbf{b}_U\|_2^2 < tol^{*2}, \quad (25)$$

where $\Delta \mathbf{m}_V = \mathbf{V}^T \Delta \mathbf{m}$, and $\Delta \mathbf{b}_U = \mathbf{U}^T (\Delta \mathbf{d} - \delta \mathbf{d})$. Then we arrive at:

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta b_{Uk})^2 = tol^{*2} - \sum_{k=r+1}^s \Delta b_{Uk}^2. \quad (26)$$

Also, the linearized equivalent region of value tol around the model \mathbf{m}_0 in the noise-free case is

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta d_{Uk})^2 = tol^2 - \sum_{k=r+1}^s \Delta d_{Uk}^2. \quad (27)$$

Reasoning in a similar way to the linear case shown in the accompanying paper Fernández-Martínez et al., (2014), the tol equivalent region in the noise-free case will have the same size as the tol^* region under the effect of noise and the relation of both tolerances is

$$tol^2 - \sum_{k=r+1}^s \Delta d_{Uk}^2 = tol^{*2} - \sum_{k=r+1}^s \Delta b_{Uk}^2. \quad (28)$$

Thus,

$$tol^*(\mathbf{m}_0, \delta \mathbf{d}) = \sqrt{tol^2 + \sum_{k=r+1}^s (\Delta b_{Uk}^2 - \Delta d_{Uk}^2)}. \quad (29)$$

In this case the value tol^* depends on the model \mathbf{m}_0 through the orthogonal matrices \mathbf{U} and \mathbf{V} given by the singular value decomposition of $\mathbf{JF}_{\mathbf{m}_0}$, and also on the level of noise.

Calling $\Delta = \sum_{k=r+1}^s (\Delta b_{Uk}^2 - \Delta d_{Uk}^2)$, we have

- If $\Delta > 0$, then $tol^*(\mathbf{m}_0, \delta \mathbf{d}) > tol$. Thus, the models have to be found in a region with higher misfit. The case $\Delta > 0$ will generally apply for models \mathbf{m}_0 with very small data misfits $\Delta d_{Uk}^2 \rightarrow 0$. Thus, relatively the region of lower misfits will decrease in size.
- If $\Delta < 0$, then $tol^*(\mathbf{m}_0, \delta \mathbf{d}) < tol$. This will happen for models with medium–high misfits. Thus, in presence of noise the models of medium–high misfits will increase in size.

The previous analysis depends on the level and type of noise that is added to the observed data, and on the model \mathbf{m}_0 that is considered. By continuity of the cost function there might exist a neutral line where both tolerances are equal $tol^*(\mathbf{m}_0, \delta \mathbf{d}) = tol$, that is, along this neutral line no deformation takes place. Inside the region limited by this line we will have $tol^*(\mathbf{m}_0, \delta \mathbf{d}) > tol$, that is, regions will decrease in size, and outside $tol^*(\mathbf{m}_0, \delta \mathbf{d}) < tol$, that is, regions will increase in size. This feature will explain why in presence of noise, to find the global minimum using search methods becomes a very hard task (“finding a needle in a haystack” (Tarantola, 2006)), and, conversely, locating the region of medium misfits is easier. Nevertheless, under the same conditions noise

in data has a regularization effect when using local optimization techniques, since it reduces the extent of the nonlinear equivalence region. These features will also preconize the use of global search methods to locate the region of medium misfits if the dimension is moderate and the forward problem is fast. Also, it is important to remark that this analysis has been simplified due to the effects of the higher order terms that have not been taken into account.

Finally, considering Eq. (2)

$$c_p(\mathbf{m}) = c(\mathbf{m}) - 2\delta\mathbf{d}^T\Delta\mathbf{d}(\mathbf{m}) + \|\delta\mathbf{d}\|_2^2, \quad (30)$$

where $\Delta\mathbf{d}(\mathbf{m}) = \mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}$ represents the modeling error for the model \mathbf{m} , the models where no deformation takes place will fulfill the condition

$$\delta\mathbf{d}^T\Delta\mathbf{d}(\mathbf{m}) = \frac{1}{2}\|\delta\mathbf{d}\|_2^2, \quad (31)$$

that is, the projection of the modeling error $\Delta\mathbf{d}(\mathbf{m})$ into the noise has to be positive and equal to the half of the squared energy of the noise in data, that is, this neutral line will exist only in case of colored noise with positive correlation with the residuals $\Delta\mathbf{d}(\mathbf{m})$.

3.1. A synthetic numerical example

To show the effect of the noise numerically, we have adopted a similar procedure with that adopted in the linear case, solving a simple nonlinear regression problem. The application of this analysis to a nonlinear geophysical inverse problem it is also possible. Nevertheless, visualization of the nonlinear equivalence region is not possible, as in the 2D synthetic case, since real nonlinear geophysical inverse problems are high dimensional. Also, the aim of this analysis is for it to become general knowledge since it has not been designed for a particular geophysical inverse problem.

In this case the data predictions in the different x_k points are given by

$$y_k = f(x_k; \alpha, \beta) = \alpha(1 - e^{-\beta x_k}) + \varepsilon, \quad (32)$$

where ε represents the effect of noise. The inverse problem is aimed at identifying the parameters (α, β) from a set of data points

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, \quad (33)$$

that have been generated adopting $(\alpha_t, \beta_t) = (20.0, 0.1)$ as the true model. As in the linear case (Fernández-Martínez et al., 2014) we have added in a first numerical test a Gaussian random variable $\varepsilon \rightarrow N(0, 0.05)$, trying to mimic the noise in the observed data during measurement. Also in a second numerical test, we have added to the previous one an extra noise term $\delta \rightarrow N(0, 0.025)$. The nonlinear inverse problem has been solved via the Gauss–Newton method using $(\alpha_0, \beta_0) = (15.0, 0.06)$ as initial guess.

Fig. 1 shows the results obtained for these two inversions. In both cases we show the topography of the cost function, the true model and the model found using local optimization, and the nonlinear and linearized equivalent regions of 9% relative error. Several interesting remarks can be made:

- (i) First, it is possible to observe the “croissant shape” of the region of equivalence, which is typical from nonlinear inverse problems (Fernández-Martínez et al., 2012). The hyper-quadric representing the linearized equivalent region calculated in the Gauss–Newton solution is also being represented and only represents locally the nonlinear uncertainty region.
- (ii) Secondly, both regions (the nonlinear and the linearized region) of 9% relative misfit, decrease in size as the level of noise increases.
- (iii) Third, the solution found in the perturbed (noisy) cases might move further away from the true solution as the level of noise

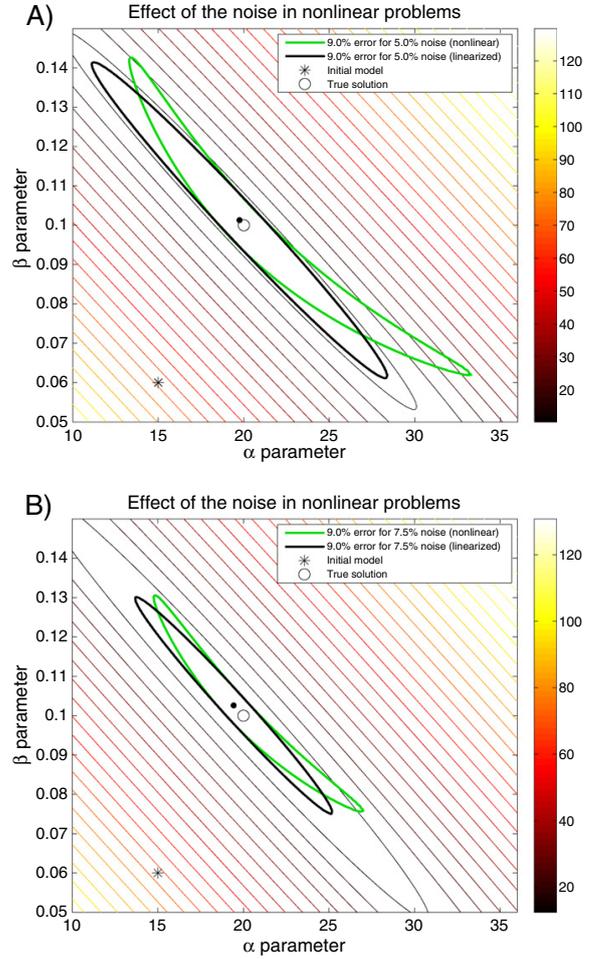


Fig. 1. Synthetic numerical example. Effect of the noise for a nonlinear inverse problem. Linearized and nonlinear equivalent region for 9% misfit tolerance and additive Gaussian noises of zero mean and standard deviations A) 5% and B) 7.5%, respectively.

increases. This depends on the statistical distribution of the noise that perturbs the observed data. In the present case this effect is not very strong. Nevertheless, the true solution is located inside the nonlinear equivalent region calculated in the model that has been adopted as the solution for the nonlinear inverse problem. The true solution has in both cases a higher misfit than the solution that has been found in presence of noise, as it has been analytically proven.

In conclusion, noise deforms non-homogeneously the topography of the cost function in nonlinear inverse problems. This deformation depends on the model that has been adopted to perform the analysis (as approximate solution of the nonlinear inverse problem) and on the level of noise. Under certain conditions, depending on the type of noise (white or colored noise) and its correlation with the model misfit, this deformation might provoke the search of the global optimum to be harder. At the same time the regions of medium misfits might increase in size, causing the search by global optimization methods in these regions to be easier. In any case, the solution of the inverse problem found in presence of noise never coincides with the true solution, that is, the hypothetical model parameters used to generate the noise-free (or true) observed data. Understanding the effect of the noise using

analytical formulas might help in the future to quantify more accurately its effect in the observed data and in nonlinear inversion, and also in designing synthetic noise to accelerate the model search for uncertainty analysis purposes.

4. The effect of the regularization in nonlinear inverse problems

Finally, we analyze the effect of the regularization in nonlinear inverse problems, as we did for the linear case (Fernández-Martínez et al., 2014).

The cost function with zero-order regularization is

$$\begin{aligned}
 c_p^r(\mathbf{m}) &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2 = \\
 &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}\|_2^2 + \|\delta\mathbf{d}\|_2^2 - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] \\
 &+ \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2 = \\
 &= c(\mathbf{m}) + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2 - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2 = \\
 &= c^r(\mathbf{m}) - 2\delta\mathbf{d}^T - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2,
 \end{aligned}
 \tag{34}$$

where the indexes p and r stand respectively for the perturbed and regularized cost function. Imposing the stationary condition we arrive at

$$\nabla c_p^r(\mathbf{m}) = \nabla c^r(\mathbf{m}) - 2\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \delta\mathbf{d} = 0.
 \tag{35}$$

Therefore

$$\nabla c^r(\mathbf{m}) = 2\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \delta\mathbf{d},
 \tag{36}$$

and the regularized solution for the perturbed case does not coincide with the solution found in the noise-free regularized case.

The nonlinear equivalent regions are related through

$$c_p^r(\mathbf{m}) \leq tol^2 \Leftrightarrow c^r(\mathbf{m}) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2 \leq tol^2.
 \tag{37}$$

Now, adopting a linearization of $c^r(\mathbf{m})$ in a model \mathbf{m}_0 located on the nonlinear equivalent region, it is possible to obtain the linearized region of equivalence for the regularized case as

$$\begin{aligned}
 \Delta\mathbf{m}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + 2(\Delta\mathbf{d} - \delta\mathbf{d})^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} \\
 + \varepsilon^2 (\Delta\mathbf{m} - \Delta\mathbf{m}^{ref})^T (\Delta\mathbf{m} - \Delta\mathbf{m}^{ref}) \\
 + (\Delta\mathbf{d} - \delta\mathbf{d})^T (\Delta\mathbf{d} - \delta\mathbf{d}) < tol^2,
 \end{aligned}
 \tag{38}$$

with $\Delta\mathbf{m}^{ref} = \mathbf{m}^{ref} - \mathbf{m}_0$.

Considering the singular decomposition of the Jacobian $\mathbf{J}\mathbf{F}_{\mathbf{m}_0} = \mathbf{U}\Sigma_{\mathbf{m}_0}\mathbf{V}^T$, the equation of the hyper-quadric referred to the \mathbf{V} base is

$$\begin{aligned}
 \Delta\mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + 2\Delta\mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + \\
 \varepsilon^2 (\Delta\mathbf{m}_V - \Delta\mathbf{m}_V^{ref})^T (\Delta\mathbf{m}_V - \Delta\mathbf{m}_V^{ref}) + \|\Delta\mathbf{b}_U\|_2^2 < tol^2.
 \end{aligned}
 \tag{39}$$

Eq. (39) can be written as follows

$$\begin{aligned}
 \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} \Delta m_{V_k} - \frac{\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \sum_{k=r+1}^n (\Delta m_{V_k} - \Delta m_{V_k}^{ref})^2 = \\
 = tol^2 + \sum_{k=1}^r \frac{(\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k})^2}{\mu_k^2 + \varepsilon^2} - \|\Delta\mathbf{b}_U\|_2^2 - \varepsilon^2 \sum_{k=1}^r (\Delta m_{V_k}^{ref})^2,
 \end{aligned}
 \tag{40}$$

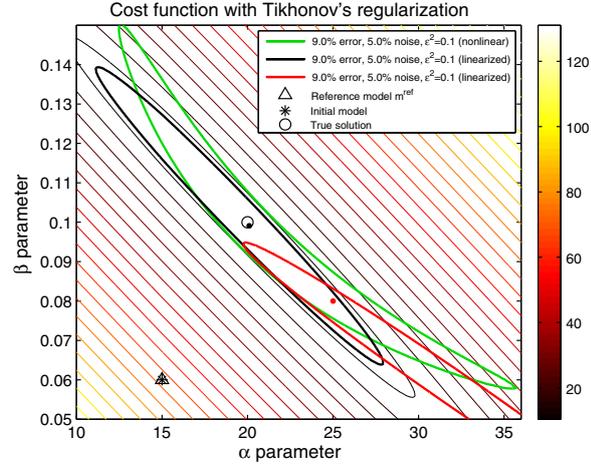


Fig. 2. Synthetic numerical example. Comparison of the nonlinear and linearized equivalent regions with Tikhonov regularization for two different models located on the nonlinear equivalent region, noise $\delta \rightarrow N(0,0.05)$, regularization parameter $\varepsilon^2 = 0.1$ and $\mathbf{m}^{ref} = (15,0.06)$.

where Δm_{V_k} represent the model increments with respect to \mathbf{m}_0 referred to the \mathbf{V} base. The center of the hyper-quadric coincides with the regularized solution of the linearized inverse problem.

Performing a similar analysis for the linearized equivalent regions with and without regularization, we arrive at a similar relationship to the one proposed for the linear case (Fernández-Martínez et al., 2014)

$$tol^*(\mu_i) = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left(tol^2 + \sum_{k=r+1}^n \Delta b_{U_k}^2 \right) + \varepsilon^2 \sum_{k=1}^r \frac{(\Delta b_{U_k} + \mu_k \Delta m_{V_k}^{ref})^2}{\mu_k^2 + \varepsilon^2}},
 \tag{41}$$

where $tol^*(\mu_i)$ is the tolerance of the equivalent region in the presence of regularization, and tol in its absence (without regularization). As in the linear case, the regularization deforms the linearized equivalent region in an anisotropic manner. Nevertheless, the main difference

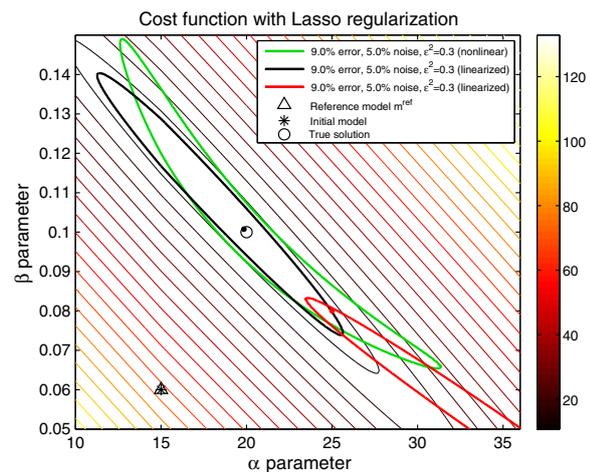


Fig. 3. Synthetic numerical example. Same comparison than in Fig. 2 when the Lasso regularization is employed and the damping parameter is increased to $\varepsilon^2 = 0.3$. Similar results are obtained for the case of Tikhonov's regularization analyzed in this paper.

resides in the fact that in the nonlinear case this deformation depends on the model \mathbf{m}_0 , through $\Delta\mathbf{m}^{ref} = \mathbf{m}^{ref} - \mathbf{m}_0$.

Fig. 2 shows, for the above mentioned synthetic numerical example, the linearized equivalent region for two different models, $\mathbf{m}_0 = (20.0, 0.1)$ and $\mathbf{m}_0 = (25.0, 0.08)$, of the nonlinear equivalent region. The numerical parameters used in this case are as follows: the initial guess used to perform the inversion coincides with the reference model that has been adopted in the regularization, $\mathbf{m}^{ref} = (15.0, 0.06)$; the damping parameter used was $\varepsilon^2 = 0.1$; and a noise term $\delta \rightarrow N(0, 0.05)$ was added to the observed data. It is possible to observe that the linearized equivalent region only represents locally the *real* nonlinear uncertainty region. Similar results are also obtained for other kinds of regularization schemes, such as Lasso (Tibshirani, 1996). This algorithm is used to look for sparse solutions by increasing the value of the regularization parameter ε^2 . This strategy has some potential dangers since the linearized equivalent region in a point of higher misfit than the global optimum might not be contained within the nonlinear equivalent region. This situation is shown in Fig. 3 for the synthetic numerical example solved with Lasso regularization, $\varepsilon^2 = 0.3$ and noise $\varepsilon \rightarrow N(0, 0.05)$. The analysis was also performed in the same models of the nonlinear equivalent region of the previous case ($\mathbf{m}_0 = (20.0, 0.1)$ and $\mathbf{m}_0 = (25.0, 0.08)$). Results are similar to those obtained using the zero-order Tikhonov's regularization (Fig. 2). In conclusion, depending on the value of the damping parameter, the

linearized equivalent region does not always represent, even locally, the nonlinear equivalent counterpart.

The conclusions of this analysis are the following:

- (i) The center of the linearized hyper-quadric provides the regularized solution of the linearized inverse problem in \mathbf{m}_0 and coincides with the Gauss–Newton solution of the nonlinear inverse problem.
- (ii) The noise and the regularization deform non-homogeneously the cost function topography. The regularization deforms the linearized equivalence region in an anisotropic way, acting differently in each model parameter. Taking into account this fact, we can conclude that the noise can be interpreted as a particular kind of regularization. In fact, the Levenberg–Marquardt's algorithm consists in adding a white noise of variance ε^2 to stabilize the inversion.
- (iii) The nonlinear and linearized equivalent regions are completely different in shape. It has been analytically proven in Fernández-Martínez et al. (2012, 2013) that the nonlinear equivalent region has a flat meandering valley shape. Also, there could be other low misfit basins in the cost function landscape if the forward operator is not injective. This is one of the main differences between the linear and nonlinear inverse problems. It can be easily understood that in the non-injective

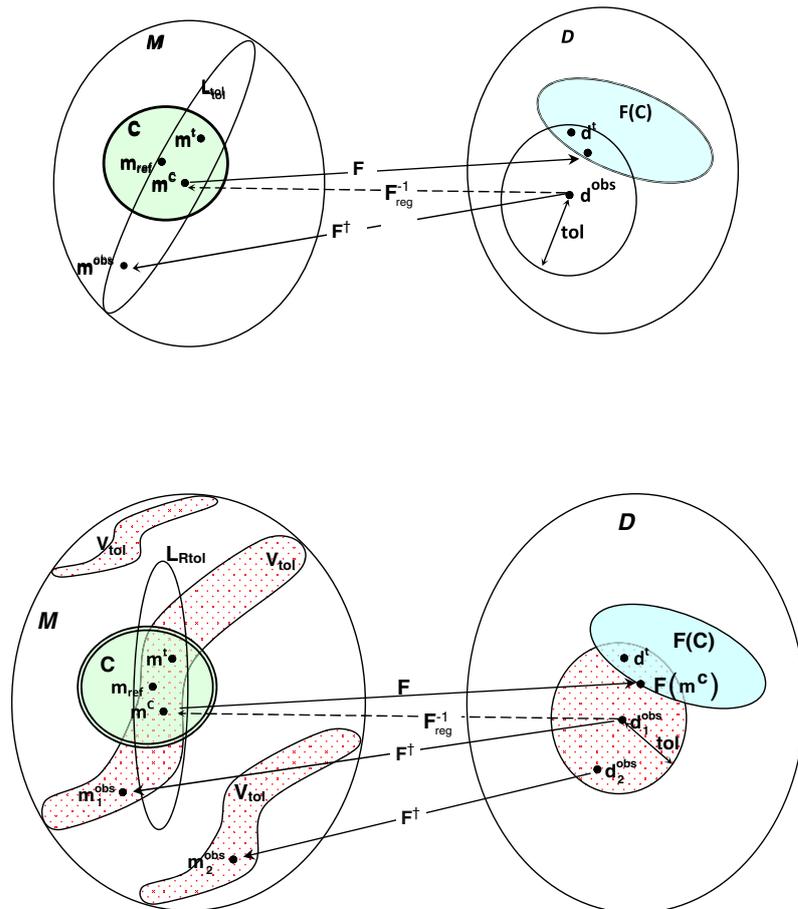


Fig. 4. Comparison of the effect of the regularization in a linear/nonlinear inverse problem. See the text for more details.

case the regularization might provide different solutions located in different basins of the nonlinear equivalent region, depending on the reference model that has been adopted. A sketch for this fact is shown in Fig. 4, where we compare the notion of correctness set in linear and nonlinear inversion. In nonlinear inversion the correctness set \mathbf{C} induced by the regularization might not contain the true solution of the nonlinear inverse problem that could be located in another different basin of the cost function topography. This is not the case in linear inverse problems, where the region of equivalence is unique (bounded or not), and the only possibility for that to happen is adopting a wrong reference model. Besides, the linearized equivalent region $L_{R_{tol}}$, only represents locally (when it does) the nonlinear equivalent region(s) V_{tol} . Thus, performing a full nonlinear uncertainty analysis is the only way to unravel the existence of very different equivalent solutions, in order to quantify risk.

5. Conclusions

Noise is an important ingredient in inverse problems, since its effect enters into the inverse problem cost function through the observed data. In this paper we have analyzed the effect of the noise and that of the regularization in nonlinear discrete inverse problems. We showed that the noise shifts the solution found by the typical nonlinear optimization methods, and deforms the topography of the cost function non-homogeneously. Under certain conditions, the noise might decrease the size of the regions of very low misfits and at the same time increase the size of the regions of medium misfits. The inverse problem solution becomes very difficult to find using search methods, but locating the regions of medium misfits is easier. The success of global optimization and searching methods in noisy problems might be related to this fact, that is, a noise-free problem is usually harder to optimize.

The role of the regularization in the linearized region of equivalence is similar to that shown for the linear case. In any case, under the presence of noise the model of minimum misfit is never the model that has generated the observed data. Although the use of regularization techniques has been presented as the panacea, and it obviously has a positive effect in decreasing the uncertainty of the solution that has been found, regularization does not cause the disappearance of the nonlinear equivalent models: regularization only serves at improving locally the ill-conditioning of the Jacobian of the forward operator. Besides, the linearized and the nonlinear problems have very different equivalence regions. Although the regularization serves to stabilize the inversion under the presence of noise, the regularized solution can be relatively far from the real one since other equivalent solutions might exist in different basins of the cost function topography.

The analysis shown in this paper is completely general and applies for any kind of nonlinear inverse problem. This knowledge might have a significant impact in nonlinear geophysical inversion, due to the omnipresence of noise in data, and to the impossibility of always having at disposal good prior information to design the regularization. Due to these facts, we can affirm that a full-nonlinear uncertainty analysis is always needed in applied geophysics in order to quantify risk in decision-making approaches. Therefore, no serious inversion research should be presented without its corresponding uncertainty analysis.

Acknowledgments

The authors would like to acknowledge the editor of the Journal of Applied Geophysics, Professor Klaus Holliger, for his support in understanding the need of theoretical analysis in geophysical inversion, and also the feedback of anonymous referees, that helped us to improve the quality of these manuscripts.

References

- Aster, R.C., Borchers, B., Thurber, C.H., 2012. *Parameter Estimation and Inverse Problems*, 2nd ed., Academic Press, New York.
- Barhen, J., Berryman, J.G., Borcea, L., Dennis, J., de Groot-Hedlin, C., Gilbert, F., Gill, P., Heinkenschloss, M., Johnson, L., McEvilly, T., Moré, J., Newman, G., Oldenburg, D., Parker, P., Porto, B., Sen, M., Torczon, V., Vasco, D., Woodward, N.B., 2000. *Optimization and Geophysical Inverse Problems*, Technical Report. Ernest Orlando Lawrence Berkeley National Laboratory, San Jose, California.
- Blackman, R.B., Tukey, J.W., 1958a. The measurement of power spectra from the point of view of communications engineering—part I. *Bell Syst. Tech. J.* 37, 185–282.
- Blackman, R.B., Tukey, J.W., 1958b. The measurement of power spectra from the point of view of communications engineering—part II. *Bell Syst. Tech. J.* 37, 485–569.
- Engl, H.W., Hanke, M., Neubauer, A., 1996. *Regularization of inverse problems*. Number 375 in *Mathematics and its Applications*, Kluwer Academic Publishers, Dordrecht.
- Fernández-Martínez, J.L., Fernández Muñiz, Z., Tompkins, M.J., 2012. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* 77, W1–W15. <http://dx.doi.org/10.1190/geo2011-0341.1>.
- Fernández-Martínez, J.L., Fernández Muñiz, Z., Pallero, J.L.G., Pedruelo-González, L.M., 2013. From Bayes to Tarantola: new insights to understand uncertainty in inverse problems. *J. Appl. Geophys.* 98, 62–72. <http://dx.doi.org/10.1016/j.jappgeo.2013.07.005>.
- Fernández-Martínez, J.L., Pallero, J.L.G., Fernández Muñiz, Z., Pedruelo-González, L.M., 2014. The effect of the noise and Tikhonov's regularization in inverse problems. Part I: the linear case. *J. Appl. Geophys.* <http://dx.doi.org/10.1016/j.jappgeo.2014.05.005>.
- Kaltenbacher, B., Neubauer, A., Scherzer, O., 2008. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, Walter de Gruyter, Berlin.
- Ospov, K., Nichols, D., Woodward, M., Zdraveva, O., Yarman, C.E., 2008. Uncertainty and resolution analysis for anisotropic tomography using iterative eigendecomposition. *SEG Technical Program Expanded Abstracts*, pp. 3244–3249. <http://dx.doi.org/10.1190/1.3064019>.
- Tarantola, A., 2006. Popper, Bayes and the inverse problem. *Nat. Phys.* 2, 492–494.
- Tarantola, A., Valette, B., 1982a. Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys. Space Phys.* 20, 219–232.
- Tarantola, A., Valette, B., 1982b. Inverse problems = quest for information. *J. Geophys.* 50, 159–170.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58, 267–288. <http://dx.doi.org/10.1190/1.1442303>.
- Tikhonov, A.N., Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems*, V. H. Winston and Sons, Washington, D. C.
- Tompkins, M.J., Fernández-Martínez, J.L., Alumbaugh, D.L., Mukerji, T., 2011a. Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: marine controlled-source electromagnetic examples. *Geophysics* 76, F263–F281. <http://dx.doi.org/10.1190/1.3581355>.
- Tompkins, M.J., Fernández-Martínez, J.L., Fernández-Muñiz, Z., 2011b. Marine electromagnetic inverse solution appraisal and uncertainty using model-derived basis functions and sparse geometric sampling. *Geophys. Prospect.* 59, 947–965. <http://dx.doi.org/10.1111/j.1365-2478.2011.00955.x>.
- Tompkins, M.J., Fernández-Martínez, J.L., Fernández-Muñiz, Z., 2013. Comparison of sparse-grid geometric and random sampling methods in nonlinear inverse solution uncertainty estimation. *Geophys. Prospect.* 61, 28–41. <http://dx.doi.org/10.1111/j.1365-2478.2012.01057.x>.
- Vugrin, K.E., 2005. *On the Effects of Noise on Parameter Identification Optimization Problems*. Ph.D. thesis, (Virginia, USA).
- Zhdanov, M.S., 2002. *Geophysical inverse theory and regularization problems*. Number 36 in *Methods in Geochemistry and Geophysics*, Elsevier.

ARTICLE 4

**Gravity inversion and uncertainty
assessment of basement relief via
Particle Swarm Optimization**

Authors:

José Luis García Pallero
Juan Luis Fernández Martínez
Sylvain Bonvalot
Olivier Fudym

JOURNAL OF APPLIED GEOPHYSICS
2015, VOL. 116, PP. 180–191
[doi:10.1016/j.jappgeo.2015.03.008](https://doi.org/10.1016/j.jappgeo.2015.03.008)

Summary

In this article, the Particle Swarm Optimization (PSO) algorithm is applied to the 2D gravity inversion problem in sedimentary basins. The gravity inverse problem is an essential tool for the Earth crust study at regional and local scales, as well as in mineral exploration and, as a potential-field based problem, has an inherent non-uniqueness, which makes it ill-posed and, then, an optimal target problem for the global search algorithms. In particular, the approach consist in the determination of the sediments-basement interface using gravity data observed at the terrain surface, imposing the density contrast. This corresponds to a nonlinear approach. After an exhaustive literature review, it is concluded that this is the first time the PSO algorithm is applied to this problem. The highlights of this research work are:

- A short summary of the problem solution via classical local optimization methods is presented.
- The problem solution via PSO is detailed. The search space determination is explained, as well as the possibility of using constraints and its comparison with their counterparts in the local optimization techniques. In this work, the GPSO, CC-PSO, CP-PSO, PP-PSO and RR-PSO family members are used.
- Synthetic models, without and with white noise, are solved. The evolution of the function error and swarm dispersion is analyzed through the iterations, which are important tasks to take into account when using PSO. Two assertions, which were theoretically proved in article 3 were verified: (i) the minimum of the cost function for the noisy problem does not coincide with the minimum of the noise-free problem, and (ii) the noise decreases the size of the equivalence region of lower misfit with respect to the noise-free problem. For that purpose, it was performed an interpolation of the cost function topography in the 2D PCA space using the models that have been sampled by PSO in the nonlinear region of equivalence of relative error lower than 20%.
- A real inversion problem using data from the Atacama Desert (north Chile) is presented. The obtained results are plausible from the geological point of view, and agree with the results obtained using local optimization methods. The PSO method adds to the problem the uncertainty analysis of the obtained solution. It is then demonstrated that the developed technique is a valid method for the 2D gravity inversion problem in sedimentary basins and is not computationally expensive.



Gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization



J.L.G. Pallero^a, J.L. Fernández-Martínez^{b,*}, S. Bonvalot^c, O. Fudym^d

^a ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Madrid, Spain

^b Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain

^c Laboratoire GET (Université de Toulouse, CNRS, IRD, CNES), Bureau Gravimétrique International (BGI), Toulouse, France

^d Université de Toulouse, Mines Albi, CNRS, Centre RAPSODEE, Albi cedex, France

ARTICLE INFO

Article history:

Received 3 December 2014

Received in revised form 3 March 2015

Accepted 4 March 2015

Available online 7 March 2015

Keywords:

Nonlinear gravity inversion

Particle Swarm Optimization

Uncertainty assessment

Sedimentary basin

ABSTRACT

Gravity inversion is a classical tool in applied geophysics that corresponds, both, to a linear (density unknown) or nonlinear (geometry unknown) inverse problem depending on the model parameters. Inversion of basement relief of sedimentary basins is an important application among the nonlinear techniques. A common way to approach this problem consists in discretizing the basin using polygons (or other geometries), and iteratively solving the nonlinear inverse problem by local optimization. Nevertheless, this kind of approach is highly dependent of the prior information that is used and lacks from a correct solution appraisal (nonlinear uncertainty analysis). In this paper, we present the application of a full family Particle Swarm Optimizers (PSO) to the 2D gravity inversion and model appraisal (uncertainty assessment) of basement relief in sedimentary basins. The application of these algorithms to synthetic and real cases (a gravimetric profile from Atacama Desert in north Chile) shows that it is possible to perform a fast inversion and uncertainty assessment of the gravimetric model using a sampling while optimizing procedure. Besides, the parameters of these exploratory PSO optimizers are automatically tuned and selected based on stability criteria. We also show that the result is robust to the presence of noise in data. The fact that these algorithms do not require large computational resources makes them very attractive to solve this kind of gravity inversion problems.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Gravity inversion is a classical tool in applied geophysics (Dobrin, 1960; Nettleton, 1976; Parker, 1994; Telford et al., 1976) to analyze the structure of sedimentary basins in mineral exploration, oil and gas upstream activities, hydrogeology, glaciology, etc. The gravity inverse problem is linear when a given geometry for the bodies under study is provided and the corresponding densities are estimated. Conversely, the inverse problem is nonlinear when the geometry of the bodies is treated as unknown, making some assumptions about the values of the corresponding densities.

Among the nonlinear techniques, the inversion of basement relief of a sedimentary basin is a relatively common task (Barbosa et al., 1997; Blakely, 1995; Chakravarthi and Sundararajan, 2007). The classical way to deal with this problem is via nonlinear optimization, where the unknowns are the depth of the basement at certain locations, or the depth and some additional parameters to take into account the density variation of the sediments. Among others, basement relief

estimation has important implications in oil and gas exploration to find out the location of possible stratigraphic traps (Silva et al., 2010), in hydrogeology studies to understand the geological structure of aquifers (Adema et al., 2007; Bohidar et al., 2001), in glaciology to infer the flow rate of discharge (Krimmel, 1970; Stern, 1978; Venteris and Miller, 1993), or in landfill analysis as a tool for density determination (Mantlik et al., 2009) and bottom relief estimation (Silva et al., 2009).

Basement relief estimation based on gravity anomalies could be high dimensional nonlinear inverse problem depending on the model parameterization. Over the years, several methods were used to approach this problem. Some of them were based on the manual modeling of the basin taken into account the adjustment of the observed anomaly by the basin gravimetric model (Bott, 1960). Local optimization techniques, particularly the Levenberg–Marquardt algorithm with Tikhonov's regularization, are commonly used in the resolution of nonlinear inverse problems: stating a prior model for the density distribution (fixed or variable with depth), and using an adequate set of constraints, a solution is achieved through iterative linearization of the cost function (Barbosa et al., 1997; Chakravarthi, 1995; Silva et al., 2006). This procedure provides a solution that is highly dependent on the initial model and on the prior information that are used. Besides, no model appraisal in the nonlinear sense is usually performed on the solution that has been found.

* Corresponding author.

E-mail addresses: jlg.pallero@upm.es (J.L.G. Pallero), jlfm@uniovi.es

(J.L. Fernández-Martínez), sylvain.bonvalot@ird.fr (S. Bonvalot), fudym@mines-albi.fr (O. Fudym).

Uncertainty assessment is a very important step in inversion (see for instance Scales and Snieder, 2000) since inverse problems always belong to decision-making processes, and they are by nature ill-posed, that is, there exist different solutions (called equivalent) that are compatible with the prior information and fit the observed data within the same error bounds. Fernández-Martínez et al. (2012a) studied the cost function landscape for linear and nonlinear inverse problems, showing that equivalent models (for a certain error tolerance) are located within the hyper-quadratic region of equivalence in the linear case. This flat elongated valley bends for nonlinear inverse problems. Also, several disconnected basins might compose the nonlinear equivalence landscape.

Uncertainty analysis consists in obtaining a representative set of model parameters in the low misfit region(s) of the cost function topography. Fernández-Martínez et al. (2013) provided a deterministic analysis of the inverse problem uncertainty, proving that the regularization does not provoke the disappearance of the equivalent models, showing the risks of adopting a wrong prior model, and highlighting the fact that linear analysis never accounts for the real uncertainty in nonlinear inverse problems.

Finally, noise in data is an important ingredient in inversion that deforms the cost function topography and it is intimately related to the uncertainty analysis of the solution and to the need of regularization. Fernández-Martínez et al. (2014a,b) have analyzed the effect of noise in inversion, showing that noise deforms the topography of the cost function in a homogeneous/inhomogeneous way, depending if the inverse problem is linear or nonlinear. As a consequence of noise the inverted model never coincides with the hypothetical true model that has generated the observed data. Nevertheless, the true model belongs to the region of equivalence having a higher misfit than the *true* solution. In conclusion, inversion and uncertainty analysis (solution appraisal) must always go hand in hand.

In this paper, a full family of Particle Swarm Optimizers is applied to the gravity inversion and uncertainty assessment of basement relief in sedimentary basins. Particle Swarm Optimization (PSO) is an interesting global optimization technique that was heuristically inspired by the social behavior of groups of animals (birds and fishes) in nature (Kennedy and Eberhart, 1995). Nevertheless, PSO was physically interpreted as a damped-mass spring system (Fernández-Martínez and García-Gonzalo, 2011b). The interest of having at disposal the PSO physical model consists in properly understanding the exploratory behavior of the different PSO family members, and relating stochastic stability of the particle trajectories with the algorithm convergence for any arbitrary statistical distribution of the PSO parameters (García-Gonzalo and Fernández-Martínez, 2014). Although no theoretical results exist to state the theoretic conditions needed for these algorithms to perform a correct posterior sampling, exploratory PSO versions have shown to perform a fairly good sampling of the nonlinear equivalent region when applied to different types of DC electrical inverse problems (Fernández-Martínez et al., 2010a,b), and also in oil and gas production history matching (Fernández-Martínez et al., 2012b). Similar results were also achieved for the Vertical Electrical Sounding inverse problem using binary genetic algorithms (Fernández-Álvarez et al., 2008). These results emphasize the fact that exploration is a key factor in performing a good approximate uncertainty analysis of the inverse problem solution (sampling while optimizing).

In this work we introduce for the first time the use of two novel PSO family optimizers (PP-PSO and RR-PSO) in geophysics, and particularly in 2D basement relief inversion, showing the application to synthetic and real data (a gravimetric profile acquired in the Atacama Desert, north Chile). We also compare the posterior sampling of the nonlinear region of equivalence performed by different PSO family members, analyzing numerically how noise in data affects the topography landscape and, therefore, the corresponding uncertainty analysis of the solution. The uncertainty analysis provided by PSO serves to answer important questions about the basin structure in a probabilistic framework, taking into account the intrinsic uncertainty of the basement relief inverse problem.

2. Classical modeling in 2D basement relief gravity inversion

Gravity inversion of basement relief in sedimentary basins can be proposed as a 2D or 3D problem depending on the model conceptualization. The 2D case is very common, and consists in the inversion of one or different profiles across the basin, generally containing its maximum depth. When the basin is elongated, a common approach consists in inverting several profiles that are orthogonal to its longest dimension, which analyzed together serve to depict a pseudo 3D model. The 2D approach assumes that the anomalous body (in this case the sediments) has an infinity extent in the perpendicular direction to the gravimetric profile.

Several ways to model a generic basin in 2D have been proposed in the literature: modeling the entire basin as a trapezoid (Rao, 1990), using polygons with an arbitrary number of vertices (Chakravarthi et al., 2001), using polynomial functions (Martín Atienza, 2001), or by accretion of rectangular prisms of known density (see for example Chakravarthi, 1995; Barbosa et al., 1997; Silva et al., 2006). Fig. 1 shows a 2D model divided into M prisms, whose density is imposed. The width of the prisms can be selected as a factor (generally between $0.5\times$ and $1.5\times$) of the average spacing between the N observation points, which do not necessarily need to be equally spaced. Obviously depending on the number of observation points and the number of prisms, the corresponding inverse problem might have an over-determined or under-determined character, that will highly impact the corresponding uncertainty analysis of the solution (Fernández-Martínez et al., 2012a).

2.1. Modeling of the density contrast

Multiple approximations have been proposed in the literature to model the density contrast $\Delta\rho = \rho_s - \rho_b$ between the sediments and the basement. For environments such as glaciers it is very common to use a constant density value (Krimmel, 1970; Stern, 1978; Venteris and Miller, 1993), although variable depth-density models could be also used (Shumskiy, 1960). For sedimentary basins, a constant density contrast is usually employed (Barbosa et al., 1997; Gabalda et al., 2005). Nevertheless, adopting a variable dependency with depth is also a common strategy, especially for deep environments. Some models for variable density contrast with depth can be seen in Table 1. The parameter needed to define these models of density variation can be estimated at inversion or being adjusted based on borehole information, which is a more realistic approach. The corresponding forward problem equations of gravity attraction following the different density models are presented in the respective papers (Chakravarthi, 1995; Rao, 1990; Rao et al., 1995). The useful exponential depth density variation model introduced in Cordell (1973) has no analytical expression in the space domain, even in the case of simple geometric bodies. Solutions to this problem have been proposed in the frequency domain (Chapell and Kusznir, 2008; Cordell, 1973; Granser, 1987).

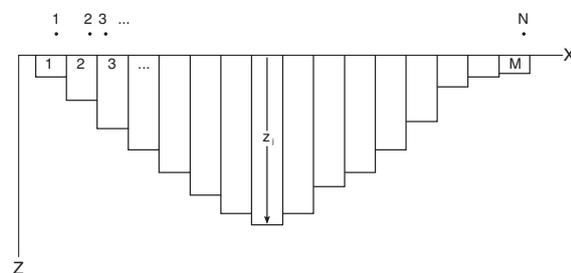


Fig. 1. Two dimensional modeling of a sedimentary basin by right rectangular prisms' accretion. The domain is divided into M rectangular prisms of known density, where depths z_i are unknown. The width of the prisms is a factor of the average separation between the N observed points.

Table 1
Several models for density contrast variation with depth.

Analytical expression	Reference
$\Delta\rho(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2$	Rao (1990)
$\Delta\rho(z) = \frac{\Delta\rho_0 \alpha^2}{(\alpha+z)^2}$	Rao et al. (1995)
$\Delta\rho(z) = \frac{\Delta\rho_0^2}{(\Delta\rho_0 - \alpha z)^2}$	Chakravarthi (1995)
$\Delta\rho(z) = \Delta\rho_0 \exp(-\alpha z)$	Cordell (1973)

In this paper the inversions have been performed using a constant density value, as suggested by the geological setup of the Atacama profile. Therefore the number of parameters in the inverse problem coincides with the number of prisms used to define the basin. In case a homogeneous depth dependency is given, the number of model parameters is increased by the number of parameters needed to model this dependency if they were considered as unknowns. The full nonlinear basement relief inverse problem allows this dependency to vary laterally across the basin. In that case it is crucially important to provide a good prior knowledge to constraint the density space variability. This problem is beyond the scope of this paper, but it will be the subject of future research work.

2.2. The 2D basement relief gravity inverse problem

One common way to deal with the inverse problem stated in Fig. 1 (valid for 2D and 3D) is exposed in Barbosa et al. (1997). Let a set of N observed gravity anomalies taken along a profile at observation points $p_i(x_i, z_i)$, and let us adopt a partition of the basin into M rectangular prisms with known constant density contrast $\Delta\rho = \rho_s - \rho_b$ (standing the subscripts s and b for sediments and basement, respectively). The gravity anomaly Δg at each observed point can be written as

$$\Delta g_i = \sum_{j=1}^M F(\Delta\rho_j, z_j, r_{ij}), \quad i = 1, 2, \dots, N, \quad (1)$$

where $F(\Delta\rho_j, z_j, r_{ij})$ is the function that solves the forward problem for an individual prism at the observation point i , $\Delta\rho_j$ is the density contrast of the j th prism, z_j is the depth of the j th prism, and r_{ij} is the horizontal position vector between the i th point and the j th prism. A detailed formulation for F is given in Appendix A.

The nonlinear inverse problem to estimate the depth model \mathbf{z} may be formulated as the minimization of the prediction error cost function

$$c^g(\mathbf{z}) = \frac{1}{N} \|\Delta\mathbf{g}^{obs} - \Delta\mathbf{g}^*(\mathbf{z})\|_2^2, \quad (2)$$

where $\Delta\mathbf{g}^{obs}$ is the vector of observed gravity anomalies, and $\Delta\mathbf{g}^*(\mathbf{z})$ the anomalies predicted by the model \mathbf{z} . The inverse problem stated in Eq. (2) is ill-posed, since there exist an infinity of solutions that fit the observed data within the same error bounds, and the determination of the solution with minimum misfit is very unstable. To alleviate the ill-posed character of the inverse problem and reducing the set of solutions to those that are compatible with the prior information, different kinds of constraints are commonly used (Barbosa et al., 1997; Silva et al., 2006, 2009). These constraints consist in imposing smoothness conditions of some degree on the basement relief (relative), or prescribing the values of the depth in some points, based on borehole information (absolute). For that purpose some regularization parameters which determine the trade-off between the different terms in the cost functions (prediction error and regularity and prior conditions) are needed. More detailed information about the regularization constraints is given in Appendix B. Different well-known iterative local optimization techniques, such as Levenberg–Marquardt (Marquardt, 1963), Gauss–Newton and Quasi-Newton methods (Nocedal and Wright, 1999) are used to solve the nonlinear regularized problem. Nevertheless, the result provided by

these algorithms depends on the initial model and on the prior information that are adopted. Besides, these methods do not provide the possibility to perform model appraisal. Only linear techniques are used in this case, but do not take into account the real uncertainty existing in a nonlinear inverse problem (Fernández-Martínez et al., 2014b). This setup is valid in two and three dimensions. The most important difference between both resides in the forward problem formulation and on the dimensionality of the model space concerning \mathbf{z} . Examples of 3D basement relief inversion can be consulted in Barbosa et al. (1997), Chakravarthi and Sundararajan (2004, 2007), and Chakravarthi et al. (2002).

The above-mentioned methodology corresponds to the solution of the relief gravity inversion using nonlinear local optimization methods with regularization. The global optimization algorithms used in this paper (a full family of PSO optimizers) do not look for the inverse operator. These algorithms perform smart search of the model space, evaluating the forward problem for different geophysical models, in order to conduct the swarm towards the low misfit regions of the cost function landscape. In the case of nonlinear optimization techniques, regularization techniques are needed to stabilize the inversion, due to the ill-conditioned character of the Jacobian of the forward operator $\mathbf{JF}(\mathbf{m})$, computed in different model candidates. One of the main differences with respect to nonlinear inversion is that regularization is not needed when using global optimization procedures because no inversion is performed. Nevertheless, prior information is still necessary to reduce the set of equivalent solutions and guiding correctly the search.

3. The Particle Swarm Optimization design

Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995) is a global optimization algorithm inspired in the behavior of bird flocks and fish schools searching for food. In this search algorithm, a swarm of particles (or models) explores the space of possible solutions in order to optimize a given cost function, that in our case is related to the inverse problem: the data prediction error shown in expression (2).

In this paper we use a full family of particle swarm optimizers that is fully described in Appendix C. Particularly we introduce the use of RR-PSO and PP-PSO in geophysical inversion. We have used the cloud versions of these algorithms (Fernández-Martínez et al., 2010a), that is, no parameter tuning of the inertia, local and global accelerations is needed, since each of the particles in the swarm has their own PSO parameters that are automatically chosen from a set of PSO parameters that are located in the neighborhood of the corresponding second order stability regions. In our opinion, these mathematical results make PSO to be a very unique algorithm, different from other heuristic approaches.

The use of PSO in applied geophysics has been introduced in Fernández-Álvarez et al. (2006), Shaw and Srivastava (2007) and Fernández-Martínez et al. (2010a,b). Particularly, Fernández-Martínez et al. (2010a) presented the application of Generalized PSO (GPSO), CC-PSO and CP-PSO to the solution and appraisal of a 1D-DC resistivity inverse problem, justifying in each case how to perform the parameter tuning. Nevertheless, in gravity inversion, the PSO algorithm has been barely used. Sanyi et al. (2009) presented some simple synthetic examples using buried cylinders (using the depths and density contrast as unknowns), and compared the results with the ones obtained via Levenberg–Marquardt, genetic algorithm (GA), simulated annealing (SA), and Ant Colony Optimization (ACO) approaches. The conclusion was that PSO outperforms the rest of the algorithms, showing higher convergence rates for PSO and ACO with respect to GA and SA. Tushmalani (2013a,b) used PSO to invert the parameters of a fault (thickness of the sheet, left and right distances to the middle of the sheet, and the fault's angle) using gravity anomalies, comparing the results with the Levenberg–Marquardt method. His general conclusion was that the PSO algorithm provides a better agreement between

the predicted model anomaly and the observations than for the Levenberg–Marquardt method. Sweilam et al. (2008) also compared the PSO algorithm with function stretching (SPSO) to genetic algorithms for synthetic and real (a salt dome) inverse gravimetric problems, finding comparable solutions for both algorithms, but with the advantage for SPSO of having less parameters to tune than the GA method. These papers that are devoted to the use of PSO in gravity inversion as a global optimizer, comparing its performance with classical inversion techniques, miss the most important feature of global optimization algorithms, which is their capability to perform an approximate nonlinear uncertainty analysis (sampling of the nonlinear uncertainty region(s)) of the solution that has been adopted. Also, a clear description about the importance of a correct tuning of the PSO parameters (Fernández-Martínez et al., 2010a,b, 2012b) is usually absent.

3.1. The search space design and the prior information

The first step before PSO sampling consists in computing low and upper bounds for the parameter search space. For the 2D basement relief gravity inverse problem, the search bounds are estimated as follows:

1. An approximate model \mathbf{z}^0 is computed using the observed anomaly and applying to each individual prism location a formulation based on the standard Bouguer's plate correction formula, which is valid in the case of a constant density contrast (Barbosa et al., 1997)

$$z_j^0 = \frac{\Delta g_j^{obs}}{2\pi G \Delta \rho}, \quad (3)$$

where z_j^0 is the initial depth for the j th prism, Δg_j^{obs} is the observed gravity anomaly at the center of the prism (if there is not an observation point at this location, its value can be interpolated), G is the universal gravitational constant, and $\Delta \rho$ is the prescribed constant density contrast.

2. Factors $k_{min} < 1.0$ and $k_{max} > 1.0$ are applied to the approximate model \mathbf{z}^0 in order to obtain the bounds of the searching space as

$$\mathbf{z} \in [k_{min} \cdot \mathbf{z}^0, k_{max} \cdot \mathbf{z}^0], \quad (4)$$

where \mathbf{z}^0 has the same dimensionality as the main unknown of the inverse problem, \mathbf{z} . Typically the number of parameters in a 2D gravity inverse problem is in the order of tens to hundreds. For very high dimensional problems, additional model reduction techniques should be used, as it has been shown for the history matching problem in oil and gas production optimization (Fernández-Martínez et al., 2012b).

Absolute constraints can be easily imposed by setting the lower and upper bounds of the depth for the involved prisms equal to the known value. The way of imposing relative constraints in PSO is different from local optimization techniques (through a regularization term in the cost function), since the use of relative constraints in global optimization is not aimed at stabilizing the inversion as in the case of local optimization methods, but to restrict the search to the family of models that are compatible with the prior information. In the PSO case, a smoothness condition for the basement relief can be achieved via a moving average filter. The size of the window filter might be fixed in the inversion, or can be considered as an additional parameter to be estimated. In this last case, the lower and upper bounds of the window size (in number of grid points) must be given, increasing in one the dimension of the search space. This kind of constraint is always needed, since the models sampled by PSO within the search space in the absence of additional constraints show a typical sawtooth shape, that might not be realistic from the geologic point of view.

3.2. PSO parameter tuning

Parameter tuning (inertia, local and global accelerations) is very important in PSO. Most of the publications use a unique parameter set for all the particles, deduced from the numerical analysis provided by Carlisle and Dozier (2001) and Clerc and Kennedy (2002). Nevertheless it has been shown (Fernández-Martínez and García-Gonzalo, 2008, 2009) that these points are not unique, and their performance is related to the fact that they are close to the upper border of the PSO second order stability limit, where the attraction from the oscillation center is weak, and the explorative behavior of PSO is very high, due to the fact that the variance of the trajectories is almost unbounded.

The algorithms used in this paper belong to the category of free-parameter tuning, since for all the PSO optimizers we have used the corresponding cloud version, where each particle of the swarm has its corresponding values of inertia, local and global accelerations, that are selected in the neighborhood of the upper border of the corresponding second order stability regions (Fernández-Martínez and García-Gonzalo, 2009, 2012). Therefore, the PSO parameters of each particle vary with iterations. Besides each particle (geophysical model) has its corresponding time step Δt , varying between $\Delta t_{min} = 0.8$ and $\Delta t_{max} = 1.2$. The Δt parameter greatly influences the explorative behavior of each PSO member: the exploration of the search space increases when Δt is greater than 1.0. Conversely the algorithm becomes more exploitative when Δt is lower than 1.0.

4. Numerical experiments with synthetic examples

First, the PSO algorithm has been tested for basement relief determination by two different 2D synthetic examples: a noise-free gravimetric inverse problem, and a second example where a white Gaussian noise of $N(0,50)$ μGal (which represents approximately 5 % of the signal) was added to the observed data. The synthetic model of the basin is composed by 50 right rectangular prisms of 100 m width. The maximum depth is located at the center of the basin and reaches 250 m, while the shallower prisms are located at the left and right borders of the basin, and reach a depth of 50 m. The density contrast between the sediments and the basement was $\Delta \rho = -150 \text{ kg/m}^3$ for all the prisms. The Bouguer anomaly was generated, using Eq. (1), at 40 regularly spaced observation points, located at the same height of the prism's upper sides. The search space was generated following Section 3.1 with $k_{min} = 0.5$ and $k_{max} = 2.0$.

4.1. The noise-free case

Different members of the PSO family (GPSO, CC-PSO, CP-PSO, PP-PSO and RR-PSO) have been used for comparative purposes, both in inversion (to analyze their convergence rate) and in model appraisal (explorative analysis). In the present case, only relative constraints were imposed, fixing the size of the average filter window to 15 points, based on different numerical experiments that have been carried out. The size of the swarm was 250 particles, and the number of iterations 150. Fig. 2 shows the real model and its generated anomaly, as well as the initial model \mathbf{z}^0 and the search space $[0.5 \cdot \mathbf{z}^0, 2.0 \cdot \mathbf{z}^0]$.

For comparative purposes the following plots were produced:

1. Convergence rate curves with iterations to compare all the members of the PSO family. Fig. 3 (top) shows that most of the members reach the 2 % equivalence region in less than 10 iterations, that is, around 2500 forward evaluations. GPSO, CC-PSO, and CP-PSO are the most exploitative versions, reaching the minimum misfit solution in less than 40 iterations, while RR-PSO and PP-PSO are more explorative.
2. Dispersion curves, Fig. 3 (bottom), that provide a measure of the swarm collapse with the iterations, given by the median distance in each iteration between the different particles of the swarm and

184

J.L.G. Pallero et al. / Journal of Applied Geophysics 116 (2015) 180–191

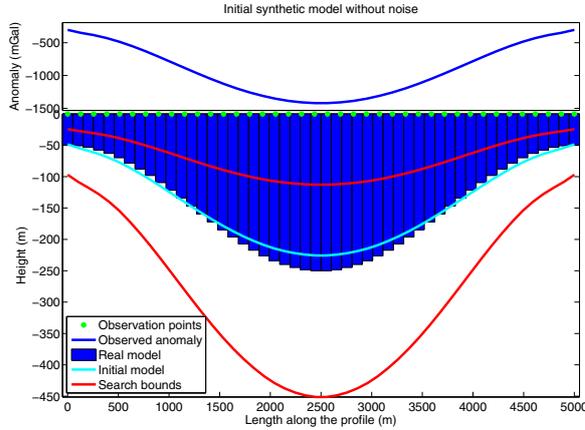


Fig. 2. Synthetic model of the 2D basin for the noise-free case. Top panel: gravity anomaly in μGal , computed at the 40 observation points. Bottom panel: real model, initial model generated from the observed anomaly using expression (3), and lower and upper bounds defining the searching space for the PSO, computed as factors $k_{\min} = 0.5$ and $k_{\max} = 2.0$ of the initial model.

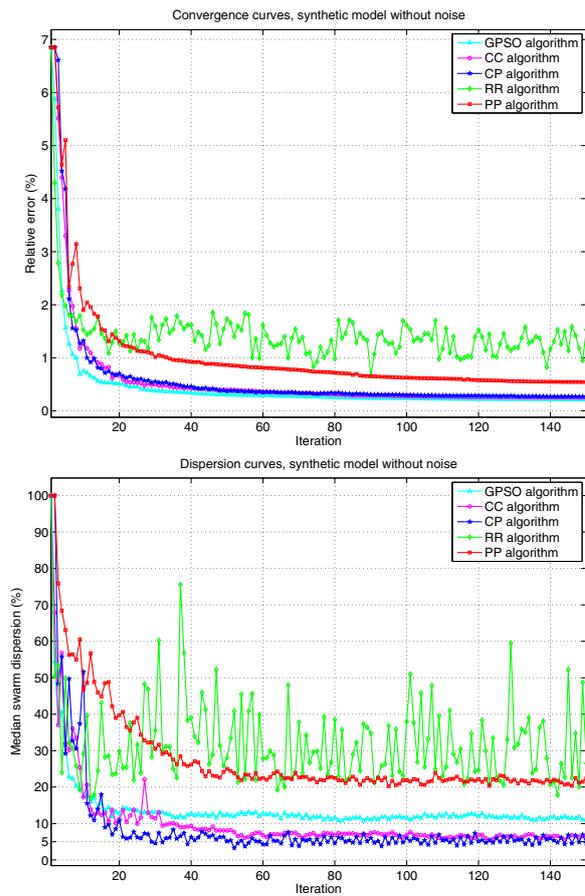


Fig. 3. Convergence (top) and dispersion (bottom) curves for the noise-free case.

their center of gravity, normalized by the median distance in the first iteration (100 % dispersion). Monitoring this plot is important, since the dynamics of the entire swarm in PSO is designed to collapse with iterations towards their corresponding oscillation center. A dispersion lower than 5 % of the initial dispersion would mean that the swarm has collapsed into a unique particle. Therefore, to control the premature collapse of the swarm, exploratory PSO versions, and a correct parameter tuning are needed. It can be observed that RR-PSO and PP-PSO are the most exploratory versions in this case. In the rest of cases, the cloud versions of these algorithms have prevented the swarm to collapse.

- Best model plots and posterior analysis to perform the solution appraisal, by considering all the models \mathbf{z} that fulfill

$$\frac{\|\Delta\mathbf{g}^{obs} - \Delta\mathbf{g}^*(\mathbf{z})\|_2}{\|\Delta\mathbf{g}^{obs}\|_2} < tol. \quad (5)$$

Fig. 4 shows the best estimated models, residuals and the limits of the 10 % equivalence region for the CP-PSO (top), PP-PSO (middle), and RR-PSO (bottom) algorithms. It can be observed that the 3 versions provide similar solutions.

- The interpolated inverse problem cost function topography in the 2D PCA (principal component analysis) space in the region of equivalence, deduced from the models projected into the 2D PCA space. Fig. 8 (top) shows the topography of the cost function deduced in this case from the CP-PSO samples. These graphics will be discussed later in the paper when compared to the noisy case, and serve to visualize in 2D the topography landscape of the cost function in the region of equivalence, that is by construction very high dimensional, and to analyze numerically how the noise affects the inversion. The algorithm to construct the topography of the cost function in the PCA space is as follows:

- Given a set of models $\{m_1, m_2, \dots, m_q\}$ in the non-linear misfit region, sampled via PSO, the experimental covariance matrix $C = (X - \mu)^T \cdot (X - \mu)$ is computed, where X is the matrix that has m_1, m_2, \dots, m_q as column vectors and μ is the model mean of this set of models.
- Diagonalization of experimental covariance matrix C provides an orthonormal set of vectors $B_v = \{v_1, v_2, \dots, v_q\}$.
- Finally, defining the matrix composed by the two first principal components $V = [v_1, v_2]$, the projection into the 2D PCA base is done as follows: $m_{PCA} = V^T \cdot X$. Given the coordinates of these q models in the 2D PCA base, and their corresponding misfit, the cost function is interpolated and plotted as a surface.

4.2. Gaussian noise case

In a second synthetic example a random white noise of normal distribution $N(0,50)$ μGal was added to the observations of the previous example, in order to analyze the performance of the PSO family in the inversion and uncertainty analysis under the effect of noise. The rest of the numerical parameters of the experiment remained unchanged. Fig. 5 shows the real and initial models, the generated anomaly, and the search bounds.

The main conclusions of this analysis are the following:

- The convergence rate curves in Fig. 6 (top) show that most of members reach the 5 % equivalent region in less than 20 iterations, that is, around 5000 forward evaluations. GPSO and CC-PSO are the most explorative versions, since they reach the minimum misfit in fewer iterations than the other algorithms. CP-PSO, RR-PSO and PP-PSO are more explorative.
- The dispersion curves, Fig. 6 (bottom) show again the largest exploratory character of the RR-PSO and PP-PSO algorithms.

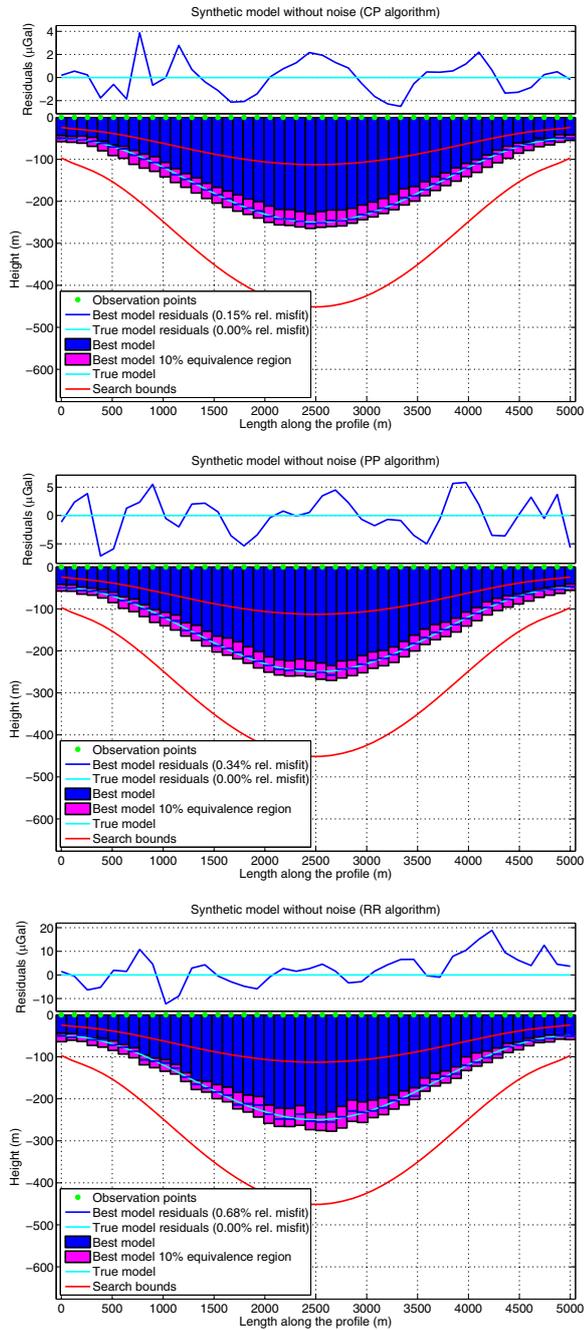


Fig. 4. Model solution for the noise-free case using the CP (top), PP (middle), and RR (bottom) algorithms. Top sub-panels: residual gravity anomaly computed at the 40 observation points. Bottom sub-panels: estimated model and its 10 % equivalence region, true model and search bounds.

3. Fig. 7 shows the best estimated model, the residual, and the posterior analysis performed in the 10 % nonlinear equivalence region for the CP-PSO (top), PP-PSO (middle) and RR-PSO (bottom). All the PSO members provide very good solutions. In this case, RR-PSO has the lowest misfit, but the model shows a slightly sawtooth profile. The

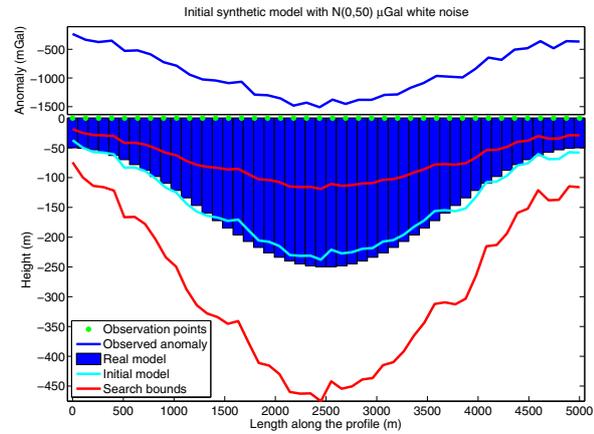


Fig. 5. Synthetic model of the 2D basin with white $N(0,50)$ μGal noise. Top panel: gravity anomaly in μGal , computed at the 40 observation points. Bottom panel: real model, initial model generated from the observed anomaly using expression (3), and lower and upper bounds defining the searching space for the PSO, computed as factors $k_{min} = 0.5$ and $k_{max} = 2.0$ of the approximate model.

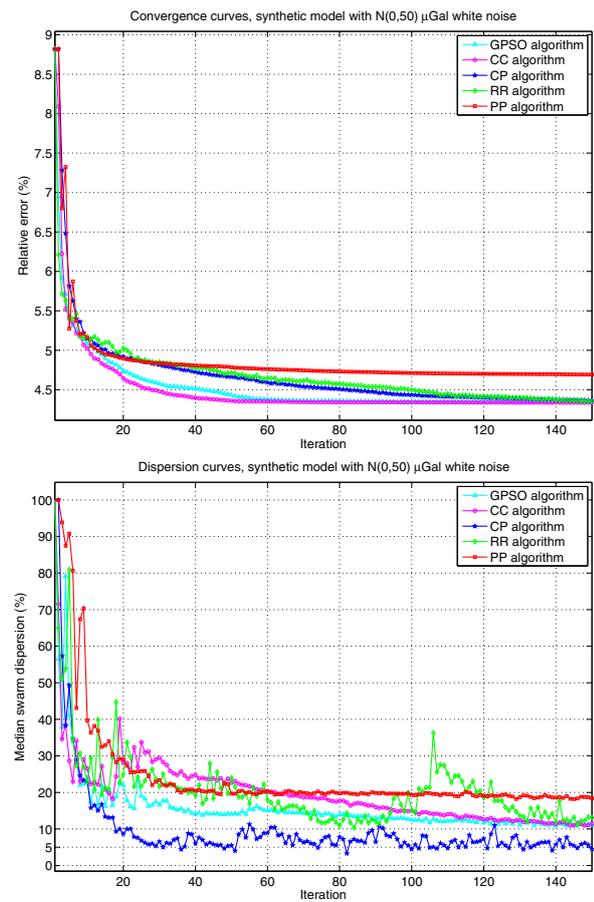


Fig. 6. Convergence (top) and dispersion (bottom) curves for the noisy case.

186

J.L.G. Pallero et al. / Journal of Applied Geophysics 116 (2015) 180–191

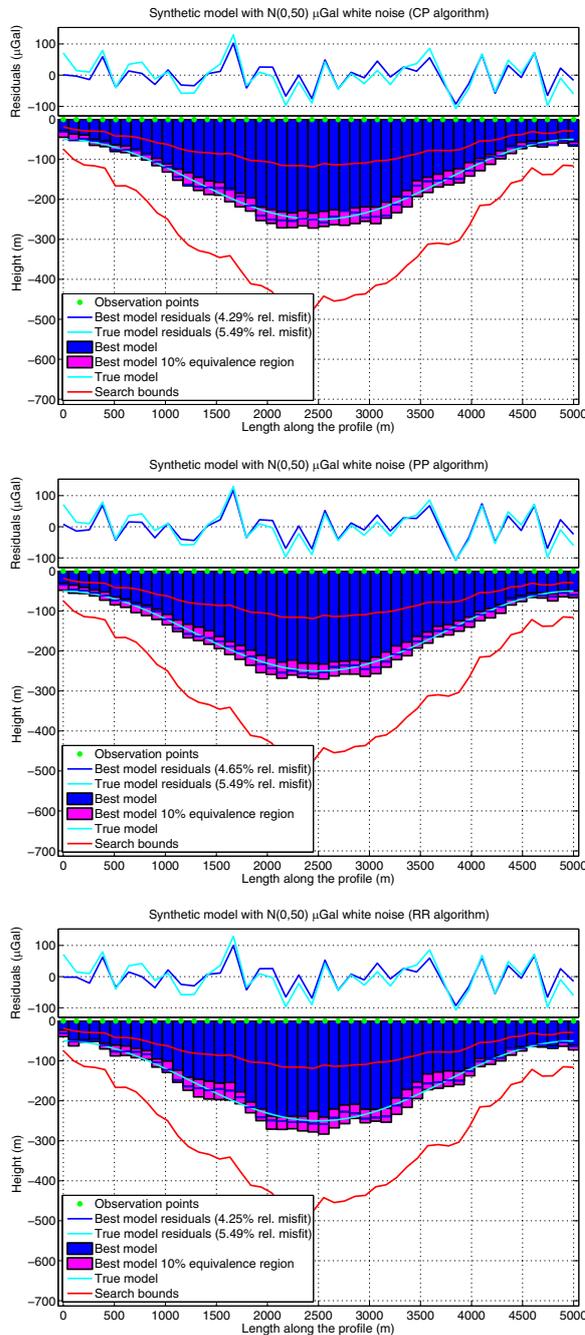


Fig. 7. Model solution of the 2D basin with white $N(0,50)$ μGal noise for CP (top), PP (middle), and RR (bottom) algorithms. Top sub-panels: residual gravity anomaly computed at the 40 observation points. Bottom sub-panels: estimated model and its 10% equivalence region, true model and search bounds.

solutions provided by the other algorithms are smoother.

4. Fig. 8 shows the topography of the cost function deduced from the CP-PSO samples. The top panel shows the noise-free case, and the bottom panel the noisy example. The noise decreases the size of the region of lower misfits and also shifts the location of the best

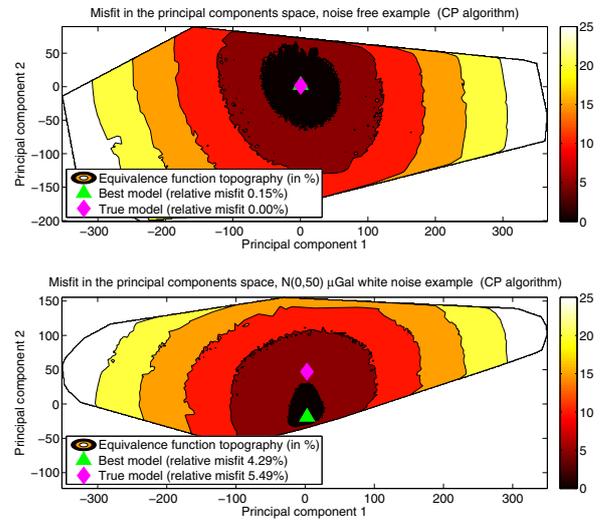


Fig. 8. Cost function topography in the principal components space for the first two principal components of the CP algorithm. Top: noise-free example. Bottom: noisy example. It can be seen how the noise makes smaller the region equivalence that the correspondent of the same value for the noise-free case.

model of lowest misfit with respect to the true model. These facts have been pointed out in Fernández-Martínez et al. (2014a,b) for linear and nonlinear inverse problems. It can also be observed the elliptic shape of the cost function landscape in the PCA plane. This fact implies that the corresponding inverse problem in the reduced PCA space is close to linearity (Fernández-Martínez et al., 2012a). Similar plots were obtained for other family members.

Finally, it can be observed that the relative misfit for the best model found is smaller than the relative misfit for the true model, which is one of the effects of noise in inverse problems (Fernández-Martínez et al., 2012a, 2013, 2014a,b). This fact also highlights the importance of the uncertainty analysis in nonlinear inversion.

5. Real example: Atacama Desert

This section presents the application of PSO to the inversion of a real gravity profile in the Atacama Desert (north Chile). In this region of central Andes, huge amounts of continental sediments (Cenozoic) produced by the erosion of the Andes mountains have filled a central depression previously formed during the built up of the Andean system (Jurassic–Cretaceous) in response to the subduction of the oceanic Nazca plate beneath the South American continent (Isacks, 1988; Mortimer, 1973). These sediments, well known as the Atacama gravel formation, are composed of low density deposits (fluvial gravels, sand and clays intercalated with ignimbrite layers) accumulated on the western flank of the main Andean volcanic arc by large coalescing alluvial fans over the pre-existing andesitic basement. Due to the extremely arid climate in the Atacama region, sedimentary basins are well preserved and their geometries (thickness and basement relief) are considered of particular interest to better understand the tectonic history of this region and the role of uplift and erosion processes in the formation of the Andes (Vernon et al., 2005). The density contrast between these sedimentary fluvial formations and the andesitic basement is around 800 kg/m^3 (Gabalda et al., 2005). Therefore, the study of the Atacama gravel formation represents a good candidate for basement relief estimation via the gravimetric inverse problem.

The gravity data has been acquired during a field survey carried out in central Chile in 2004 (Gabalda et al., 2005) and aimed at estimating

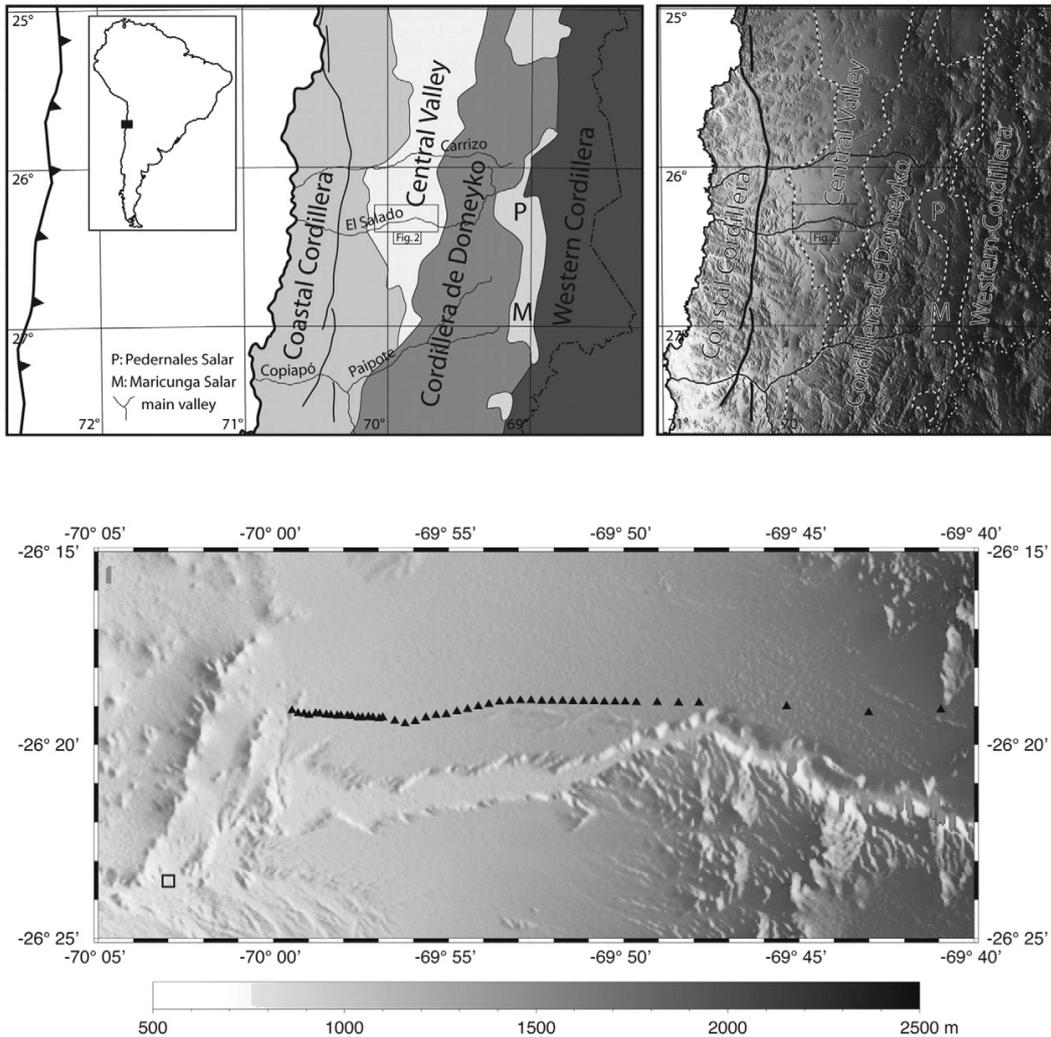


Fig. 9. Location of the test area used for the 2D gravity modeling of the Atacama gravel sedimentary formation (top), and location of the gravity stations (dark triangles) over the shaded SRTM topography (bottom). After Gabalda et al. (2005).

the basement relief of the Atacama gravel formation in El Salado valley (Fig. 9), where irregular sediment thickness was suspected (presence of paleo-valleys). The gravity observations (53 points), acquired along a 30 km long profile using a relative Scintrex CG-3M gravity meter (#9408267), have been processed via CG3TOOL (Gabalda et al., 2003). The station coordinates were determined by differential GPS measurements using full wavelength data (dual-frequency receivers) from a reference base station located in the El Salado valley (profile topography ranged between 1050 m and 1900 m). The accuracies of gravity and GPS measurements are estimated to be better than 20 μGal and 5 cm respectively (Gabalda et al., 2005). Finally, a terrain correction was applied using the SRTM¹ 90 m topographic model and the GeoSoft Oasis Montaj² software. The resulting complete Bouguer anomaly profile is oriented perpendicularly to the main NS orientation of the Andes and is thus dominated by a W–E gravity gradient of

–3.23 mGal/km produced by the gravity effect of the sub-ducting plate. This gradient was estimated jointly with the basin model by the PSO family members.

In order to apply PSO to this gravimetric profile, we have used a gravimetric model composed of 63 prisms—500 m width, which is the approximate mean separation between observations. For the initial model computation, a first approximation of the regional trend was computed and subtracted from the original complete Bouguer anomaly profile in order to isolate the residual anomaly. This residual signal is assumed to be produced by the sedimentary filling of the Atacama gravel formation in a local pre-existing morphological basin. The approximate trend was computed assuming that the anomaly at the extreme points is null, since these ending observation points were located on the outcropping andesitic basement. The initial model was computed using Eq. (3) for a density contrast of -800 kg/m^3 , and the factors $k_{\min} = 0.5$ and $k_{\max} = 2.0$ were applied to create the searching bounds. The numerical parameters used for PSO coincide with previous synthetic experiments, that is, the size of the swarm was 250 individuals, the number of iterations

¹ <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1>.

² <http://www.geosoft.com/products/oasis-montaj/>.

150, and the size of the filtering window was set to 15. The parameters Δg_0 and m of the linear regional trend $\Delta g^{reg}(x) = \Delta g_0 + m \cdot x$, where x is the length along the profile, were also imaged by PSO.

Fig. 10 shows the involved anomaly profiles. The top panel shows the original complete Bouguer anomaly and the middle panel shows the regional trend estimated by PP-PSO. The final estimated slope was -3.23 mGal/km W–E. The bottom panel contains the residual gravity anomaly generated by the estimated basin itself. It reaches a minimum of -7.38 mGal at a point (the 30th km from the profile) at 10.36 km far from the westward extreme. Fig. 11 shows the basin model derived from the inversion of the observed gravity data via PP-PSO. The top panel shows the residuals between the observations generated by the gravimetric model and the original complete Bouguer anomaly (after subtraction of the regional trend), producing a relative misfit of 5.40%. The basin geometry (bottom panel) shows that the maximum filling of low density materials occurred in the westward half of the profile. The analysis of the 10% equivalence region shows that the deepest point lies in the range of [205, 232] m depth from the surface. Using the uncertainty analysis provided by PSO it is possible to establish probability thresholds, such as the probability that the depth of the deepest point in the basin exceeds a certain depth d_c as follows: $P(z_{max} \geq d_c) = N_c/N_t$, where N_c is the number of equivalent models whose maximum depth is greater than d_c , and N_t is the total number of models that have been sampled.

These results are similar to those that were previously published by Gabalda et al. (2005) using the methodology proposed by Cady (1980). Nevertheless these authors did not explore the uncertainty in the inversion of the Atacama profile. From a pure geophysical point of view, it is not straightforward to prove how the evaluation of these uncertainties improves the understanding of the basin. Nevertheless, two important features from this application can be highlighted: (i) a more accurate estimation of the interface is provided here as it takes into account the station elevation into the inversion process (not done in the previous approximation), and (ii) the uncertainties provided here are believed to be very useful as a priori constraints on basin geometry in other numerical modeling of the tectonic, climatic and morphological processes performed to explain the formation and evolution of the sedimentary basins in this area (see for instance Fariás et al., 2008).

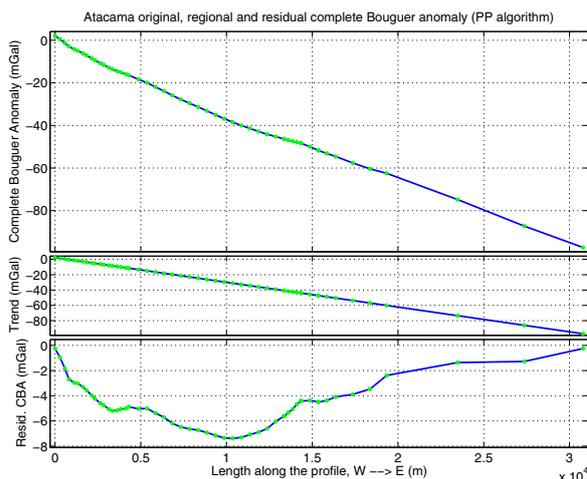


Fig. 10. Complete Bouguer anomaly profiles (West to East). Top panel: original complete Bouguer anomaly in mGal. Middle panel: regional trend $\Delta g^{reg} = 2.358 - 3.228 \cdot X$ mGal, where the slope has units of mGal/km, estimated using the PP-PSO algorithm. Bottom panel: residual anomaly after subtracting the regional trend to the original anomaly. The observation points are depicted along the profiles as the small circles.

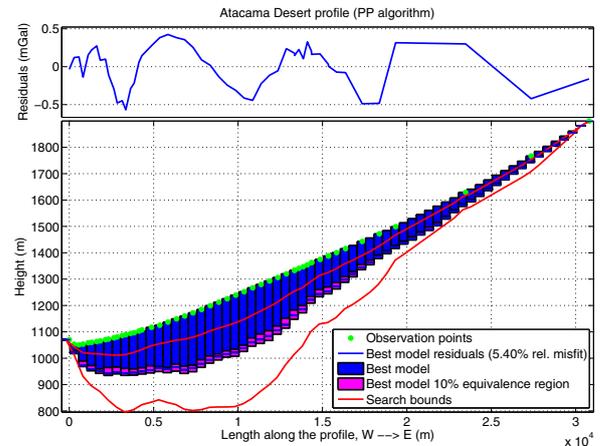


Fig. 11. Atacama 2D profile model. PP algorithm. Top panel: residual gravity anomaly computed at the observation points. Bottom panel: estimated model and its 10% equivalence region, and search bounds.

6. Conclusions

In this paper we presented the design and application of a whole family of PSO optimizers to the 2D gravity inversion and model appraisal (uncertainty assessment) of basement relief estimation in sedimentary basins, using a classical approach consisting in describing the basins as aggregates of rectangular prisms. The performance of these algorithms is first analyzed for two synthetic data sets (noise free and white Gaussian noise cases) showing that all the PSO family members provide similar results for inversion and nonlinear uncertainty analysis when explorative versions are used (sampling while optimizing). All of these algorithms used a free-parameter tuning philosophy, automatically choosing the PSO parameters from regions that are close to their respective second order stability upper limits (cloud versions). Particularly, monitoring their respective dispersion curves is of paramount importance to ensure a correct sampling of the nonlinear uncertainty region. Also the effect of noise in inversion has been analyzed via principal components analysis of the nonlinear equivalent models that have been sampled, illustrating the theoretical results that have been shown in Fernández-Martínez et al. (2014a,b), and urging for the need of a correct nonlinear uncertainty analysis.

Finally, a real profile from Atacama Desert (north Chile) has been inverted using the same methodology, and demonstrating that PSO generates geologically plausible results that are consistent with those that have been obtained by other well known classical approaches, but without uncertainty assessment. Compared to such techniques already proposed for 2D nonlinear gravity inversion, the PSO technique proposed here appears as a powerful tool for estimating the basement relief of sedimentary basins and their uncertainties in a fast³ optimized procedure. Moreover, as the inversion procedure takes into account the actual topography of the gravity observations (i.e. calculations are performed at the station elevation) and enables a separation of regional/residual anomalies, it is believed to be of special interest in many geological and geophysical applications requiring accurate estimation of basement relief of sedimentary basins.

This work is a first approximation to a future free Matlab-based package for sedimentary basin gravity inversion using PSO.

³ The PSO, implemented as a Matlab program, took for the real example around 2 min of processing time in a laptop equipped with an Intel Core i7-4800MQ 2.70 GHz processor, running a Debian GNU/Linux operating system.

Acknowledgments

JLGP acknowledges the support of the École des Mines d'Albi-Carmaux, that allowed him to develop this research in Albi during several months in 2014, and also the collaboration and help of the Laboratoire GET (Université de Toulouse).

We would also thank Germinal Gabalda and Thierry Nalpas for their contribution to the acquisition of the gravity data used in this study.

Appendix A. Gravity attraction of a 2D right rectangular prism with constant density

Let a right rectangular 2D prism of constant density ρ , defined by its extreme coordinates x minimum x_m , x maximum x_M , z of the top side z_t , and z of the bottom side z_b . The gravity attraction generated by this prism at a point $P(x,y)$ can be computed (Barbosa and Silva, 1994; Telford et al., 1976) as

$$F = F(x, z, x_m, x_M, z_t, z_b, \rho) \\ = G\rho \left[A \ln \frac{A^2 + D^2}{A^2 + C^2} - B \ln \frac{B^2 + D^2}{B^2 + C^2} - 2C \left(\arctan \frac{A}{C} - \arctan \frac{B}{C} \right) \right. \\ \left. + 2D \left(\arctan \frac{A}{D} - \arctan \frac{B}{D} \right) \right], \quad (\text{A.1})$$

where $A = x - x_m$, $B = x - x_M$, $C = z - z_t$, $D = z - z_b$, and G is the Newton's constant. A schematic view of the situation can be seen in Fig. A.1.

Although this formulation is stated for a constant density, other density distributions, such as those presented in Table 1, could be used if the original prism is discretized into sub-prisms, each of which with its correspondent density contrast computed by the right equation. Then, Eq. (A.1) can be applied to each sub-prism of constant density and the final attraction computed as the sum of each individual attractions. Also, this computation strategy could be used taking into account the known density values from boreholes. The size of the sub-prisms must be chosen such that the error in the approximation is less than a specified tolerance.

Appendix B. Relative and absolute constraints in the right-rectangular prism model

The inverse problem stated in Eq. (2), consisting in minimizing the prediction error, is ill posed, since there exist an infinity of solutions that fit the observed data within the same error bounds. Also, the determination of the solution with minimum misfit is very unstable. Different kinds of constraints are commonly used to win stability at inversion,

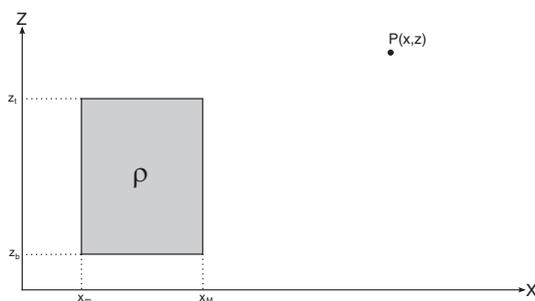


Fig. A.1. Two dimensional schematic situation of a right rectangular prism and an attracted point $P(x,y)$. ρ : density, x_m : minimum x coordinate, x_M : maximum x coordinate, z_t : top side z coordinate, and z_b : bottom side z coordinate.

and reducing the set of solutions to those that are compatible with the prior information.

Relative constraints (Barbosa et al., 1997; Silva et al., 2006, 2009) consist in imposing smoothness conditions of some degree to the basement relief. This smoothness is introduced via the regularization term

$$c^r(\mathbf{z}) = \frac{1}{M-1} \sum_{j=1}^{M-1} (z_j - z_{j+1})^2. \quad (\text{B.1})$$

A more general formulation of Eq. (B.1) is

$$c^r(\mathbf{z}) = \frac{1}{L} \|\mathbf{R}\mathbf{z}\|_2^2, \quad (\text{B.2})$$

where \mathbf{R} is a $L \times M$ weighting matrix for the regularity conditions to be imposed on the model parameters. Expression (B.2) includes the case of Tikhonov's regularizations of order 1 and 2, to impose regularity conditions on the first and second order derivatives of the model \mathbf{z} .

Absolute constraints (Barbosa et al., 1997) consist in fixing the depth value of some prisms to prescribed values based on borehole or seismic profile information. The corresponding regularization term is

$$c^a(\mathbf{z}) = \frac{1}{K} \|\mathbf{z}^{obs} - \mathbf{A}\mathbf{z}\|_2^2, \quad (\text{B.3})$$

where \mathbf{A} is a $K \times M$ (with $K < M$) zeros and ones matrix, where 1 indicates the positions of the prisms whose depth has to be fixed, and \mathbf{z}^{obs} is a vector containing the observed depths.

The final cost function with both regularization terms is written as

$$c(\mathbf{z}) = c^g(\mathbf{z}) + \mu^r c^r(\mathbf{z}) + \mu^a c^a(\mathbf{z}), \quad (\text{B.4})$$

where μ^r and μ^a are regularization parameters which determine the trade-off between the different terms in the cost function (prediction error and relative and absolute constraints).

Appendix C. The family of PSO optimizers

PSO applied to optimization problems is very simple: individuals, or particles, are represented by vectors whose length is the number of degrees of freedom of the optimization problem. To start, a population of particles is initialized with random positions \mathbf{x}^0 and velocities \mathbf{v}^0 . The same objective function (Eq. (2)) is used to compute the fitness (or prediction error) of each particle in the swarm. As time advances, the position and velocity of each particle are updated as a function of its fitness and the fitness values of its neighbors. At time-step $k + 1$, the algorithm updates positions \mathbf{x}^{k+1} and velocities \mathbf{v}^{k+1} of the individuals as

$$\mathbf{v}_i^{k+1} = \omega \mathbf{v}_i^k + \phi_1 (\mathbf{g}^k - \mathbf{x}_i^k) + \phi_2 (\mathbf{I}_i^k - \mathbf{x}_i^k), \quad (\text{C.1}) \\ \mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1},$$

where \mathbf{g}^k is the global best position in the whole swarm, \mathbf{I}_i^k is the i th particle's best position, $\phi_1 = r_1 a_g$ and $\phi_2 = r_2 a_l$ are the random global and local accelerations, and ω is a real constant called inertia weight. Finally, r_1 and r_2 are random variables uniformly distributed in $(0,1)$, to weight the global and local acceleration constants a_g and a_l . Tuning of the PSO parameters implies providing suitable values for the inertia constant ω and for the local and global accelerations a_g and a_l , in order to achieve exploration and also convergence towards the low misfit regions.

The PSO algorithm, as it was originally proposed in Eq. (C.1), was not dimensionally correct, since velocities and accelerations, and positions and velocities are directly added. The solution to this simple enigma came from physical analysis. PSO can be interpreted as a double stochastic gradient algorithm in the model space, and is the particular

case of the generalized PSO (GPSO) algorithm (Fernández-Martínez and García-Gonzalo, 2008) for $t = k$ and $\Delta t = 1$:

$$\begin{aligned} \mathbf{v}(t + \Delta t) &= \mathbf{v}(t)[1 - (1 - \omega)\Delta t] \\ &\quad + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] \\ &\quad + \phi_2 \Delta t [\mathbf{I}(t) - \mathbf{x}(t)], \\ \mathbf{x}(t + \Delta t) &= \mathbf{x}(t) + \mathbf{v}(t + \Delta t)\Delta t. \end{aligned} \quad (\text{C.2})$$

The GPSO algorithm was derived using a mechanical analogy: a damped mass-spring system with unit mass, damping factor $1 - \omega$, and total stiffness constant $\phi = \phi_1 + \phi_2$, i.e., the so-called PSO continuous model

$$\begin{cases} \mathbf{x}''(t) + (1 - \omega)\mathbf{x}'(t) + \phi\mathbf{x}(t) = \phi_1\mathbf{g}(t) + \phi_2\mathbf{I}(t), \\ \mathbf{x}(0) = \mathbf{x}_0, \\ \mathbf{x}'(0) = \mathbf{v}_0, \end{cases} \quad (\text{C.3})$$

where $\mathbf{x}(t), t \in \mathbb{R}$ is the position of any particle of the swarm in continuous time. In this model, particles interact through the local and global attractors $\mathbf{I}(t)$ and $\mathbf{g}(t)$, and mean particle trajectories oscillate around the particle position (Fernández-Martínez and García-Gonzalo, 2008)

$$\mathbf{o}_i(t) = \frac{\phi_1\mathbf{g}(t) + \phi_2\mathbf{I}(t)}{\phi_1 + \phi_2}. \quad (\text{C.4})$$

The full stochastic analysis of the PSO continuous and discrete models (GPSO) was performed in a collection of 3 papers (Fernández-Martínez and García-Gonzalo, 2008, 2009, 2011a). This analysis shed light about the relationship between PSO convergence and the PSO parameter tuning: good PSO parameters (ω, a_g and a_l) are located in the neighborhood of the upper border of the second order stability region. Therefore, no magical tuning points with five decimal figures exist. The answer is simpler than that: stochastic stability and convergence are intimately related in the PSO case. Using the above-mentioned mechanical analogy a full family of particle swarm optimizers was also derived (Fernández-Martínez and García-Gonzalo, 2009, 2012; García-Gonzalo et al., 2014). Numerical analysis using different analytical benchmark functions has shown that the most performing algorithm of the family in terms of the balance of its exploration/exploitation capabilities was RR-PSO, that is the PSO member family obtained by adopting both regressive finite difference schemes of the PSO continuous model (Eq. (C.3)) for the acceleration and the velocity.

The RR-PSO algorithm can be written as

$$\begin{cases} \mathbf{v}(t + \Delta t) = \frac{\mathbf{v}(t) + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{I}(t) - \mathbf{x}(t)]}{1 + (1 - \omega)\Delta t + \phi \Delta t^2}, \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t)\Delta t, \\ \mathbf{x}(0) = \mathbf{x}_0, \\ \mathbf{v}(0) = \mathbf{v}_0, \end{cases} \quad (\text{C.5})$$

where $(t, \Delta t) \in \mathbb{R}$. This algorithm has regions of first and second order stochastic stability that are unbounded. It has been numerically shown that good parameter sets in the RR-PSO case are concentrated around the line $\bar{\phi} = 3(\omega - 3/2)$, mainly for inertia values $\omega > 2$. This line is independent of the cost function that is optimized, and remains invariant when the number of optimization parameters increases. Furthermore, this line is located in a region of medium attenuation and very high frequency for the swarm particle trajectories. This last property confers to RR-PSO a good balance between exploration and exploitation, allowing for a very efficient and explorative search around the oscillation center of each particle in the swarm.

Two other explorative members of the PSO family are also used in this paper (Fernández-Martínez and García-Gonzalo, 2009, 2012):

• The CP-PSO:

$$\begin{cases} \mathbf{v}(t + \Delta t) = \frac{(1 + \phi \Delta t^2)\mathbf{v}(t) + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{I}(t) - \mathbf{x}(t)]}{1 + (1 - \omega)\Delta t}, \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t. \end{cases} \quad (\text{C.6})$$

• The PP-PSO:

$$\begin{cases} \mathbf{v}(t + \Delta t) = [1 - (1 - \omega)\Delta t]\mathbf{v}(t) \\ \quad + \phi_1 \Delta t [\mathbf{g}(t) - \mathbf{x}(t)] + \phi_2 \Delta t [\mathbf{I}(t) - \mathbf{x}(t)], \\ \mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t. \end{cases} \quad (\text{C.7})$$

PP-PSO has the same velocity rule than PSO, but the positions of the particles are written in time t , instead of $t + 1$. This small difference has big implications in the stability analysis.

References

- Adema, G.W., Breckenridge, R.M., Sprengle, K.F., 2007. Gravity, morphology, and bedrock depth of the Rathdrum Prairie, Idaho. Tech. Rep., Idaho Geological Survey. University of Idaho.
- Barbosa, V.C.F., Silva, J.B.C., 1994. Generalized compact gravity inversion. *Geophysics* 95 (1), 57–68.
- Barbosa, V.C.F., Silva, J.B.C., Medeiros, W.E., 1997. Gravity inversion of basement relief using approximate equality constraints on depths. *Geophysics* 62 (6), 1745–1757.
- Blakely, R.J., 1995. *Potential Theory in Gravity and Magnetic Applications*. Cambridge University Press.
- Bohidar, R.N., Sullivan, J.P., Hermance, J.F., 2001. Delineating depth to bedrock beneath shallow unconfined aquifers: a gravity transect across the Palmer river basin. *Ground Water* 39 (5), 729–736.
- Bott, M.H.P., 1960. The use of rapid digital computing methods for direct gravity interpretation of sedimentary basins. *Geophys. J. R. Astron. Soc.* 3, 63–67.
- Cady, J.W., 1980. Calculation of gravity and magnetic anomalies of finite length-right polygonal prisms. *Geophysics* 45 (10), 1507–1512.
- Carlisle, A., Dozier, G., 2001. An Off-The-Shelf PSO. *Proceedings of the Particle Swarm Optimization Workshop*.
- Chakravarthi, V., 1995. Gravity interpretation of nonoutcropping sedimentary basins in which the density contrast decreases parabolically with depth. *Pure Appl. Geophys.* 145 (2), 327–335.
- Chakravarthi, V., Sundararajan, N., 2004. Automatic 3-D gravity modeling of sedimentary basins with density contrast varying parabolically with depth. *Comput. Geosci.* 30 (6), 601–607.
- Chakravarthi, V., Sundararajan, N., 2007. 3D gravity inversion of basement relief—a depth-dependent density approach. *Geophysics* 72 (2), 123–132.
- Chakravarthi, V., Singh, S.B., Babu, G.A., 2001. INVER2DBASE—a program to compute basement depths of density interfaces above which the density contrast varies with depth. *Comput. Geosci.* 27 (10), 1127–1133.
- Chakravarthi, V., Raghuram, H.M., Singh, S.B., 2002. 3-D forward gravity modeling of basement interfaces above which the density contrast varies continuously with depth. *Comput. Geosci.* 28 (1), 53–57.
- Chapell, A., Kusznir, N., 2008. An algorithm to calculate the gravity anomaly of sedimentary basins with exponential density–depth relationships. *Geophys. Prospect.* 56, 249–258.
- Clerc, M., Kennedy, J., 2002. The particle swarm—exploration, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* 6 (1), 58–73.
- Cordell, L., 1973. Gravity analysis using an exponential density–depth function—San Jacinto Graben, California. *Geophysics* 38, 684–690.
- Dobrin, M.B., 1960. *Introduction to Geophysical Prospecting*. McGraw-Hill.
- Farias, M., Charrier, R., Carretier, S., Martinod, J., Fock, A., Campbell, D., Cáceres, J., Comte, D., 2008. Late Miocene high and rapid surface uplift and its erosional response in the Andes of central Chile (33°–35°S). *Tectonics* 27 (1).
- Fernández-Álvarez, J.P., Fernández-Martínez, J.L., García-Gonzalo, E., Menéndez-Pérez, C.O., 2006. Application of the particle swarm optimization algorithm to the solution and appraisal of the vertical electrical sounding inverse problem. 11th Annual Conference of the International Association of Mathematical Geology (IAMG'06), Liège, Belgium.
- Fernández-Álvarez, J.P., Fernández-Martínez, J.L., Menéndez-Pérez, C.O., 2008. Feasibility analysis of the use of binary genetic algorithms as importance samplers application to a 1-D DC resistivity inverse problem. *Math. Geosci.* 40 (4), 375–408.
- Fernández-Martínez, J.L., García-Gonzalo, E., 2008. The generalized PSO: a new door for PSO evolution. *J. Artif. Evol. Appl.* 15.
- Fernández-Martínez, J.L., García-Gonzalo, E., 2009. The PSO family: deduction, stochastic analysis and comparison. *Swarm Intell.* 3, 245–273.
- Fernández-Martínez, J.L., García-Gonzalo, E., 2011a. Stochastic stability analysis of the linear continuous and discrete PSO models. *IEEE Trans. Evol. Comput.* 15 (3), 405–423.
- Fernández-Martínez, J.L., García-Gonzalo, E., 2011b. What Makes Particle Swarm Optimization a Very Interesting and Powerful Algorithm. Springer.

- Fernández-Martínez, J.L., García-Gonzalo, E., 2012. Stochastic stability and numerical analysis of two novel algorithms of PSO family: PP-PSO and RR-PSO. *Int. J. Artif. Intell. Tools* 21 (3), 20.
- Fernández-Martínez, J.L., García-Gonzalo, E., Fernández Álvarez, J.P., Kuzma, H.A., Menéndez Pérez, C.O., 2010a. PSO: a powerful algorithm to solve geophysical inverse problems. Application to a 1D-DC resistivity case. *J. Appl. Geophys.* 71, 13–25.
- Fernández-Martínez, J.L., García-Gonzalo, E., Naudet, V., 2010b. Particle swarm optimization applied to solving and appraising the streaming-potential inverse problem. *Geophysics* 75 (4), WA3–WA15.
- Fernández-Martínez, J.L., Fernández-Muñiz, M.Z., Tompkins, M.J., 2012a. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* 77 (1), W1–W15.
- Fernández-Martínez, J.L., Mukerji, T., García-Gonzalo, E., Suman, A., 2012b. Reservoir characterization and inversion uncertainty via a family of particle swarm optimizers. *Geophysics* 77 (1), M1–M16.
- Fernández-Martínez, J.L., Fernández-Muñiz, Z., Pallero, J.L.G., Pedruelo-González, L.M., 2013. From Bayes to Tarantola: new insights to understand uncertainty in inverse problems. *J. Appl. Geophys.* 98, 62–72.
- Fernández-Martínez, J.L., Pallero, J.L.G., Fernández-Muñiz, Z., Pedruelo-González, L.M., 2014a. The effect of noise and Tikhonov's regularization in inverse problems. Part I: the linear case. *J. Appl. Geophys.* 108, 176–185.
- Fernández-Martínez, J.L., Pallero, J.L.G., Fernández-Muñiz, Z., Pedruelo-González, L.M., 2014b. The effect of noise and Tikhonov's regularization in inverse problems. Part II: the nonlinear case. *J. Appl. Geophys.* 108, 186–193.
- Gabalda, G., Bonvalot, S., Hipkin, R., 2003. CG3TOOL: an interactive computer program for Scintrex CG-3M gravity data processing. *Comput. Geosci.* 29 (2), 155–171.
- Gabalda, G., Nalpas, T., Bonvalot, S., 2005. The Base of the Atacama Gravels Formation (26°S, Northern Chile): first results from gravity data. 6th International Symposium on Andean Geodynamics (ISAG 2005, Barcelona). IRD, Paris, pp. 286–289 (Extended Abstracts).
- García-Gonzalo, E., Fernández-Martínez, J.L., 2014. Convergence and stochastic stability analysis of particle swarm optimization variants with generic parameter distributions. *Appl. Math. Comput.* 249, 286–302.
- García-Gonzalo, E., Fernández-Martínez, J.L., Cernea, A., 2014. Four-points particle swarm optimization algorithms. *J. Mult. Valued Logic Soft Comput.* 22 (3), 239–266.
- Granser, H., 1987. Three-dimensional interpretation of gravity data from sedimentary basins using an exponential density–depth function. *Geophys. Prospect.* 35, 1030–1041.
- Isacks, B.L., 1988. Uplift of the Central Andean Plateau and bending of the Bolivian Orocline. *J. Geophys. Res.* 93 (B4), 3211–3231.
- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. *IEEE International Conference on Neural Networks* Vol. 4, pp. 1942–1948.
- Krimmel, R.M., 1970. Gravimetric ice thickness determination, South Cascade Glacier, Washington. *Northwest Sci.* 44 (3), 147–153.
- Mantlik, F., Matias, M., Lourenço, J., Grangeia, C., Tareco, H., 2009. The use of gravity methods in the internal characterization of landfills—a case study. *J. Geophys. Eng.* 6, 357–364.
- Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* 11 (2), 431–441.
- Martín Aienza, B., 2001. Modelado e inversión en 2D y 3D de anomalías gravimétricas producidas por cuerpos cuya geometría y densidad de masa se describen utilizando funciones polinómicas: aplicaciones a datos gravimétricos de Canadá y México. (Ph.D. thesis, Universidad Complutense de Madrid, Facultad de Ciencias Físicas. Departamento de Física de la Tierra, Astronomía y Astrofísica I).
- Mortimer, C., 1973. The Cenozoic history of the southern Atacama Desert, Chile. *J. Geol. Soc.* 129 (5), 505–526.
- Nettleton, L.L., 1976. *Gravity and Magnetics in Oil Prospecting*. McGraw-Hill.
- Nocedal, J., Wright, S.J., 1999. *Numerical Optimization*. 1st edition. Springer, New York.
- Parker, R.L., 1994. *Geophysical Inverse Theory*. Princeton University Press.
- Rao, D.B., 1990. Analysis of gravity anomalies of sedimentary basins by an asymmetrical trapezoidal model with quadratic density function. *Geophysics* 55 (2), 226–231.
- Rao, C.V., Raju, M.L., Chakravarthi, V., 1995. Gravity modelling of an interface above which the density contrast decreases hyperbolically with depth. *J. Appl. Geophys.* 34 (1), 63–67.
- Sanyi, Y., Shangxu, W., Nan, T., 2009. Swarm intelligence optimization and its application in geophysical data inversion. *Appl. Geophys.* 6 (2), 166–174.
- Scales, J.A., Snieder, R., 2000. The anatomy of inverse problems. *Geophysics* 65 (6), 1708–1710.
- Shaw, R., Srivastava, S., 2007. Particle swarm optimization: a new tool to invert geophysical data. *Geophysics* 72, F75–F83.
- Shumskiy, P.A., 1960. Density of glacier ice. *J. Glaciol.* 3 (27), 568–573.
- Silva, J.B.C., Costa, D.C.L., Barbosa, V.C.F., 2006. Gravity inversion of basement relief and estimation of density contrast variation with depth. *Geophysics* 71 (5), J51–J58.
- Silva, J.B.C., Teixeira, W.A., Barbosa, V.C.F., 2009. Gravity data as a tool for landfill study. *Environ. Geol.* 57, 749–757.
- Silva, J.B.C., Oliveira, A.S., Barbosa, V.C.F., 2010. Gravity inversion of 2D basement relief using entropic regularization. *Geophysics* 75 (3), I29–I35.
- Stern, T.A., 1978. Gravity Survey of the Taylor Glacier, Victoria Land, Antarctica. Tech. Rep. 8, Geology Department, Victoria University of Wellington.
- Sweilam, N.H., Gobashy, M.M., Hashem, T., 2008. Using particle swarm optimization with function stretching (SPSO) for integrating gravity data: a visibility study. *Proc. Math. Phys. Soc. Egypt* 86 (2), 259–281.
- Telford, W.M., Geldart, L.P., Sheriff, R.E., Keys, D.A., 1976. *Applied Geophysics*. Cambridge University Press.
- Toushmalani, R., 2013a. Comparison result of inversion of gravity data of a fault by particle swarm optimization and Levenberg–Marquardt methods. *SpringerPlus* 2, 462.
- Toushmalani, R., 2013b. Gravity inversion of a fault by Particle Swarm Optimization (PSO). *SpringerPlus* 2, 315.
- Venteris, E., Miller, M., 1993. Gravitational Profiles on the Taku Glacier System. *Glaciological and Arctic Sciences Institute, University of Idaho, Open File Report*.
- Vernon, A., Nalpas, T., Dabard, M.-P., Mpodozis, C., Riquelme, R., Hérail, G., 2005. Preservation of the Miocene Atacama gravels: climatic/depositional–erosional balance in the El Salvador area, North Chilean Andes. 6th International Symposium on Andean Geodynamics (ISAG 2005, Barcelona). IRD, Paris, pp. 783–786 (Extended Abstracts).

ARTICLE 5

**3D gravity inversion and uncertainty
assessment of basement relief via
Particle Swarm Optimization**

(SUBMITTED)

Authors:

José Luis García Pallero
Juan Luis Fernández Martínez
Sylvain Bonvalot
Olivier Fudym

JOURNAL OF APPLIED GEOPHYSICS
2016, VOL. XXX, PP. XXX–XXX
doi:[XXXXXX](#)

Summary

In this last article, the Particle Swarm Optimization (PSO) algorithm is applied to the 3D gravity inversion problem in sedimentary basins. The main differences between this work and the one presented in article 4 are the increase of the problem dimensions and a specific treatment of the density contrast for taking into account its horizontal and depth variations. As for the 2D case, this is the first time that the PSO algorithm is applied to this kind of problem. The principal features of this research work are:

- A regional trend adjustment during the inversion is added to the basin modeling, which is composed by a set of right rectangular prisms. The horizontal and depth-dependence of the density contrast can be taken into account.
- A filtering window is introduced, which acts as relative constraints, acting as a generalization of the first order Tikhonov regularization in local optimization methods.
- The search space construction is explained and the possibility of introduction of absolute constraints is discussed, comparing it with the same task used in local optimization algorithms. As in the 2D case, the GPSO, CC-PSO, CP-PSO, PP-PSO and RR-PSO family members are used.
- Synthetic examples, without and with noise are solved, in both cases with horizontal and vertical density contrast variation. The obtained solutions and uncertainty regions are compared with the real models, concluding that PSO is a valid method for the solution of the 3D gravity basement relief inversion problem in sedimentary basins. Similarly to the 2D case, it has been shown numerically the effect of the noise in the topography cost function.
- A real problem with data from the Argelès-Gazost basin (France) is solved. The geophysical model that has been obtained is consistent from a geological point of view and agree with the results obtained via classical inversion. The PSO method adds to the problem the uncertainty analysis of the obtained solution. It is then demonstrated that the developed technique is a valid method for the 3D gravity inversion problem in sedimentary basins.

Subject: Your recent submission to APPGEO
From: "J.Appl. Geophysics" <appgeo-eo@elsevier.com>
Date: 04/03/16 13:05
To: jlg.pallero@upm.es

Dear Dr. J.L.G. Pallero,

You have been listed as a Co-Author of the following submission:

Journal: Journal of Applied Geophysics
Corresponding Author: Juan Luis Fernandez Martinez
Co-Authors: J.L.G. Pallero, MSc; Sylvain Bonvalot, Professor; Olivier Fudym, Professor
Title: 3D gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization

If you did not co-author this submission, please contact the Corresponding Author of this submission at jlfmuniovi@gmail.com; jlfm@uniovi.es; do not follow the link below.

An Open Researcher and Contributor ID (ORCID) is a unique digital identifier to which you can link your published articles and other professional activities, providing a single record of all your research.

We would like to invite you to link your ORCID ID to this submission. If the submission is accepted, your ORCID ID will be linked to the final published article and transferred to CrossRef. Your ORCID account will also be updated.

To do this, visit our dedicated page in EES. There you can link to an existing ORCID ID or register for one and link the submission to it:

<http://ees.elsevier.com/appgeo/l.asp?i=29617&l=UY060TCR>

More information on ORCID can be found on the ORCID website, <http://www.ORCID.org>, or on our help page: http://help.elsevier.com/app/answers/detail/a_id/2210/p/7923

Like other Publishers, Elsevier supports ORCID - an open, non-profit, community based effort - and has adapted its submission system to enable authors and co-authors to connect their submissions to their unique ORCID IDs.

Thank you,

Journal of Applied Geophysics

3D gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization

J.L.G. Pallero^a, J.L. Fernández-Martínez^{b,*}, S. Bonvalot^c, O. Fudym^d

^a*ETSI en Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid. Madrid, Spain*

^b*Departamento de Matemáticas. Universidad de Oviedo. Oviedo, Spain*

^c*Laboratoire GET (Université de Toulouse, CNRS, IRD, CNES) - Bureau Gravimétrique International (BGI). Toulouse, France*

^d*Université de Toulouse, Mines Albi, CNRS, Centre RAPSODEE. Albi cedex, France*

Abstract

Nonlinear gravity inversion in sedimentary basins is a classical problem in applied geophysics. Although a 2D approximation is widely used, 3D models have been also proposed to take better into account the basin geometry. A common nonlinear approach to this 3D problem consists in modeling the basin as a set of right rectangular prisms with prescribed density contrast, whose depths are the unknowns. Then, the problem is iteratively solved via local optimization techniques from an initial model computed using some simplifications or being estimated using prior geophysical models. Nevertheless, this kind of approach is highly dependent on the prior information that is used, and lacks from a correct solution appraisal (nonlinear uncertainty analysis). In this paper, we show the application of Particle Swarm Optimization (PSO) to the 3D gravity inversion and model appraisal of the solution that is adopted for basement relief estimation in sedimentary basins. Synthetic and real cases are illustrated, showing that robust results are obtained. Therefore, PSO seems to be a very good alternative for 3D gravity inversion and uncertainty assessment of basement relief when used in a sampling while optimizing modality. That way important geological questions can be answered probabilistically in order to perform risk assessment in the decisions that are made.

Keywords: Nonlinear gravity inversion, Particle Swarm Optimization, Uncertainty assessment, Sedimentary basin, Pyrenees-Argelès-Gazost

1. Introduction

Gravity in geophysics is a commonly used tool to analyze the Earth crust structure, in mineral exploration, oil and gas upstream activities, hydrogeology, glaciology, etc. (Dobrin, 1960; Nettleton, 1976; Telford et al., 1976; Parker, 1994; Hinze et al., 2013). Gravity inversion is defined as a procedure that constructs a model of subsurface density variations or geometries from gravity measured data and other prior information (Nabighian et al., 2005). The gravity inverse problem can be stated as linear (if the geometry is provided and the densities are estimated) or nonlinear (if the densities are postulated and the geometry is treated as unknown).

Bottom relief determination in sedimentary basins is a common task (Silva et al., 2006; Chakravarthi and Sundararajan, 2007) as a former step for different posterior studies as oil and gas exploration (Silva et al., 2010), aquifers structure (Adema et al., 2007; Bohidar et al., 2001), etc. Although a 2D approximation to the inverse problem is frequently used in practice (Chakravarthi et al., 2001; Silva et al., 2006, 2009), there have been also proposed 3D methods that produce more realistic

models (Barbosa et al., 1997; Chakravarthi and Sundararajan, 2007).

Estimation of basement relief based on observed gravity anomalies could be a high dimensional nonlinear inverse problem, especially for 3D approaches. Most resolution methods are based on local optimization techniques, such as the well-known Levenberg-Marquardt's algorithm with Tikhonov's regularization: stating a prior model for the densities and (optionally) a set of constraints, a solution is achieved through iterative linearization of the cost function. This procedure provides a solution that is highly dependent on the initial model and on the prior information that are used. Besides, no model appraisal in the nonlinear sense is usually performed on the solution that has been found. The fact that exist different (equivalent) models that fit the observations with the same error bounds makes uncertainty assessment a key step in any inversion process (Scales and Snieder, 2000; Fernández-Martínez et al., 2012a, 2013).

In this paper, we applied a full family of Particle Swarm Optimizers to the 3D gravity inversion and uncertainty assessment of basement relief in sedimentary basins. It can be understood as the second part of Pallero et al. (2015), where the 2D problem with PSO was addressed with satisfactory results. In addition to the 3D extension, some aspects as density contrast variation with depth and laterally are developed in a more detailed way.

The organization of the paper is as follows: section 2 explains the basin modelling. The criteria for basin discretization,

*Corresponding author

Email addresses: jlg.pallero@upm.es (J.L.G. Pallero), jlfm@uniovi.es (J.L. Fernández-Martínez), sylvain.bonvalot@ird.fr (S. Bonvalot), fudym@mines-albi.fr (O. Fudym)

density selection, regional trend modelling and cost function design are detailed. In section 3, the PSO design for the model inversion is introduced with regard to the searching bounds design and the PSO family members. Sections 4 and 5 are devoted to work with synthetic and real examples, respectively. The synthetic examples served to analyze the behavior of the different PSO family members and also helped to understand the importance of the different inversion parameters. As real example we present the inversion of the Argelès-Gazost basin (France). Therefore, the algorithm behavior in real environments is tested, showing that PSO is a serious alternative for gravity inversion and uncertainty assessment in 3D basement relief inversion. Finally, the general conclusions are summarized in the last section.

2. 3D basement relief modelling

Following [Barbosa et al. \(1997\)](#), we have modelled the basin as an accretion of M right rectangular prisms, considering the density contrast $\Delta\rho$ between the sediments and the basement a known parameter (fixed or variable with depth and/or horizontally). The horizontal dimensions of the prisms can be calculated as a factor (generally between $0.5\times$ and $1.5\times$) of the average spacing between the observations, and do not need to be equal along the X and Y axis. The top sides of the prisms are situated at the terrain surface level, being the height of the prisms the main problem unknowns. The set of prisms generates an attraction over each of the observed points that can be computed as the sum of the attraction of each individual prism. The formulation is provided in [Appendix A](#).

A key step in the problem setup is the decision concerning the density contrast between the sediments and the basement, $\Delta\rho = \rho_s - \rho_b$, to be used. The literature shows examples of using a constant density value ([Krimmel, 1970](#); [Stern, 1978](#); [Venteris and Miller, 1993](#); [Barbosa et al., 1997](#); [Gabalda et al., 2005](#)), although a variable density dependency with depth due to compaction is a common strategy, especially for deep environments. Some models for density variation with depth are based on binomial ([Chakravarthi, 1995](#)), quadratic ([Rao, 1990](#)), hyperbolic ([Litinsky, 1989](#); [Rao et al., 1995](#); [Silva et al., 2009](#)), parabolic ([Chakravarthi et al., 2002](#); [Işik and Şenel, 2009](#)) or exponential ([Cordell, 1973](#)) functions.

In this paper, we propose a numerical approximation to the problem of dealing with the density variation with depth. The original prism is divided into sub-prisms with their corresponding density value interpolated according to the selected density distribution. Then, Eqns. (A.2) or (A.3) are applied to each sub-prism and the final attraction is computed as the sum of each individual attractions. This strategy can be used with analytical density distributions and also with known density values from boreholes or other estimation methods. The size of the sub-prisms must be chosen such as the error in the approximation is less than a given tolerance.

In real environments, the density varies not only with depth but also laterally at global and local scales, due to different heterogeneities in the Earth crust. The common practice to estimate this lateral density variations is using linear inversion

techniques setting the depths, layer thicknesses, dividing the domain into prisms, and then working with densities as unknowns ([Tiberi et al., 2005](#); [Li et al., 2012](#)). This procedure requires a good prior knowledge of the subsurface geometry, which is not always at disposal.

In this work, we allow lateral density variations to be used at inversion. If the density (fixed or variable with depth) is known at various locations throughout the basin, for example at both sides of a fault, the domain (in top view) can be divided into polygons and the corresponding density contrast can be used for the gravity computations in the corresponding prisms.

Finally, a regional trend affecting the data might also be considered. The gravity anomalies Δg contain information not only due to the sedimentary basin mass anomaly under study, but also due to mass distribution in large areas of the depth Earth crust and upper mantle, which affect at regional scale. These regional contribution mask the local anomaly, and need to be removed prior to the inversion or considered as an additional unknown in the process. In this paper, we consider a plane surface as a regional trend whose coefficients will be treated as unknowns. Also a second order polynomial [Chakravarthi and Sundararajan \(2007\)](#) or even higher degree surfaces could be used.

The gravity anomaly derived at each of the $i = 1, \dots, N$ observed points is computed as the sum of the attraction of each of the M individual prisms plus the regional trend contribution:

$$\Delta g_i^{com} = \sum_{j=1}^M \left[F(\Delta\rho(z)_j, p_{ij}) + A + B \cdot x_i + C \cdot y_i \right], \quad (1)$$

where F is the forward problem defined in Eqns. (A.2) or (A.3), p_{ij} is the prism position relative to the observation point, $\Delta\rho(z)_j$ is the density contrast (fixed or depth dependent) attributed to each prism, x_i and y_i are the coordinates of the observation point, and A , B and C are the trend coefficients.

The nonlinear inverse problem to estimate the bottom side coordinates \mathbf{z}_b of the model prisms may be formulated as the minimization of the cost function

$$c^g(\mathbf{z}_b) = \|\Delta \mathbf{g}^{obs} - \Delta \mathbf{g}^{com}\|_2^2, \quad (2)$$

where $\Delta \mathbf{g}^{obs}$ is the vector of observed gravity anomalies, and $\Delta \mathbf{g}^{com}$ the model predicted anomalies.

The problem stated in Eq. (2) is ill-posed because there exist many equivalent solutions that fit the observed data within the same error bounds. The classical way to deal with this kind of problem is the use of the well-known iterative local optimization techniques, such as Levenberg-Marquardt ([Marquardt, 1963](#)), Gauss-Newton or Quasi-Newton methods ([Nocedal and Wright, 1999](#)), which need an initial approximate model to solve the problem. Also, in order to stabilize the inversions these methods require different types of constraints and regularization parameters. Nevertheless, the results provided depend strongly on the used initial model and on the prior information that is adopted. Besides, these methods do not provide the possibility of performing a nonlinear uncertainty analysis of the solution that is adopted ([Fernández-Martínez et al., 2014b](#)).

Instead of local optimization algorithms we have used a full family of global PSO optimizers that do not look for the inverse operator, but perform a smart search of the model space, evaluating the forward problem for different models in order to conduct the swarm towards the low misfit regions of the cost function landscape. The algorithm has to be fast in locating the low misfit region (or regions) to avoid the solution of many forward problems and the need of very high computational requirements. Regularization is not needed in this case, but prior information is still necessary in order to reduce the set of equivalent solutions and guiding correctly the search. The inversion of the real dataset shown in this paper have been done with a personal computer and took less than 3 minutes of CPU time. The program could be further parallelized.

3. The Particle Swarm Optimization inversion scheme

Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995) is a global optimization algorithm inspired in the behavior of bird flocks and fish schools searching for food. In this search algorithm a swarm of models explores the space of possible solutions in order to optimize a given cost function, that in our case is the data prediction error shown in Eq. (2).

Although PSO has been used in applied geophysics (Fernández-Álvarez et al., 2006; Shaw and Srivastava, 2007; Fernández-Martínez et al., 2010a,b), in gravity inversion has been barely used. Some approximations have been proposed in Sweilam et al. (2008); Sanyi et al. (2009); Touthmalani (2013b,a), which are focused in the comparison of the PSO with other global optimizers and classical local optimizations techniques, but do not deal the capability to perform an approximate nonlinear uncertainty analysis (sampling of the nonlinear uncertainty regions) of the solution that has been adopted. Also, a clear description about the importance of a correct tuning of the PSO parameters (Fernández-Martínez et al., 2010a,b, 2012b) is usually absent. In Pallero et al. (2015), we applied PSO to the 2D determination of basement relief and model appraisal. The effect of noise in inversion was also analyzed, illustrating the theoretical results shown in Fernández-Martínez et al. (2014a,b).

The key task before PSO sampling consists in computing low and upper bounds for the parameter search space. These parameters are, in our case: (i) the regional trend coefficients, and (ii) the prisms bottom side heights. For the 3D basement relief gravity inverse problem, the search bounds are estimated as follows:

1. A first approximation to the regional trend is estimated via least squares fitting of a plane to the observed anomaly. Search bounds for the plane parameters A , B and C are also needed, and can be computed as follows: (i) selected a small gravity value, at the level of the observation error for example, it is sequentially added to each observed anomaly and also sequentially a new plane is adjusted via least squares; (ii) from the obtained collection of planes, the maximum and minimum values for A , B and C parameters are extracted, conforming the $[A_{min}, A_{max}]$, $[B_{min}, B_{max}]$, and $[C_{min}, C_{max}]$ search bounds for them.

2. The next task consist in computing a first approximation to the prisms bottom side height. After subtracting the computed regional trend from the observations, the remaining signal corresponds to the basin under study (plus the always present observational noise). Using these data, an approximate depth model \mathbf{d}^0 is computed using the residual anomaly and applying to each individual prism a formulation based on the standard Bouguer's plate correction formula, assuming a constant density contrast $\Delta\rho_0$ (Barbosa et al., 1997) as follows:

$$d_j^0 = \frac{|\Delta g_j^{res}|}{2\pi G \Delta\rho_0}, \quad (3)$$

where d_j^0 is the initial depth for the j -th prism (from 1 to M), Δg_j^{res} is the residual gravity anomaly at the top center of the prism, G is the Newton's constant, and $\Delta\rho_0$ is the prescribed constant density contrast (taking into account the possible lateral variations). Then, factors $k_{min} < 1.0$ and $k_{max} > 1.0$ are applied to the approximate model \mathbf{d}^0 in order to obtain its bounds as $\mathbf{d}_{min} = k_{min} \cdot \mathbf{d}^0$, and $\mathbf{d}_{max} = k_{max} \cdot \mathbf{d}^0$. Combined with the prisms top side heights \mathbf{z}_t , the basement relief absolute height search bounds are $\mathbf{z}_b^{min} = \mathbf{z}_t - \mathbf{d}_{max}$, and $\mathbf{z}_b^{max} = \mathbf{z}_t - \mathbf{d}_{min}$.

3. Finally, the complete model search bounds are created as the combination of both, the bottom relief heights and the regional trend parameters, as

$$\mathbf{m} \in \left\{ \left[\mathbf{z}_b^{min}, A_{min}, B_{min}, C_{min} \right], \left[\mathbf{z}_b^{max}, A_{max}, B_{max}, C_{max} \right] \right\}. \quad (4)$$

Prior information about known depths at particular locations can be introduced in the search bounds. The search bounds for the prisms containing these locations are modified according the prescribed density values. This kind of search bounds modification is equivalent to the *absolute constraints* employed by Barbosa et al. (1997).

Barbosa et al. (1997); Silva et al. (2006, 2009) imposed smoothness conditions for the basement relief through a regularization term in the cost function. The purpose of this term, called *relative constraints* in the literature, is stabilizing the inversion and to produce smooth and geologically plausible models. The way of imposing relative constraints in PSO is different from local optimization techniques. This kind of constraint is needed in the solution of the basement relief inverse problem via PSO, since the sampled models in absence of additional constraints show a typical sawtooth shape that is not realistic from a geological point of view. In this case the smoothness can be achieved via a moving average filter (Pallero et al., 2015).

The coefficients for a filter window of determined size are imposed, and each model is smoothed (only the prisms bottom side heights) before computing its misfit via Eq. (2). The optimal window filter size can be also considered as an additional inversion parameter.

3.1. PSO parameter tuning

Parameter tuning (inertia, local and global accelerations) is crucial important in PSO to achieve good results. Most of

the publications use a unique parameter set for all the particles, that was deduced from numerical analysis (Carlisle and Dozier, 2001; Clerc and Kennedy, 2002). Nevertheless it has been shown (Fernández-Martínez and García-Gonzalo, 2008, 2009) that these points are not unique, and their performance is due to the fact that are located close to the upper border of the PSO second order stability limit, where the attraction from the oscillation center is weak, and the exploratory behavior of PSO is very high because the variance of the trajectories, considered as stochastic processes, is almost unbounded.

The algorithms used in this paper, as we did in Pallero et al. (2015), belong to the category of free-parameter tuning, since for all the PSO optimizers use their corresponding cloud version, where each particle of the swarm have its corresponding values of inertia, local and global accelerations, that are selected in the neighborhood of the upper border of the corresponding second order stability regions (Fernández-Martínez and García-Gonzalo, 2009, 2012). Besides, each particle (geophysical model) has its corresponding time step Δt , in this paper varying randomly at each iteration between $\Delta t_{min} = 0.3$ and $\Delta t_{max} = 1.5$. The Δt parameter greatly influences the exploratory behavior of each PSO member: the exploration of the search space increases when Δt is greater than 1.0. Conversely the algorithm becomes more exploitative when Δt is lower than 1.0.

4. Numerical experiments with synthetic examples

In this section, the methodology will be tested using synthetic models for the noise-free and noisy cases. The synthetic basin relief model is based on a quadratic function and is composed by 400 right rectangular prisms of 250×250 m in X and Y directions, which top sides are all located at 1000 m height. The maximum depth is located at the center of the basin and reaches 500 m, while the shallower prisms are located at the NW, NE, SW and SW corners, reaching a depth of 20 m (Fig. 1 shows the depth map). The density contrast between the sediments and the basement was selected to be variable with depth, and it is composed by two layers of constant density. Also, lateral density variations are considered for the half North and South sectors. The contrast density model can be seen in Table 1.

4.1. Noise-free case

The Bouguer's anomaly was generated, using Eq. (3), at 229 observation points, located at the same height of the prisms upper sides (1000 m). 225 of them are regularly spaced forming a 321×321 m grid, and the last 4 are located at the corners, outside the model. A plane regional trend of parameters $A = -114.00$ mGal, $B = -1.263158 \cdot 10^{-3}$ mGal/m, and $C = 5.894737 \cdot 10^{-3}$ mGal/m was also added to the generated observations. Fig. 2 (upper) shows the original anomaly and the observation points. Then, the regional trend search bounds are estimated using all the observed anomalies as it was explained previously in Sect. 3, obtaining as search limits the following intervals:

- $A \in [-114.55, -113.95]$ mGal,

- $B \in [-1.29, -1.24] \cdot 10^{-3}$ mGal/m, and
- $C \in [5.86, 6.13] \cdot 10^{-3}$ mGal/m.

The initial residual gravity corresponding to the basin mass anomaly is computed by subtracting the adjusted regional trend from the observed anomaly and can be shown in Fig. 2 (middle). Although the true model has horizontal symmetry, it can be observed that the minimum is not located at the basin center, due to the lateral density variations. Finally, the initial depth model computed using the residual anomaly and Eq. (3) to design the search bounds is shown in Fig. 2 (bottom). The corresponding search space was generated following Sect. 3 with $k_{min} = 0.5$ and $k_{max} = 2.0$.

As we did in Pallero et al. (2015) different members of the PSO family (GPSO, CC-PSO, CP-PSO, PP-PSO and RR-PSO) have been analyzed. In this case, only relative constraints were imposed, using a filter window of size 5×5 points, and values

$$\begin{bmatrix} 0.7669 & 0.8471 & 0.8757 & 0.8471 & 0.7669 \\ 0.8471 & 0.9358 & 0.9673 & 0.9358 & 0.8471 \\ 0.8757 & 0.9673 & 1.0000 & 0.9673 & 0.8757 \\ 0.8471 & 0.9358 & 0.9673 & 0.9358 & 0.8471 \\ 0.7669 & 0.8471 & 0.8757 & 0.8471 & 0.7669 \end{bmatrix}, \quad (5)$$

which was created based on a bi-dimensional normal distribution with $\sigma = 2$. The swarm size was fixed to 35 particles (models), and the number of iterations was 150. Therefore, 5250 forward evaluations were performed in each experiment.

Different plots were produced:

1. Fig. 3 (top) shows the relative error for the best model through iterations. This convergence curves shows a high variability in the CP-PSO member, which implies a very exploratory character. Also CC-PSO and GPSO are exploratory versions and get stabilized in the region of 2% relative error. RR-PSO and PP-PSO have a more exploitative character, showing a constant relative error decreasing behavior.
2. Fig. 3 (bottom) shows the dispersion curves, which provide a measure of the swarm collapse with the iterations given by the median distance in each iteration between the different particles of the swarm and their center of gravity (normalized by the median distance in the first iteration, 100% dispersion). A dispersion lower than 5% would mean that the swarm has collapsed into an unique particle. This situation must be avoided (using a correct parameter tuning) to perform a fear sampling of the nonlinear equivalent region. GPSO, CC-PSO and CP-PSO show clearly an exploratory character. RR-PSO and PP-PSO show dispersions below 5% after iterations 60 – 100, respectively. Moreover, the Δt parameter selection was set randomly between the limits $\Delta t_{min} = 0.3$ and $\Delta t_{max} = 1.5$. The Δt parameter serves to disperse the swarm and exploring the uncertainty space when the swarm is about to collapse. This fact can be observed for example in the PP-PSO case at iteration 80th, when the dispersion jumps up till 10%.
3. Fig. 4 shows the best model estimated by PP-PSO. Fig. 4 (top) shows the residual gravity anomaly produced by this

model, while Fig. 4 (bottom) represents the basin estimated depths. The relative error for this best model is 1.07%, and was computed as

$$c^r(\mathbf{z}_b) = \frac{\|\Delta\mathbf{g}^{obs} - \Delta\mathbf{g}^{com}\|_2}{\|\Delta\mathbf{g}^{obs}\|_2} \cdot 100, \quad (6)$$

where $\Delta\mathbf{g}^{obs}$ is the observed anomaly and $\Delta\mathbf{g}^{com}$ the computed one with the best depth model. The regional trend best coefficients for the PP-PSO solution were $A = -114.01$ mGal, $B = -1.277962 \cdot 10^{-3}$ mGal/m, and $C = 5.998566 \cdot 10^{-3}$ mGal/m. Fig. 5 shows two profiles from the PP-PSO best model plus the 10% equivalence region, the true model and the residuals in the observation points. Fig. 5 (top) is a West-East profile crossing the deepest basin point (index $i = 10$, according to Fig. 1), while Fig. 5 (bottom) is the North-South profile of index $j = 10$. It can be seen that the best estimated model matches the true one. In the North-South profile (bottom) it can be seen a systematically slightly over-determination in the northern depths and under-determination in the south. This can be interpreted as an effect of the joint regional trend determination and the lateral density variation: the estimated parameter C , the Y axis plane slope shows the greatest difference compared with its corresponding true value. The gravitational effect of the lateral density variation can be partially absorbed as trend signal, so this fact must be taken into account in the posterior interpretation of the results. In order to mitigate this effect in real environments, we propose to work with observed points not only contained in the basin, but also to consider observations in the surroundings, located farthest from the working area. These points are not so affected as the interior ones by the lateral density variations, and can be used for the initial estimation of the regional trend parameters and the searching bounds. Fig. 6 (top) shows the North-South profile ($j = 10$) corresponding to RR-PSO. It shows similar characteristics as the PP-PSO solution, with slightly lower relative error. Fig. 6 (bottom) shows the North-South profile $j = 10$ found by CC-PSO. Although the relative error is 1.92%, the bottom relief shape deviates from the true solution.

4. Finally, Fig. 12 (top) shows the interpolated cost function topography in the 2D-PCA space, in the region of equivalence (see details in [Pallero et al. \(2015\)](#)). This graph also serves to analyze numerically the effect of the noise in the inversion. It will be discussed later in the paper, when analyzing the noisy case.

4.2. Noisy case

In this example, the basin is the same as in the previous one that can be seen in Fig. 1. The observation points are also 229 (located also at height of 1000 m), but in this case they are not evenly distributed and the observations have been perturbed by a random white noise with normal distribution $N(0, 500)$ μ Gal. The relative misfit for the true model was 5.95%. The swarm size was 70 models, and the number of iterations 150. The

rest of the numerical parameters of the experiment remained unchanged with respect to the previous example.

Fig. 7 (upper) shows the original anomaly and the observation points. The regional trend searching bounds are in this case

- $A \in [-114.25, -113.89]$ mGal,
- $B \in [-1.29, -1.20] \cdot 10^{-3}$ mGal/m, and
- $C \in [5.80, 5.98] \cdot 10^{-3}$ mGal/m.

Fig. 7 (middle) shows the initial residual gravity corresponding to the basin mass anomaly minus the adjusted regional trend. Fig. 7 (bottom) shows the initial depth model, from which the search bounds for depths are computed using $k_{min} = 0.5$ and $k_{max} = 2.0$.

The main conclusions of this analysis are the following:

1. The convergence curves are shown in Fig. 8 (top). In this case, CC-PSO, PP-PSO and RR-PSO are the most exploitative algorithms. The algorithm that reaches the minimum relative error is RR-PSO, as in the noise-free example. CP-PSO is again the more exploratory algorithm.
2. Fig. 8 (bottom) shows the dispersion curves. As in the noise-free case, CC-PSO, GPSO and CP-PSO show a clear exploratory character. PP-PSO and RR-PSO dispersion is below 5% at the end of the exploration process, but have a good exploratory character until iterations 80 – 100.
3. Fig. 9 shows the best model estimated by PP-PSO (top), RR-PSO (middle), and CC-PSO (bottom). All the PSO versions estimate a maximum depth greater than the true depth, and also not in the true location. The CC-PSO solution has also a non-smooth character. PP-PSO provided the best solution. The estimated best regional trend for the PP-PSO solution are $A = -113.95$ mGal, $B = -1.279687 \cdot 10^{-3}$ mGal/m, and $C = 5.904982 \cdot 10^{-3}$ mGal/m. Fig. 10 shows the West-East profile number 8 (top), and the North-South one number 12 (bottom) provided by PP-PSO. It can be seen that the true model is contained in the 10% equivalent region. Fig. 11 (top) shows the North-South profile number 13 for the RR-PSO solution. In this case, the true model is almost contained in the 10% equivalent region, but there is more prisms with the true depths outside the 10% equivalent region than in the PP-PSO case. Fig. 11 (bottom) corresponds to the North-South profile number 13 of the CC-PSO solution. In this case, the solution is not as smoother as the PP-PSO and RR-PSO solutions.
4. Fig. 12 shows the cost function topography deduced from the PP-PSO samples. The top panel shows the noise-free case, and the bottom panel, the noisy case. As it was pointed in ([Fernández-Martínez et al., 2014a,b](#)) the noise decreases the size of the region of lower misfits and shifts the location of the maximum depth with respect to its true position. Similar plots were obtained for other PSO family members. It can be also observed that relative misfit for the best model is smaller than the relative misfit for the true solution, as it was pointed in ([Fernández-Martínez et al., 2012a, 2013, 2014a,b](#)).

5. Real example: The Argelès-Gazost basin

In this section, the proposed PSO method is applied to the inversion of available gravity data on the Argelès-Gazost basin, an ancient glacial valley located in the French Pyrenees mountain range, that is now occupied by the Gave de Pau river (Fig. 13). This area is characterized by a low density sedimentary filling of a pre-existing basement depression (older substratum), and represents a suitable case for the basement relief modeling of a sedimentary basin. It has been also subjected to recent geophysical studies, including 2D and 3D gravity data inversion. Therefore, it provides a good reference for testing the proposed 3D inversion methodology.

The Argelès-Gazost depression was formed and occupied by a glacier that belonged to a larger ice tongue overlying this Pyrenean region in the Quaternary times. Nowadays, this valley mainly surrounded by limestones and flyschs, and filled by Quaternary deposits (Alimen, 1964) and lying at an elevation of about 450 m is almost flat with width varying between 1 km and 2 km for a length of about 7 km (Fig. 13).

Within the last few years, two gravity surveys have been carried out in this valley with the aim to determine the 2D or 3D geometry of the basement relief (Perrouy, 2008; Moussirou, 2013). These new observations were acquired using relative Scintrex CG3/CG5 gravity meters (#9408268 and #9002136 from GET/IRD) and tied to absolute gravity stations recently determined in this area by IGN France and available from the BGI global Absolute gravity database (Wilmes et al., 2009; Wziontek et al., 2012). Stations coordinates were obtained from precise GPS positioning (differential measurements using dual-frequency GPS receivers) and processing (GAMIT/GLOBK software¹). The whole gravity data set, complemented with previous observations available from the Bureau Gravimétrique International (Drewes et al., 2012), has been reduced using CG3TOOL (Gabalda et al., 2003) and GEOSOFT Oasis Montaj software² in order to provide precise complete Bouguer anomalies, including terrain corrections computed from the SRTM 90 m resolution³. The final gravity data set used in this study is shown in Fig. 13 and contains 117 points. Among them 69 points are located inside the basin, while the remaining are located in the surroundings. Due to the uncertainties on both, gravity and coordinates data acquisition, and processing and according to Perrouy et al. (2015), the final accuracy of the Bouguer's gravity anomaly (computed for a density reference value of 2600 kg/m³) is estimated to be less than 0.5 mGal in zones of steep relief, and probably higher in zones of flat relief where the effect of topography is reduced.

The same dataset was also interpreted by Moussirou (2013). In order to ease the comparison of the results, the same reference value of $\Delta\rho = 600 \text{ kg/m}^3$ for density contrast between the sediments and the basement was used. No absolute constraints on the basement depths were set up. The average distance between the observed points being about 280 m. Therefore,

prisms of horizontal size equal to $225 \times 225 \text{ m}$ were selected for the basin discretization, resulting in a set of 300 prisms. The following PSO parameters were used here: (i) PP-PSO algorithm, (ii) swarm size of 200 models, (iii) 200 iterations, and (iv) plane approximation for modeling the regional trend. For the relative constraints, a filter window of size 5×5 points was used, whose coefficients were

$$\begin{bmatrix} 0.3459 & 0.4220 & 0.2652 & 0.0858 & 0.0143 \\ 0.4220 & 0.7669 & 0.7176 & 0.3459 & 0.0858 \\ 0.2652 & 0.7176 & 1.0000 & 0.7176 & 0.2652 \\ 0.0858 & 0.3459 & 0.7176 & 0.7669 & 0.4220 \\ 0.0143 & 0.0858 & 0.2652 & 0.4220 & 0.3459 \end{bmatrix}, \quad (7)$$

which was created based on a bi-dimensional normal distribution of $\sigma_1 = 1$ in NW-SE direction, and $\sigma_2 = 0.5$ along the NE-SW direction. The basin has its main dimension in the NW-SE direction, and the residual anomaly shows a similar trend. The NW-SE aligned prisms are constrained stronger than the NE-SW prisms, since this direction corresponds to the narrowest part of the basin.

Fig. 14 shows the residual anomaly (left) produced by the inverted best model, and the corresponding depths (right). As expected, the overall pattern of the gravity effect of our model is very similar to those previously found by other authors. The minimum anomaly is about -5.7 mGal and the regional trend parameters corresponding to this best model are $A = -63.61 \text{ mGal}$, $B = 3.607915 \cdot 10^{-4} \text{ mGal/m}$, and $C = 3.486091 \cdot 10^{-3} \text{ mGal/m}$, (A is referred to the point $P = (738158.216, 4765095.343) \text{ m}$). Considering the previous results, Perrouy (2008) and Moussirou (2013) deduced from their modeling a minimum gravity effect of -5.2 mGal and -5.6 mGal respectively (Perrouy (2008) considered a 3rd degree surface and Moussirou (2013) a plane as regional models in the computation of the residual gravity field produced by the basin itself). Moreover, the dataset used by Moussirou (2013) contained 24 additional observations with respect to the dataset used by Perrouy (2008). In this paper we have used the former dataset (Moussirou, 2013).

Fig. 14 shows the best depth model inferred from the 3D inversion. It can be observed a mean depth of about 150 m with a maximum value reaching 332.7 m. Fig. 15 shows different sections of this depth model, including the equivalence region of 15% relative error (the best model has a relative error of 9.72%). The West-East profile is shown in the top of this figure, and the North-South is shown in the bottom, and contains the deepest point in the basin at prisms indexes $i = 36$ and $j = 13$ (see Fig. 14). The analysis of the 15% equivalence region shows that the deepest point lies in the range of [288, 379] m depth from the surface. This uncertainty at 15% level of relative error deduced from the PSO inversion encompasses the depth value obtained for the same point (300m) by Moussirou (2013) using the 2D1/2 modeling (Cady, 1980). Perrouy (2008) obtained a shallower maximum depth of 250 m but as stated above, this model was constrained by a lower number of observations. It is interesting to note that the maximum depth value (234 m) estimated by Perrouy et al. (2015) using the same dataset and density contrast falls out of the 15% equivalence region. Even if

¹<http://www-gpsg.mit.edu/~simon/gtgk/index.htm>.

²<http://www.geosoft.com/products/oasis-montaj/>.

³<http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-dataset>.

a distinct 3D modeling technique was used (Virtual Prism magnetic and gravity inversion) following Fullagar et al. (2008), it is more likely that such difference is not related to the inversion process but to the fact that the 3rd degree surface used by Perrouy et al. (2015) to model the regional trend may have absorbed part of the Bouguer's signal. This is confirmed by the amplitude of the residual anomaly obtained by Perrouy et al. (2015) for the Argelès-Gazost valley, that reaches only -4.2 mGal, which leads to a significant underestimation of the thickness of the sedimentary basin.

Therefore we can conclude that the result achieved by PSO is consistent with previous results obtained for the Argelès-Gazost basin. Moreover, it served to quantify the uncertainty in the inverted depth model. The proposed 3D inversion methodology includes a regional trend modeling and the consideration of the topography of the gravity stations, as we previously did in our 2D approach (Pallero et al., 2015), and appears to be a suitable tool for fast and accurate 3D dimensional modeling of basement relief in sedimentary basins.

6. Conclusions

In this paper we have presented the design and application of a whole family of PSO optimizers to the 3D gravity inversion and model appraisal (uncertainty assessment) of basement relief estimation in sedimentary basins, that are discretized as aggregates of right rectangular prisms. Density variability with depth and laterally and regional trend estimation are taking into account. The algorithm performance is analyzed for two synthetic data sets (noise-free and white Gaussian noise cases) showing very acceptable results. The parameter tuning of all these algorithms is automatically performed based on stability considerations (cloud versions). Particularly, monitoring their respective dispersion curves is of paramount importance to ensure a correct sampling of the nonlinear uncertainty region. The effect of noise in inversion has also been analyzed via principal components analysis of the nonlinear equivalent models that have been sampled, illustrating the theoretical results that have been shown in Fernández-Martínez et al. (2014a,b). This analysis serves to urge for the need of performing a correct nonlinear uncertainty analysis in inversion.

Finally, a data set from the Argelès-Gazost basin (France) has been modeled using this methodology, to demonstrate that PSO is able to generate geologically plausible models that are consistent with those that have obtained by other well known classical approaches, but without uncertainty assessment. The PSO methodology that is proposed in this paper appears as a powerful tool for estimating the basement relief of sedimentary basins and their corresponding uncertainties. Moreover, as the inversion procedure takes into account the actual topography of the gravity observations and enables a separation of regional/residual anomalies, it is believed to be of special interest in many geological and geophysical applications requiring accurate estimation of basement relief of sedimentary basins. Further efforts will be made to make this software publically available.

7. Acknowledgements

JLGP acknowledges the support of the École des Mines d'Albi-Carmaux, that allowed him to develop this research in Albi during a month in 2015, and also the collaboration and help of the Laboratoire Géosciences Environnement Toulouse (GET) and Bureau Gravimétrique International (BGI).

We would also thank Joseph Martinod, Bernard Monod, Germain Gabalda, Moussirou Moussirou Bérangé, and Stéphane Perrouy for their contribution to the acquisition of the gravity data used in this study.

Appendix A. Gravity attraction of a right rectangular prism with constant density

Let a right rectangular prism of constant density ρ , defined by its extreme coordinates (x_m, y_m, z_m) and (x_M, y_M, z_M) , where the m subscript stands for *minimum*, and the subscript M for *Maximum*. Let $P(x_p, y_p, z_p)$ be an exterior point, as shown in Fig. 16. Let the boundary coordinates of the prism referred to the point P be

$$\begin{aligned} (x_1, y_1, z_1) &= (x_m, y_m, z_m) - (x_p, y_p, z_p), \\ (x_2, y_2, z_2) &= (x_M, y_M, z_M) - (x_p, y_p, z_p). \end{aligned} \quad (\text{A.1})$$

Then, the exact vertical gravitational attraction generated by the prism at P can be written (Nagy et al., 2000) as

$$F_N = G\rho \left\| \left\| x \ln(y+r) + y \ln(x+r) - z \arctan \frac{xy}{zr} \right\| \right\|_{x_1, y_1, z_1}^{x_2, y_2, z_2}, \quad (\text{A.2})$$

where G is the Newton's constant, and $r = \sqrt{x^2 + y^2 + z^2}$.

The evaluation of \ln and \arctan functions in Eq. A.2 is an intensive computational task since the exact gravity attraction computation involves a high number of prisms and observation points. When the distance from the point P to the prism is large, the gravitational attraction can be approximated (MacMillan, 1958; Fullea et al., 2008) as follows:

$$\begin{aligned} F_M &= G\rho \Delta x \Delta y \Delta z \left\{ \frac{1}{r^2} + \frac{5}{24r^6} [V_1 + V_2 \right. \\ &\quad \left. + V_3 \left(1 - \frac{2r^2}{5z^2} \right) \right\} \frac{z}{r} + o\left(\frac{1}{r^{11}}\right), \end{aligned} \quad (\text{A.3})$$

where

$$\begin{aligned} V_1 &= (2\Delta x^2 - \Delta y^2 - \Delta z^2) x^2, \\ V_2 &= (2\Delta y^2 - \Delta x^2 - \Delta z^2) y^2, \\ V_3 &= (2\Delta z^2 - \Delta x^2 - \Delta y^2) z^2, \\ r &= \sqrt{x^2 + y^2 + z^2}, \\ (\Delta x, \Delta y, \Delta z) &= (x_2 - x_1, y_2 - y_1, z_2 - z_1), \\ (x, y, z) &= \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}, \frac{z_1 + z_2}{2} \right), \end{aligned} \quad (\text{A.4})$$

and (x_1, y_1, z_1) and (x_2, y_2, z_2) are the coordinates defined in Eq. (A.1).

References

- Adema, G. W., Breckenridge, R. M., Sprengle, K. F., 2007. Gravity, Morphology, and Bedrock Depth of the Rathdrum Prairie, Idaho. Tech. rep., Idaho Geological Survey, University of Idaho.
- Alimen, M. H., 1964. Le Quaternaire des Pyrénées de la Bigorre. Mémoires pour servir à l'explication de la carte géologique de la France. Paris. Imprimerie nationale.
- Barbosa, V. C. F., Silva, J. B. C., Medeiros, W. E., 1997. Gravity inversion of basement relief using approximate equality constraints on depths. *Geophysics* 62 (6), 1745–1757.
- Bohidar, R. N., Sullivan, J. P., Hermance, J. F., 2001. Delineating Depth to Bedrock Beneath Shallow Unconfined Aquifers: A Gravity Transect Across the Palmer River Basin. *Groundwater* 39 (5), 729–736.
- Cady, J. W., 1980. Calculation of gravity and magnetic anomalies of finite length-right polygonal prisms. *Geophysics* 45 (10), 1507–1512.
- Carlisle, A., Dozier, G., 2001. An Off-The-Shelf PSO. In: Proceedings of the Particle Swarm Optimization Workshop.
- Chakravarthi, V., 1995. Gravity Interpretation of Nonoutcropping Sedimentary Basins in which the Density Contrast Decreases Parabolically with Depth. *Pure and Applied Geophysics* 145 (2), 327–335.
- Chakravarthi, V., Raghuram, H. M., Singh, S. B., 2002. 3-D forward gravity modeling of basement interfaces above which the density contrast varies continuously with depth. *Computers and Geosciences* 28 (1), 53–57.
- Chakravarthi, V., Singh, S. B., Babu, G. A., 2001. INVER2DBASE—A program to compute basement depths of density interfaces above which the density contrast varies with depth. *Computers and Geosciences* 27 (10), 1127–1133.
- Chakravarthi, V., Sundararajan, N., 2007. 3D gravity inversion of basement relief—A depth-dependent density approach. *Geophysics* 72 (2), 123–132.
- Clerc, M., Kennedy, J., 2002. The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space. *IEEE Transactions on Evolutionary Computation* 6 (1), 58–73.
- Cordell, L., 1973. Gravity analysis using an exponential density-depth function—San Jacinto Graben, California. *Geophysics* 38, 684–690.
- Dobrin, M. B., 1960. Introduction to geophysical prospecting. McGraw-Hill.
- Drewes, H., Hornik, H., Ádám, J., Rózsa, S., 2012. The Geodesist's Handbook. *Journal of Geodesy* 86 (10), 787–974.
- Fernández-Álvarez, J. P., Fernández-Martínez, J. L., García-Gonzalo, E., Menéndez-Pérez, C. O., 2006. Application of the particle swarm optimization algorithm to the solution and appraisal of the vertical electrical sounding inverse problem. In: 11th Annual Conference of the International Association of Mathematical Geology (IAMG '06), Liège, Belgium.
- Fernández-Martínez, J. L., Fernández-Muñiz, M. Z., Tompkins, M. J., 2012a. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* 77 (1), W1–W15.
- Fernández-Martínez, J. L., Fernández-Muñiz, Z., Pallero, J. L. G., Pedruelo-González, L. M., 2013. From Bayes to Tarantola: New insights to understand uncertainty in inverse problems. *Journal of Applied Geophysics* 98, 62–72.
- Fernández-Martínez, J. L., García-Gonzalo, E., 2008. The Generalized PSO: A New Door for PSO Evolution. *Journal of Artificial Evolution and Applications* 2008, 15 pages.
- Fernández-Martínez, J. L., García-Gonzalo, E., 2009. The PSO family: deduction, stochastic analysis and comparison. *Swarm Intelligence* 3, 245–273.
- Fernández-Martínez, J. L., García-Gonzalo, E., 2012. Stochastic Stability and Numerical Analysis of Two Novel Algorithms of PSO Family: PP-PSO and RR-PSO. *International Journal on Artificial Intelligence Tools* 21 (3), 20 pages.
- Fernández-Martínez, J. L., García-Gonzalo, E., Fernández Álvarez, J. P., Kuzma, H. A., Menéndez Pérez, C. O., 2010a. PSO: A powerful algorithm to solve geophysical inverse problems. Application to a 1D-DC resistivity case. *Journal of Applied Geophysics* 71, 13–25.
- Fernández-Martínez, J. L., García-Gonzalo, E., Naudet, V., 2010b. Particle swarm optimization applied to solving and appraising the streaming-potential inverse problem. *Geophysics* 75 (4), WA3–WA15.
- Fernández-Martínez, J. L., Mukerji, T., García-Gonzalo, E., Suman, A., 2012b. Reservoir characterization and inversion uncertainty via a family of particle swarm optimizers. *Geophysics* 77 (1), M1–M16.
- Fernández-Martínez, J. L., Pallero, J. L. G., Fernández-Muñiz, Z., Pedruelo-González, L. M., 2014a. The effect of noise and Tikhonov's regularization in inverse problems. Part I: The linear case. *Journal of Applied Geophysics* 108, 176–185.
- Fernández-Martínez, J. L., Pallero, J. L. G., Fernández-Muñiz, Z., Pedruelo-González, L. M., 2014b. The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case. *Journal of Applied Geophysics* 108, 186–193.
- Fullagar, P. K., Pears, G. A., McMonnies, B., 2008. Constrained inversion of geologic surfaces—pushing the boundaries. *The Leading Edge* 27 (1), 98–105.
- Fulla, J., Fernández, M., Zeyen, H., 2008. FA2BOUG—A FORTRAN 90 code to compute Bouguer gravity anomalies from gridded free-air anomalies: Application to the Atlantic-Mediterranean transition zone. *Computers and Geosciences* 34, 1665–1681.
- Gabalda, G., Bonvalot, S., Hipkin, R., 2003. CG3TOOL: An interactive computer program for Scintrex CG-3M gravity data processing. *Computers & Geosciences* 29 (2), 155–171.
- Gabalda, G., Nalpas, T., Bonvalot, S., 2005. The Base of the Atacama Gravels Formation (26°S, Northern Chile): First results from gravity data. In: 6th International Symposium on Andean Geodynamics (ISAG 2005, Barcelona). Extended Abstracts. Paris, IRD, pp. 286–289.
- Hinze, W. J., von Frese, R. R. B., Saad, A. H., 2013. Gravity and Magnetic Exploration. Principles, Practices and Applications, 1st Edition. Cambridge University Press.
- Işik, M., Şenel, H., 2009. 3D gravity modeling of Büyük Menderes basin in Western Anatolia using parabolic density function. *Journal of Asian Earth Sciences* 34, 317–325.
- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: IEEE International Conference on Neural Networks. Vol. 4, pp. 1942–1948.
- Krimmel, R. M., 1970. Gravimetric Ice Thickness Determination, South Cascade Glacier, Washington. *Northwest Science* 44 (3), 147–153.
- Li, J., Andrews-Hanna, J. C., Sun, Y., Phillips, R. J., Plaut, J. J., Zuber, M. T., 2012. Density variations within the south polar layered deposits of Mars. *Journal of Geophysical Research* 117 (E4), 1–13.
- Litinsky, V. A., 1989. Concept of effective density: Key to gravity determination for sedimentary basins. *Geophysics* 54, 1474–1482.
- MacMillan, W. D., 1958. The theory of the potential, 1st Edition. Dover Publications, New York.
- Marquardt, D. W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 11 (2), 431–441.
- Moussirou, B., 2013. Quantification du remplissage quaternaire des vallées glaciaires des Pyrénées par la méthode gravimétrique. Master's thesis, Université Toulouse III—Paul Sabatier.
- Nabighian, M. N., Ander, M. E., Grauch, V. J. S., Hansen, R. O., LaFehr, T. R., Li, Y., Peirce, J. W., Phillips, J. D., Ruder, M. E., 2005. Historical development of the gravity method in exploration. *Geophysics* 70 (6), 63ND–89ND.
- Nagy, D., Papp, G., Benedek, J., 2000. The gravitational potential and its derivatives for the prism. *Journal of Geodesy* 74, 552–560.
- Nettleton, L. L., 1976. Gravity and magnetics in oil prospecting. McGraw-Hill.
- Nocedal, J., Wright, S. J., 1999. Numerical Optimization, 1st Edition. Springer, New York.
- Pallero, J. L. G., Fernández-Martínez, J. L., Bonvalot, S., Fudym, O., 2015. Gravity inversion and uncertainty assessment of basement relief via particle swarm optimization. *Journal of Applied Geophysics* 116, 180–191.
- Parker, R. L., 1994. Geophysical Inverse Theory. Princeton University Press.
- Perrouy, S., 2008. Mesures géophysiques du remplissage sédimentaire Quaternaire dans les vallées de Bagnères de Bigorre et d'Argelès-Gazost (Hautes-Pyrénées). Master's thesis, Université Toulouse III—Paul Sabatier.
- Perrouy, S., Moussirou, B., Martinod, J., Bonvalot, S., Carretier, S., Gabalda, G., Monod, B., Héral, G., Regard, V., Remy, D., 2015. Geometry of two glacial valleys in the northern Pyrenees estimated using gravity data. *Comptes Rendus Geoscience* 347 (1), 13–23.
- Rao, C. V., Raju, M. L., Chakravarthi, V., 1995. Gravity modelling of an interface above which the density contrast decreases hyperbolically with depth. *Journal of Applied Geophysics* 34 (1), 63–67.
- Rao, D. B., 1990. Analysis of gravity anomalies of sedimentary basins by an asymmetrical trapezoidal model with quadratic density function. *Geophysics* 55 (2), 226–231.
- Sanyi, Y., Shangxu, W., Nan, T., 2009. Swarm intelligence optimization and its application in geophysical data inversion. *Applied Geophysics* 6 (2), 166–174.
- Scales, J. A., Snieder, R., 2000. The Anatomy of Inverse Problems. *Geophysics* 65 (6), 1708–1710.

- Shaw, R., Srivastava, S., 2007. Particle swarm optimization: A new tool to invert geophysical data. *Geophysics* 72, F75–F83.
- Silva, J. B. C., Costa, D. C. L., Barbosa, V. C. F., 2006. Gravity inversion of basement relief and estimation of density contrast variation with depth. *Geophysics* 71 (5), J51–J58.
- Silva, J. B. C., Oliveira, A. S., Barbosa, V. C. F., 2010. Gravity inversion of 2D basement relief using entropic regularization. *Geophysics* 75 (3), I29–I35.
- Silva, J. B. C., Teixeira, W. A., Barbosa, V. C. F., 2009. Gravity data as a tool for landfill study. *Environmental Geology* 57, 749–757.
- Stern, T. A., 1978. Gravity Survey of the Taylor Glacier, Victoria Land, Antarctica. Tech. Rep. 8, Geology Department, Victoria University of Wellington.
- Sweilam, N. H., Gobashy, M. M., Hashem, T., 2008. Using Particle Swarm Optimization with Function Stretching (SPSO) For Integrating Gravity Data: A Visibility Study. *Proceedings of the Mathematical and Physical Society of Egypt* 86 (2), 259–281.
- Telford, W. M., Geldart, L. P., Sheriff, R. E., Keys, D. A., 1976. *Applied Geophysics*. Cambridge University Press.
- Tiberi, C., Ebinger, C., Ballu, V., Stuart, G., Oluma, B., 2005. Inverse models of gravity data from the Red Sea–Aden–East African rifts triple junction zone. *Geophysical Journal International* 163, 775–787.
- Toushmalani, R., 2013a. Comparison result of inversion of gravity data of a fault by particle swarm optimization and Levenberg-Marquardt methods. *SpringerPlus* 2, 462.
- Toushmalani, R., 2013b. Gravity inversion of a fault by Particle Swarm Optimization (PSO). *SpringerPlus* 2, 315.
- Venteris, E., Miller, M., 1993. Gravitational profiles on the Taku glacier system. Open File Report, Glaciological and Arctic Sciences Institute, University of Idaho.
- Wilmes, H., Wziontek, H., Falk, R., Bonvalot, S., 2009. AGrav–The New International Absolute Gravity Database of BGI and BKG and its benefit for the Global Geodynamics Project (GGP). *Computers and Geosciences* 48, 304–309.
- Wziontek, H., Wilmes, H., Bonvalot, S., 2012. AGrav: An International Database for Absolute Gravity Measurements. In: Kenyon, S., Pacino, C. M., Marti, U. (Eds.), *Geodesy for Planet Earth: Proceedings of the 2009 IAG Symposium*. Buenos Aires, Argentina, 31 August–4 September 2009, pp. 1037–1042.

LIST OF CAPTIONS

Table 1: Density contrast model for the synthetic examples. Row indices start at 1 in the North border.

Figure 1: True basin depth model for synthetic examples.

Figure 2: Noise-free problem model. Top: observations (model gravity anomaly plus regional trend). Middle: initial model gravity anomaly (observations minus initial regional trend). Bottom: initial depth model based on Eq. (3). Gravity values have been interpolated at the prisms centers only for visualization purposes.

Figure 3: Convergence (top) and dispersion (bottom) curves for the noise-free case.

Figure 4: Gravity anomaly due to the best model (top) and depths (bottom) for the PP-PSO noise-free best solution.

Figure 5: Best model solution profiles for the noise-free case using the PP-PSO algorithm. Top: WE profile number 10 (according to the indices in Fig. 1). Bottom: NS profile number 10. The central prisms in both profiles coincide with the basin deepest area.

Figure 6: North-South best model solution profiles for the noise-free case using the RR-PSO algorithm (top), and CC-PSO (bottom).

Figure 7: Noisy $N(0, 500)$ μGal problem model. Top: observations (model gravity anomaly plus regional trend). Middle: initial model gravity anomaly (observations minus initial regional trend). Bottom: initial depth model based on Eq. (3).

Figure 8: Convergence (top) and dispersion (bottom) curves for the noisy $N(0, 500)$ μGal case.

Figure 9: Best model depths for the PP-PSO noisy $N(0, 500)$ μGal case (top), RR-PSO (middle), and CC-PSO (bottom).

Figure 10: Best model solution profiles for the noisy $N(0, 500)$ μGal case using the PP-PSO algorithm. Top: WE profile number 8. Bottom: NS profile number 12.

Figure 11: North-South best model solution profiles for the noisy $N(0, 500)$ μGal case using the RR-PSO algorithm (top), and CC-PSO (bottom). Both are NS profile number 13.

Figure 13: Localization of the Argelès-Gazost basin (France) and distribution of gravity data (black dots) used as a real case application for this study. Isolines denote the topographic variations in meters and names correspond to the main cities. The blue contour underlines the limit of the sedimentary filling of the basin considered in the gravity inversion.

Figure 14: Computed anomaly (left) and depth model (right) for the Argelès-Gazost sedimentary filling corresponding to the best model estimated from the gravity inversion using the PP-PSO algorithm (9.72% relative mistfit).

Figure 15: Best model solution profiles crossing the maximum basin depth for the Argelès-Gazost sedimentary basin model presented in Fig. 14. Top: WE profile along the prism index $i = 36$. Bottom: NS profile along the prism index $j = 13$. The 15% equivalence region estimated for the depth estimation is reported.

Figure 16: Three dimensional schematic situation of a right rectangular prism and an attracted exterior point $P(x_p, y_p, z_p)$.

ρ : prism density, (x_m, y_m, z_m) : minimum coordinates, and
 (x_M, y_M, z_M) : maximum coordinates of the prism sides.

LIST OF TABLES

Table 1:

Row index	Depth (m)	$\Delta\rho$ (kg/m ³)
$i = 1, \dots, 10$	0.0	-600.0
	199.999	-600.0
	200.001	-500.0
$i = 11, \dots, 20$	0.0	-650.0
	299.999	-650.0
	300.001	-550.0

LIST OF FIGURES

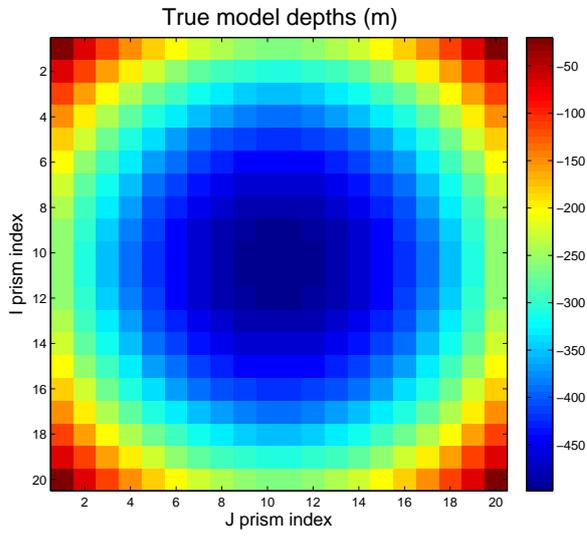


Figure 1:

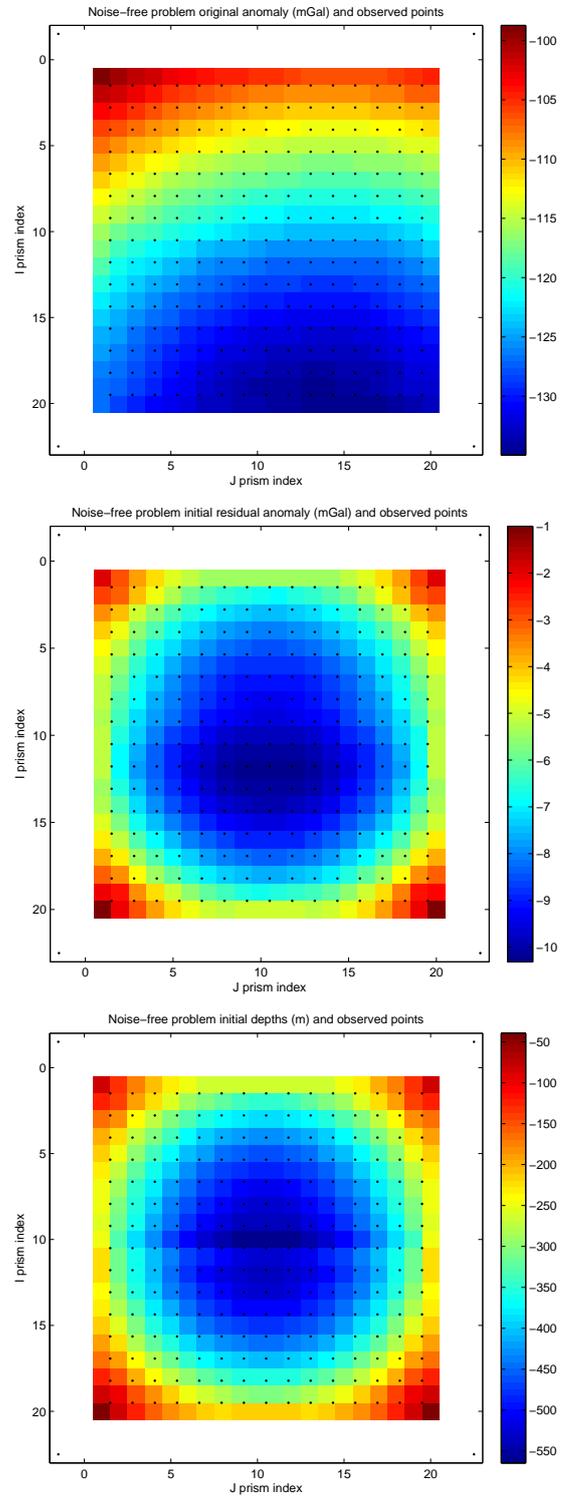


Figure 2:

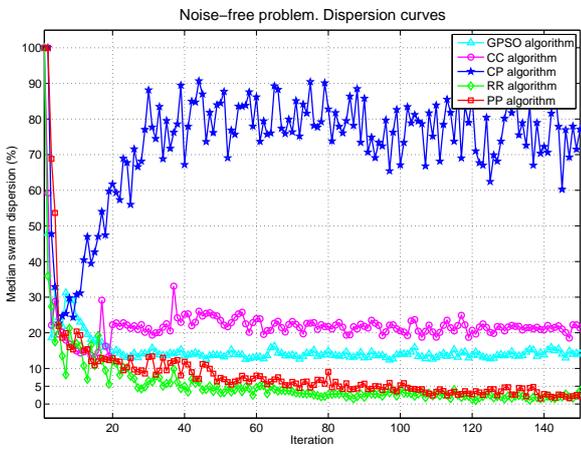
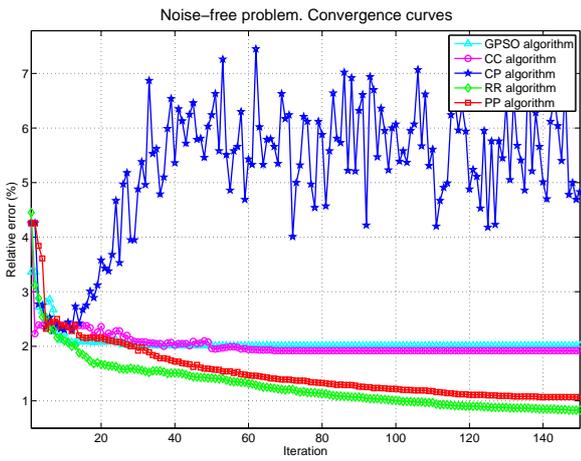


Figure 3:

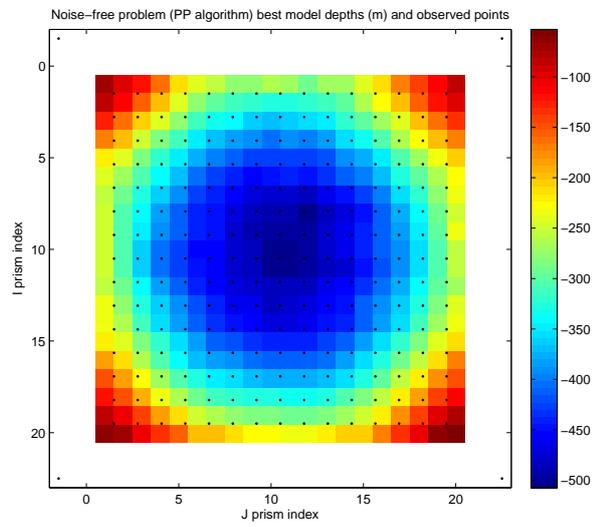
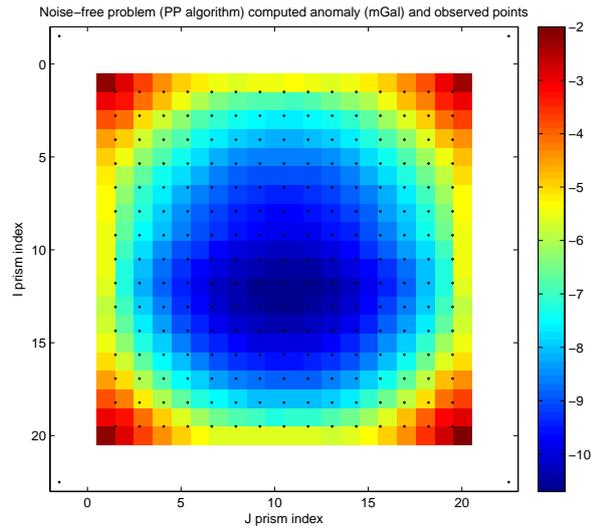


Figure 4:

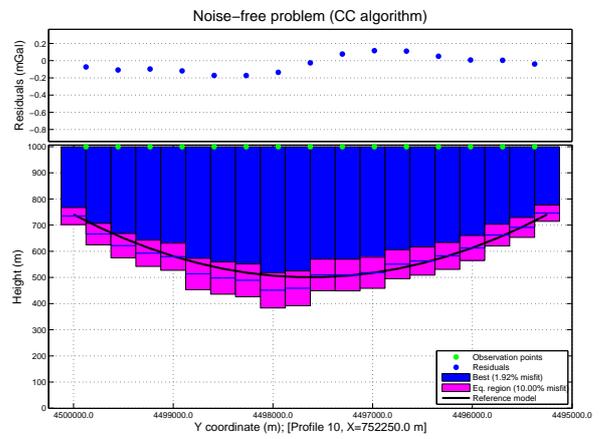
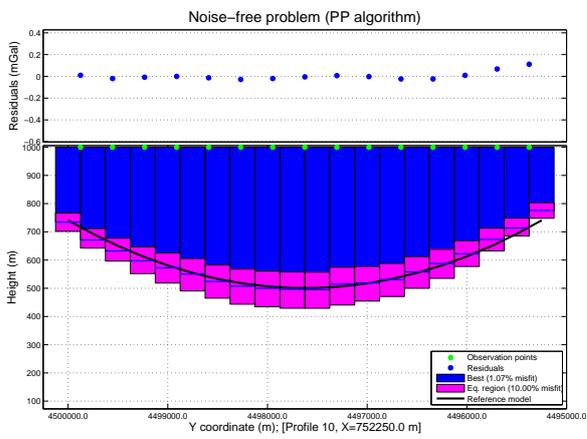
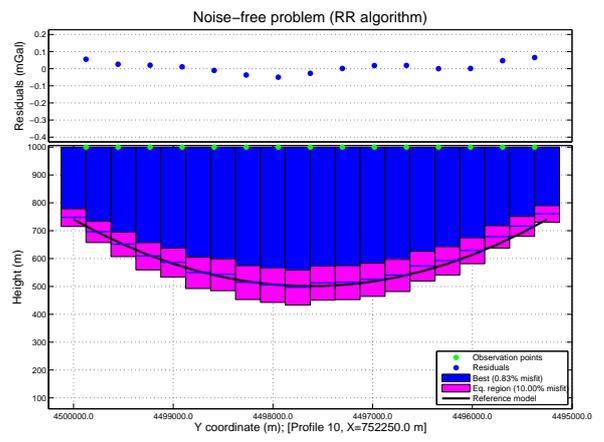
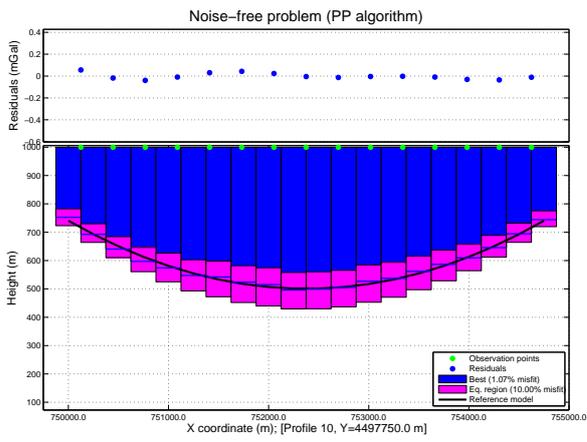


Figure 5:

Figure 6:

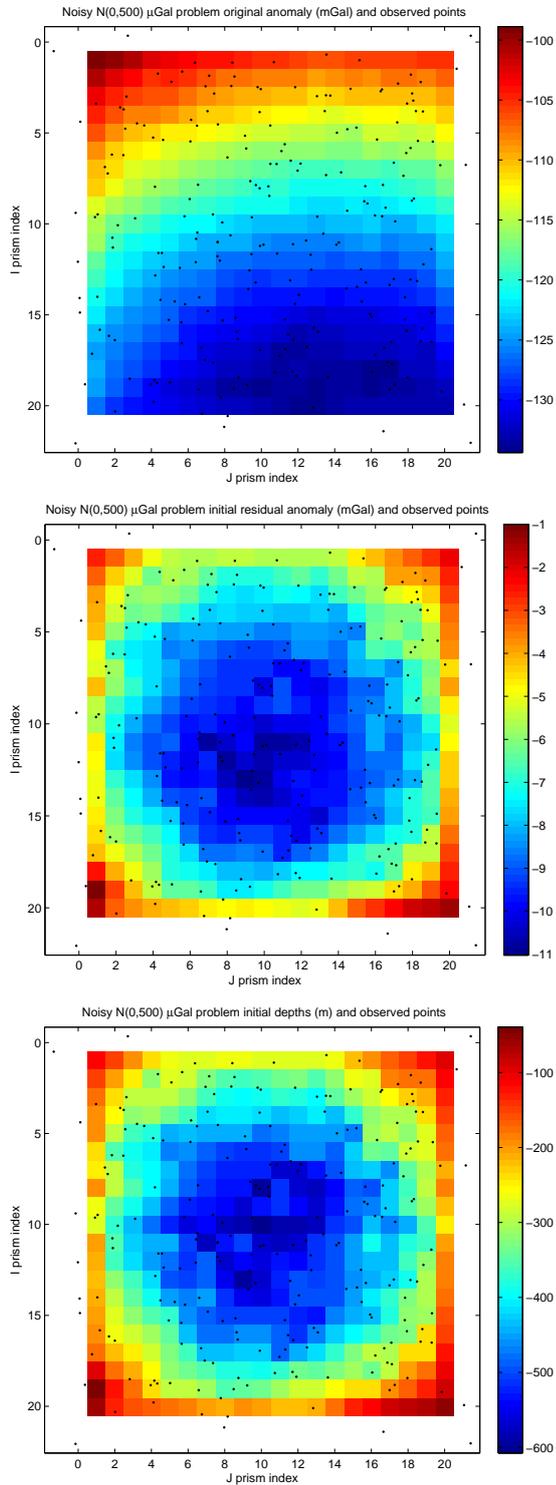


Figure 7:

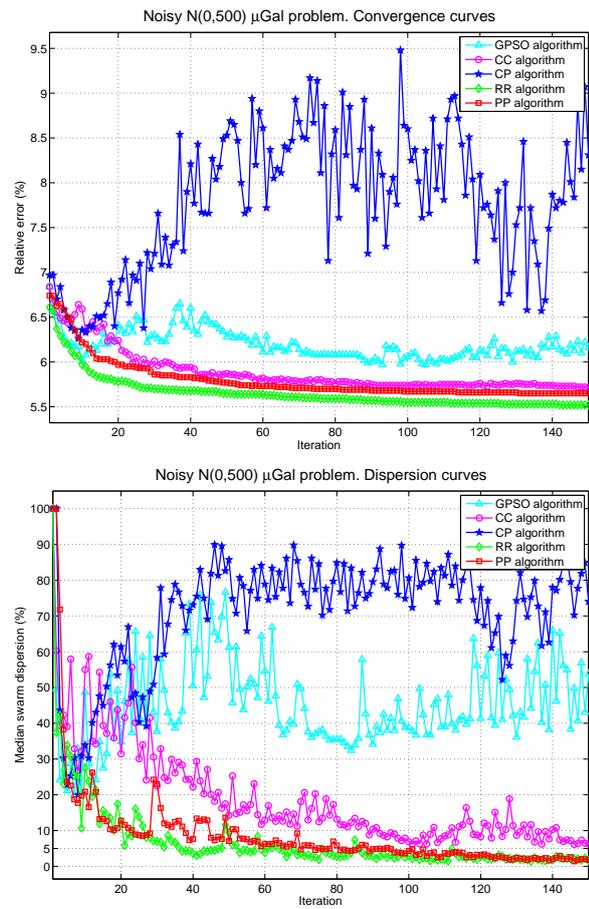


Figure 8:

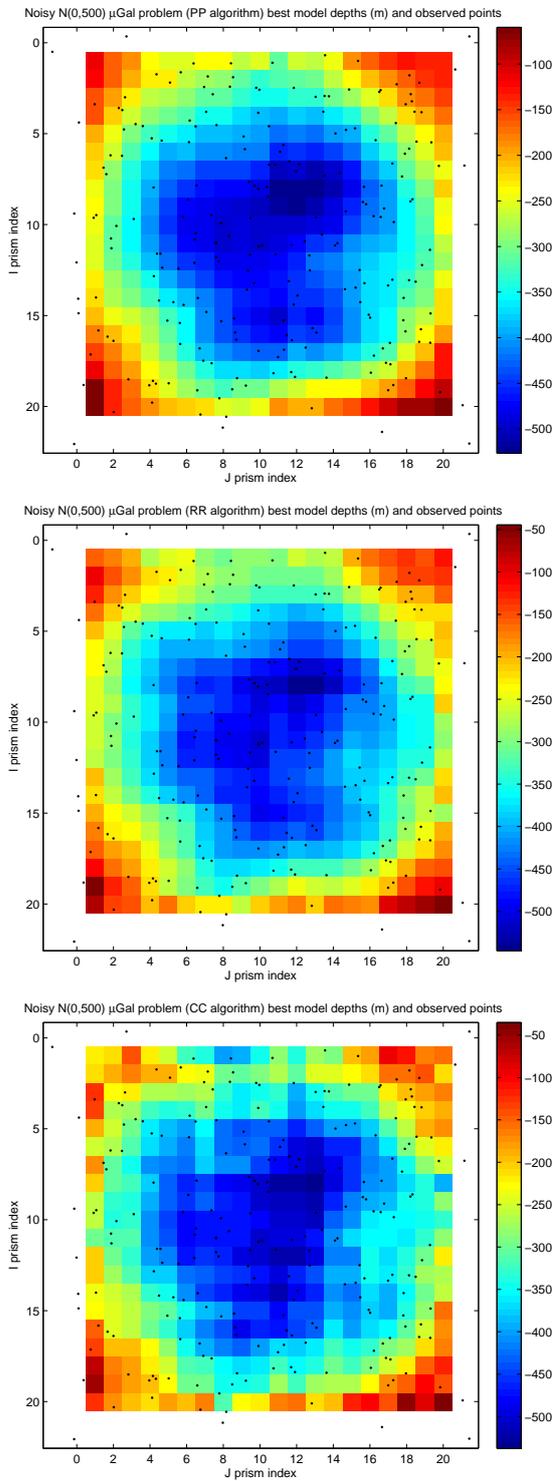


Figure 9:

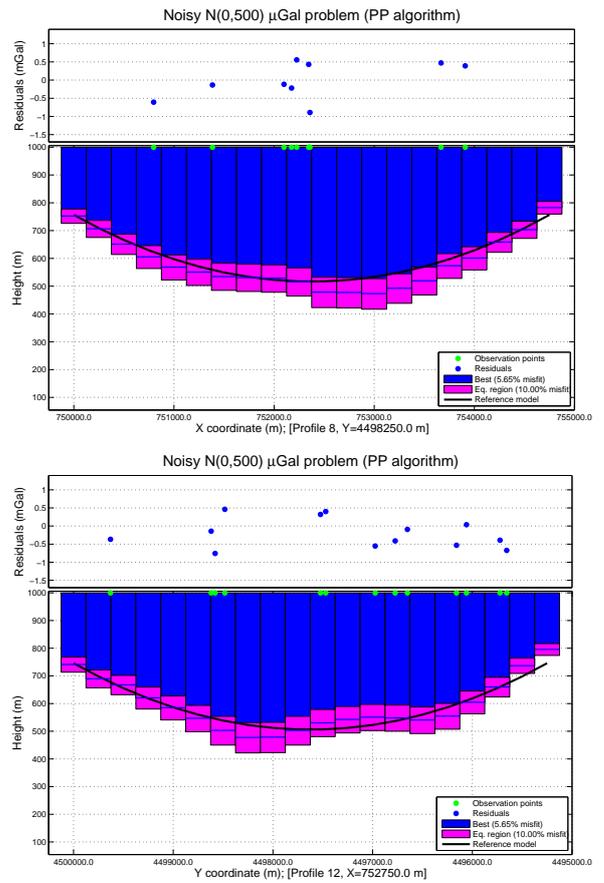


Figure 10:

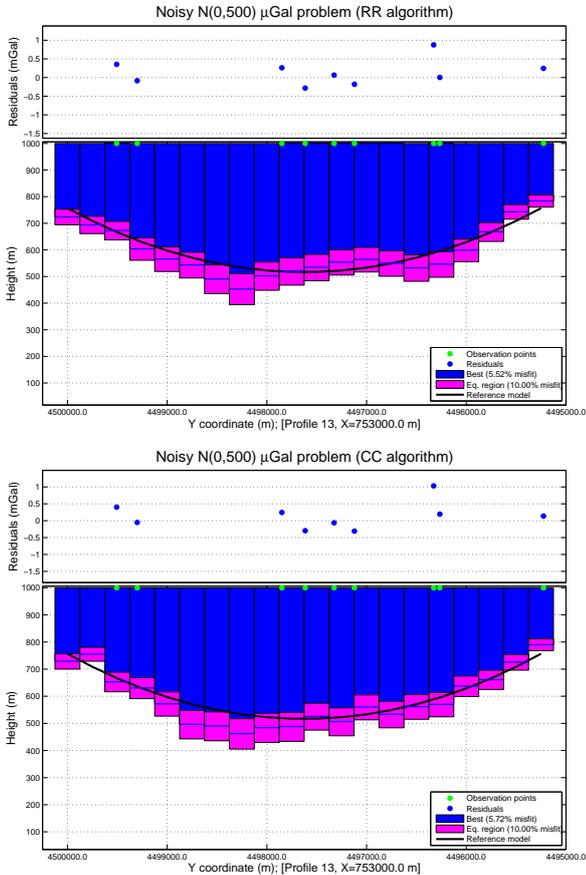


Figure 11:

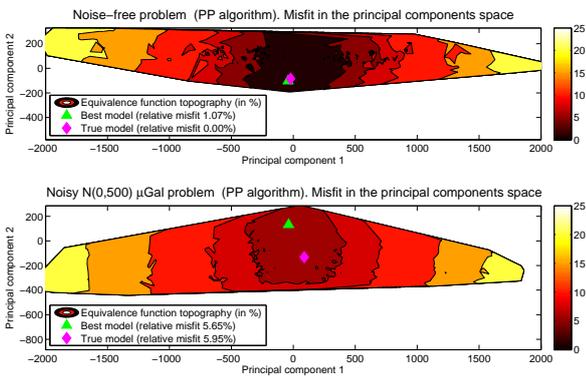


Figure 12:

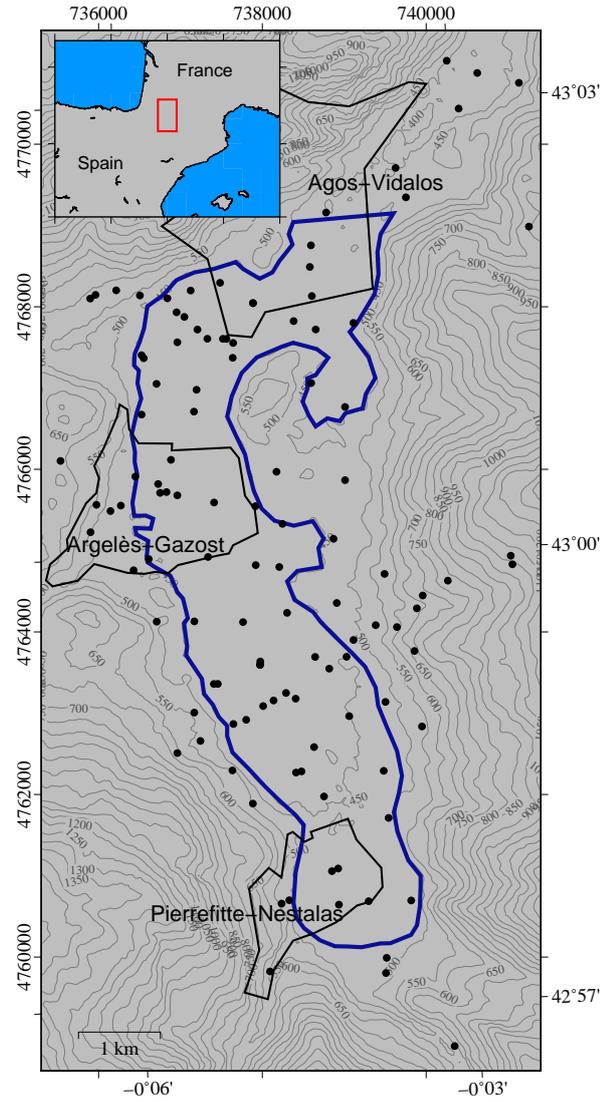


Figure 13:

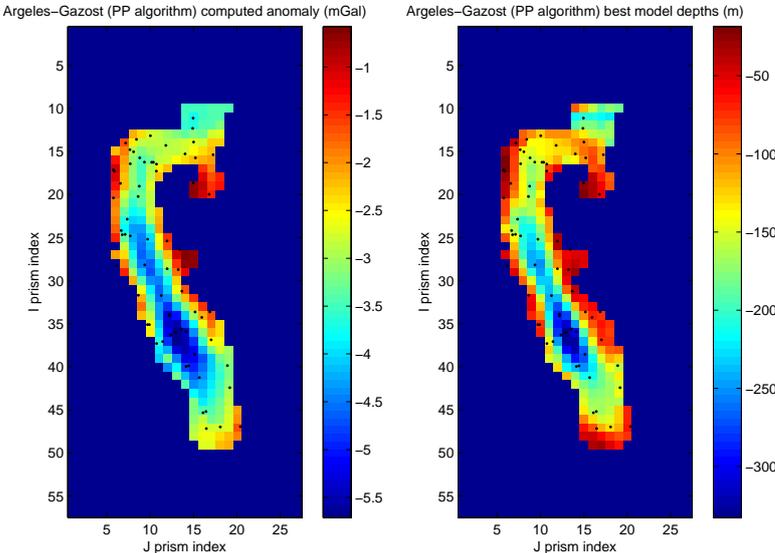


Figure 14:

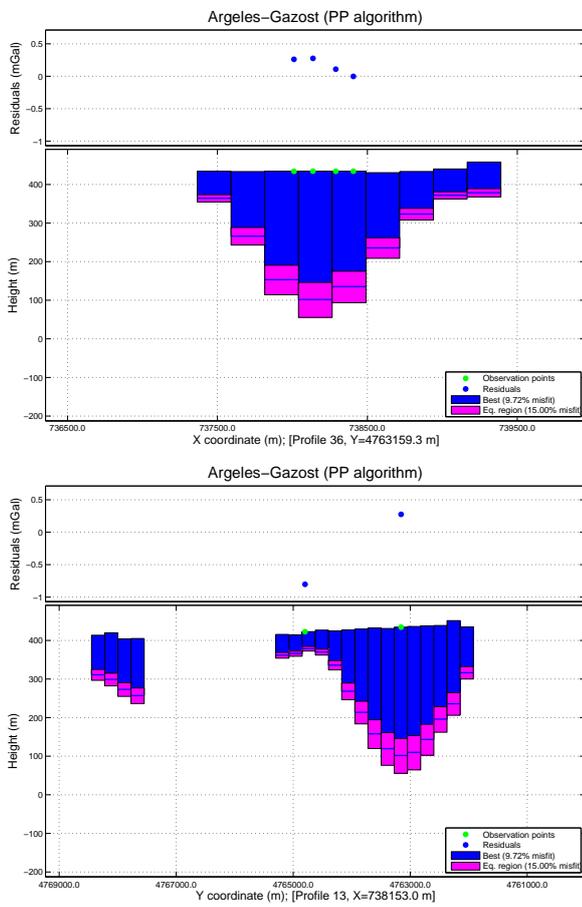


Figure 15:

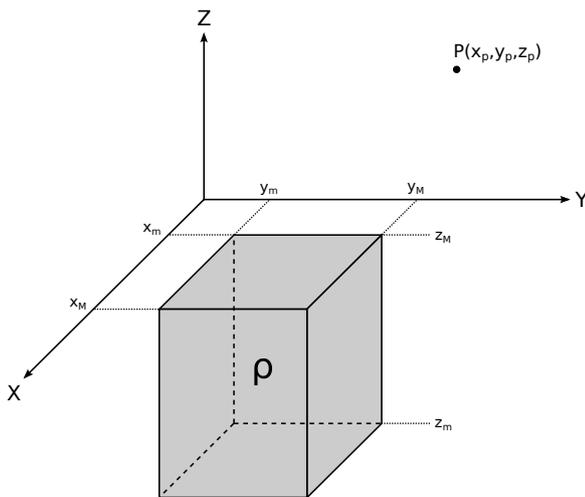


Figure 16:

Chapter 4

Other publications

Introduction

This chapter includes two international congress contributions, which complement the main core of the PhD Thesis. In both cases, the contribution was in the modality of *oral communication* plus *extended abstract*, being the documents published in the corresponding proceedings. These contributions were preliminary results of articles 2 and 3, in preparation at the time when the symposia took place.

Regarding the quality indicators, both congress were international and the contributions were peer-reviewed before their approval. One of these contributions (the work presented at the IAMG symposium) is included in the *Conference Proceedings Citation Index*.

PUBLICATION 1

**Noise, regularization and
uncertainty: new insights for linear
and nonlinear inverse problems**

Authors:

Juan Luis Fernández Martínez
José Luis García Pallero
María Zulima Fernández Muñiz
Luis Mariano Pedruelo González

4TH INVERSE PROBLEMS, DESIGN AND OPTIMIZATION
SYMPOSIUM (IPDO-2013)
ALBI (FRANCE), JUNE 26–28, 2013
isbn: 979-10-91526-01-2

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

NOISE, REGULARIZATION AND UNCERTAINTY: NEW INSIGHTS FOR LINEAR AND NONLINEAR INVERSE PROBLEMS

Juan Luis Fernández-Martínez^a, J. L. G. Pallero^b, M. Zulima Fernández-Muñiz^a, and Luis Mariano Pedruelo-González^a

^aDepartamento de Matemáticas, Área de Matemática Aplicada. Universidad de Oviedo, Oviedo, Spain.
jlfm@uniovi.es, zulima@uniovi.es, marianopg@educastur.princast.es

^bETSI en Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid, Madrid, Spain.
jlg.pallero@upm.es

Abstract

Inverse problems are a special kind of optimization problems since the cost function involves observed data that is always affected by noise. This paper analyses the effect of noise in data and the effect of the regularization in the cost function topography for linear and nonlinear inverse problems. In the case of ill-conditioned inverse problems the noise in data is amplified back to the model parameters providing spurious solutions if no regularization techniques are used. Furthermore, the noise shifts the least squares solution of the linear system and deforms homogeneously the region of equivalence which is limited by a hyper-quadric surface. The zero-order Tikhonov's regularization with a model of reference serves to limit the axes of hyper-quadric surface in the directions that span the kernel of the forward operator and to inform the model components of the solution that originally resided in the kernel. In the case of nonlinear inverse problems the effects are similar to the linear case, but the noise deforms the topography non-homogeneously. The effect of the regularization in the linearized inverse problem is similar to the linear case, although it does not provoke the disappearance of the nonlinear equivalent models. Finally, the linearized and nonlinear uncertainty analyses generally provide very different risk assessment conclusions.

Nomenclature

Roman symbols

$c(\mathbf{m})$	Cost function
\mathbf{d}^{obs}	Observed data
\mathbf{d}^{true}	Noise-free data
\mathbf{F}	Forward operator
$\mathbf{HC}_{\mathbf{m}_0}$	Hessian matrix
$\mathbf{JF}_{\mathbf{m}_0}$	Jacobian of \mathbf{F}
L_{tol}	Linear equivalence region

\mathbf{m}	Model parameters
\mathbf{m}^{ref}	Reference model
\mathbf{m}^{true}	True model
tol	Tolerance value
V_{tol}	Nonlinear equivalence region

Greek symbols

$\delta\mathbf{d}$	Noise in data
ε^2	Damping parameter
Σ	Singular value matrix from SVD

Introduction

Most inverse problems in science and engineering can be written in discrete form as $\mathbf{F}(\mathbf{m}) \approx \mathbf{d}^{obs}$, where $\mathbf{m} = (m_1, m_2, \dots, m_n) \in \mathbf{M} \subset \mathbb{R}^n$ are the model parameters, $\mathbf{d}^{obs} \in \mathbb{R}^s$ are the observed data, and $\mathbf{F}(\mathbf{m}) = [f_1(\mathbf{m}), f_s(\mathbf{m}), \dots, f_s(\mathbf{m})]$ represents the forward model, where $f_i(\mathbf{m})$ is the scalar field that serves to predict the i -th data. The inverse problem is linear if the forward operator is linear, that is, the forward operator \mathbf{F} depends linearly on the model parameters \mathbf{m} , in which case \mathbf{F} is represented by a rectangular matrix of s rows and n columns.

Uncertainty in the solution is an intrinsic feature in linear and nonlinear problems. In linear inverse problems the region of equivalence $L_{tol} = \{\mathbf{m} : c_p(\mathbf{m}) = \|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2 \leq tol\}$ is the set of models that are located within the hyper-quadric surface in \mathbb{R}^n ([1])

$$\mathbf{m}^T \mathbf{F}^T \mathbf{F} \mathbf{m} - 2\mathbf{m}^T \mathbf{F}^T \mathbf{d}^{obs} + \mathbf{d}^{obs^T} \mathbf{d}^{obs} - tol^2 = 0. \quad (1)$$

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

The fact that the matrix of the hyper-quadric $\mathbf{F}^T \mathbf{F}$ is semi-definite positive (with positive or null singular values) implies that the hyper-quadric containing the equivalent models is an hyper-ellipsoid or an hyper-elliptical cylinder. In the case of purely overdetermined linear inverse problems the hyper-quadric is a single hyper-ellipsoid centered at the least-squares solution of the linear system. The lengths of the semi-axes of the hyper-quadric of equivalence are $1/\mu_k$, which implies that the model parameters are more uncertain in the directions corresponding to the smaller singular values. For nonlinear inverse problems, it can be applied the same kind of reasoning as we have for the linear inverse problem to characterize local regions of equivalence around a certain model \mathbf{m}_0 belonging to the nonlinear region of equivalence ([1]). Let us define the region of nonlinear equivalence of value tol as $V_{tol} = \{\mathbf{m} : c_p(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2 \leq tol\}$, and let us suppose that we have at our disposal a model $\mathbf{m}_0 \in V_{tol}$. Considering a Taylor series expansion in \mathbf{m}_0 , we have

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_0) + \mathbf{J}\mathbf{F}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + \mathbf{o}(\|\mathbf{m} - \mathbf{m}_0\|_2), \quad (2)$$

where $\mathbf{J}\mathbf{F}_{\mathbf{m}_0} = \left(\frac{\partial \mathbf{F}}{\partial m_1}, \frac{\partial \mathbf{F}}{\partial m_2}, \dots, \frac{\partial \mathbf{F}}{\partial m_n} \right)_{\mathbf{m}_0}$ is the Jacobian matrix of $\mathbf{F}(\mathbf{m})$ at \mathbf{m}_0 , and $\mathbf{o}(\|\mathbf{m} - \mathbf{m}_0\|_2)$ is a vector function that vanishes faster than $\|\mathbf{m} - \mathbf{m}_0\|_2$.

The first-order approximation of the nonlinear equivalent region V_{tol} is the linearized hyper-quadric

$$\begin{aligned} L_{tol}(\mathbf{m}_0) &= \\ &= \left\{ \mathbf{m} : c_p(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) + 2\Delta \mathbf{d}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T (\mathbf{m} - \mathbf{m}_0) + \|\Delta \mathbf{d}\|_2^2 \leq tol^2 \right\}, \end{aligned} \quad (3)$$

where $\Delta \mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{obs}$.

$\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}$ represents the Gauss-Newton approximation to the Hessian term of the cost function since

$$\begin{aligned} \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T \mathbf{H}\mathbf{C}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) &= \\ &= (\mathbf{m} - \mathbf{m}_0)^T \left\{ \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} + \sum_{k=1}^s [f_k(\mathbf{m}_0) - d_k] Hf_k(\mathbf{m}_0) \right\} (\mathbf{m} - \mathbf{m}_0) \approx \\ &\approx (\mathbf{m} - \mathbf{m}_0)^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) \end{aligned} \quad (4)$$

The region of nonlinear equivalence V_{tol} has a valley shape, and cannot be straight since the nonlinearity of the forward operator \mathbf{F} causes the Jacobian to change continuously with the model \mathbf{m}_0 . In conclusion, the nonlinear region of equivalence has a meandering valley or croissant-shape. Further details can be consulted in [1]. The term $\sum_{k=1}^s [f_k(\mathbf{m}_0) - d_k] Hf_k(\mathbf{m}_0)$ might cause the exact Hessian to lose its positive semi-definite character, that is, the nonlinear equivalent region can be composed of several disconnected basins of equivalent solutions. It is worth mentioning that most of the local optimization methods approximate the curvature of the cost function by the linearized Hessian $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}$, which is semi-definite positive. Thus, the saddle points of the cost function cannot be found in this analysis, and the uncertainty is reduced to the linearized hyper-quadric. Otherwise said, the local optimization methods are not designed to look for different basins of solutions due to the numerical approximations used in their designed.

1. The effect of the noise in linear inverse problems

Noise is one of the main causes of uncertainty in inverse problems, together with incomplete data coverage and the approximated forward physics. The observed data \mathbf{d}^{obs} are always affected by noise $\delta \mathbf{d}$, that causes the linear system to have no solution. This incompatibility is also caused by the fact that the linear operator \mathbf{F} is only an approximation of the reality. Due to this fact the inverse problem is solved via linear least squares. The generalized solution of the least-squares problem is provided by the Moore-Penrose pseudo-inverse (see for instance [2]).

Let us call $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta \mathbf{d}$, where \mathbf{d}^{true} is the observed data that would be measured in absence of noise (noise-free data), and \mathbf{m}^{true} the hypothetical model that has generated these data, that is, $\mathbf{F}\mathbf{m}^{true} =$

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

\mathbf{d}^{true} . Then $\mathbf{m}^\dagger = \mathbf{F}^\dagger \mathbf{d}^{obs} = \mathbf{F}^\dagger (\mathbf{d}^{true} + \delta \mathbf{d}) = \mathbf{m}^{true} + \Delta \mathbf{m}$, with $\Delta \mathbf{m} = \mathbf{F}^\dagger \delta \mathbf{d} = \mathbf{V} \Sigma^\dagger \mathbf{U}^T \delta \mathbf{d} = \sum_{k=1}^r \frac{\delta d_{Uk}}{\mu_k} \mathbf{v}_k$, being r the rank of \mathbf{F} , and $\delta \mathbf{d}_U = \mathbf{U}^T \delta \mathbf{d}$ the noise referred to the data orthonormal base \mathbf{U} , provided by the singular value decomposition of the forward operator $\mathbf{F} = \mathbf{U} \Sigma \mathbf{V}^T$. The model perturbation $\Delta \mathbf{m}$ can be split in two parts : one that provides the stable part, and other that causes the noise in data to be amplified back into the model parameters providing equivalent solutions that in some cases are considered to be spurious, if these solutions violate some physical constraints. This second part is associated to the \mathbf{v}_k vectors of the \mathbf{V} base corresponding to the vanishing singular values $\mu_k \rightarrow 0$. This effect is known as the ill-conditioning of the discrete linear inverse problems. Also, in presence of noise the solution \mathbf{m}^\dagger found never coincides with true solution \mathbf{m}^{true} that has generated the observed data. The ill-conditioned character of the discrete inverse problems does not only originate the numerical determination of the model of lower misfit to be unstable, more important, is the origin of the uncertainty problem. The condition number of the matrix \mathbf{F} is defined as the ratio of the maximum and minimum singular values of \mathbf{F}

$$\kappa(\mathbf{F}) = \frac{\mu_{max}}{\mu_{min}} = \frac{1/\mu_{min}}{1/\mu_{max}}, \quad (5)$$

and $\kappa(\mathbf{F})$ is related to the eccentricity of the linear region of equivalence as

$$\varepsilon = \sqrt{1 - \left(\frac{\mu_{min}}{\mu_{max}} \right)^2} = \sqrt{1 - \left(\frac{1}{\kappa(\mathbf{F})} \right)^2}. \quad (6)$$

Ill-conditioned linear problems are those whose region of equivalence eccentricity tends to 1 and the equivalent models are located in oblong and elongated rectilinear valleys of the cost function topography. Thus, ill-conditioning should be also viewed as a geometric property of the linear region of equivalence (hyper-quadric). Thus, the following alternative definition might be proposed : *A linear inverse problem $\mathbf{F}(\mathbf{m}) \approx \mathbf{d}^{obs}$ is ill-conditioned if the eccentricity of its region of equivalence is close to 1.*

The noise also affects the size of the hyper-quadric, that is, the boundary of L_{tol} . In the case of an incompatible overdetermined or a rank-deficient linear system, the hyper-quadric is

$$\sum_{k=1}^r (\mu_k m_{V_k} - d_{U_k}^{obs})^2 = tol^{*2} - \sum_{k=r+1}^s d_{U_k}^{obs2}, \quad (7)$$

with $d_{U_k}^{obs} = d_{U_k}^{true} + \delta d_{U_k}$. It can be observed that

1. The noise in data does not affect the matrix of the hyper-quadric $\mathbf{F}^T \mathbf{F}$. Thus its principal axes \mathbf{V} remain unchanged.
2. The noise shifts the center of the hyper-quadric, which coincides with the least squares solution of the linear inverse problem.
3. The noise in data deforms homogeneously the topography of the unperturbed cost function in the region of equivalent models, bounded by L_{tol} .

To prove the last assert, let us consider the hyper-quadric for the noise-free case

$$\sum_{k=1}^r (\mu_k m_{V_k} - d_{U_k}^{true})^2 = tol^2 - \sum_{k=r+1}^s d_{U_k}^{true2}. \quad (8)$$

The following condition has to be fulfilled for both hyper-quadrics (with and without noise) to have the same axes length

$$tol^* = \sqrt{tol^2 + \sum_{k=r+1}^s \left[(d_{U_k}^{true} + \delta d_{U_k})^2 - d_{U_k}^{true2} \right]} = \sqrt{tol^2 + \sum_{k=r+1}^s (\delta d_{U_k}^2 + 2\delta d_{U_k} d_{U_k}^{true})}. \quad (9)$$

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

Calling $\Delta tol = \sum_{k=r+1}^s (\delta d_{U_k}^2 + 2\delta d_{U_k} d_{U_k}^{true})$, we have that the regions of equivalence will decrease in size if $\Delta tol > 0$. Particularly this happens for white noise $\sum_{k=r+1}^s \delta d_{U_k} d_{U_k}^{true} = 0$. In this case, locating the equivalence region of low misfits by means of global optimization methods becomes a harder task because it has smaller size. Nevertheless, due to the presence of noise, the posterior search is usually performed in regions of higher misfit than the ones corresponding to the noise-free case. This contributes to the feeling that finding the regions of medium misfits becomes easier under the presence of noise when in fact these region do decrease in size. Conversely, locating the global optimum of the perturbed cost function inside this region by local optimization methods is easier since the minimum is better localized. Although the noise $\delta \mathbf{d}$ does not affect the conditioning of the forward operator \mathbf{F} , it could be interpreted as a regularization since for the same value of the error tolerance the axis of the hyper-quadric become smaller. Finally, in the case of color noise Δtol might be negative and the regions of equivalence will increase in size.

These theoretical results were confirmed via a synthetic numerical example concerning a linear regression problem of the kind $y = \alpha x + \beta + \delta$, where (α, β) are the unknown model parameters and δ represents the noise. The experiment consisted in generating a dataset of 100 different (x_i, y_i) points, starting as true (noise-free) model parameters $\alpha_t = 4$ and $\beta_t = 2$. Two different levels of white Gaussian noise $\delta_1 \rightarrow N(0, 0.05)$ and $\delta_2 \rightarrow N(0, 0.075)$ were added to the noise-free data. Fig. 1 (left) shows the 9% contour lines of relative misfit for these three cases. According to Eq. (9) the original 9% equivalent region corresponds in size to the regions of 10.12% and 11.88% for δ_1 and δ_2 respectively. Also the least squares solutions in the noisy cases are shifted to the models $(\alpha_1 = 3.90, \beta_1 = 2.14)$ and $(\alpha_2 = 3.84, \beta_2 = 2.25)$.

2. The effect of the regularization in linear inverse problems

The zero-order Tikhonov regularization ([3]) is the most commonly method used to regularize ill-conditioned inverse problems, stabilizing their solution by adding an extra-term to the cost functional that penalizes the size of the solution

$$\|\mathbf{F}\mathbf{m} - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m}\|_2^2 = \|\mathbf{m}\|_2^2 + \mathbf{m}^T (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n) \mathbf{m} - 2\mathbf{m}^T \mathbf{F}^T \mathbf{d}^{obs}, \quad (10)$$

where ε^2 is the damping parameter and is close to zero. Now, the normal equations are $(\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n) = \mathbf{F}^T \mathbf{d}^{obs}$, and the operator $\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n$ becomes full rank so $(\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n)^{-1}$ exists and is continuous. This is the reason why zero-order Tikhonov's regularization stabilizes the inversion. Geometrically, Tikhonov's regularization has two main effects :

1. The linear hyper-quadric becomes a hyper-ellipsoid whose longer axes are $1/\varepsilon$ in the directions of the \mathbf{V} vectors spanning the null space of \mathbf{F} , $\ker(\mathbf{F})$.
2. The center of the hyper-quadric is shifted to $\mathbf{m}_c^\varepsilon = (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{I}_n)^{-1} \mathbf{F}^T \mathbf{d}^{obs}$, that coincides with the regularized solution of the linear system. Also the axes of the hyper-quadric are slightly rotated.

Obviously, these two effects are almost negligible when the regularization parameter tends to zero ($\varepsilon^2 \rightarrow 0$). Thus, the main effect of regularization is to limit the length of the misfit valley, causing the linear region of equivalence to become bounded. Although the zero order regularization is a practical way of stabilizing the determination of the center of the linear hyper-quadric, its location is perturbed by the presence of noise in data. In terms of the optimal solution found, the zero-order Tikhonov's regularization with no model of reference, does not add to the solution any effect from the base terms (the set of vectors $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ defined by the singular value decomposition of \mathbf{F}) spanning the null space of \mathbf{F} , since referred to the \mathbf{V} and \mathbf{U} bases the solution is

$$\mathbf{m}_{cV}^\varepsilon = (\mathbf{\Sigma}^T \mathbf{\Sigma} + \varepsilon^2 \mathbf{I}_n)^{-1} \mathbf{\Sigma}^T \mathbf{d}_U = \sum_{k=1}^r \frac{\mu_k d_{U_k}}{\mu_k^2 + \varepsilon^2} \mathbf{v}_k. \quad (11)$$

Nevertheless, when a prior model \mathbf{m}^{ref} is used in the regularization term

$$\min_{\mathbf{m} \in \mathbb{R}^n} \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2, \quad (12)$$

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

the solution referred to the \mathbf{V} base becomes

$$\mathbf{m}_V^\varepsilon = \mathbf{m}_V^{ref} + (\boldsymbol{\Sigma}^T \boldsymbol{\Sigma} + \varepsilon^2 \mathbf{I}_n)^{-1} \boldsymbol{\Sigma}^T \Delta \mathbf{d}_U^{ref}, \quad (13)$$

where $\Delta \mathbf{d}_U^{ref}$ is the reference model data misfit referred to the \mathbf{U} base, that is, $\Delta \mathbf{d}_U^{ref} = \mathbf{U}^T (\mathbf{d}^{obs} - \mathbf{F} \mathbf{m}^{ref})$. The solution can be expressed as

$$\mathbf{m}_V^\varepsilon = \sum_{k=1}^r \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}}{\mu_k^2 + \varepsilon^2} \mathbf{v}_k + \sum_{k=r+1}^n m_{V_k}^{ref} \mathbf{v}_k. \quad (14)$$

All the vectors in the \mathbf{V} base play a role in the regularized solution \mathbf{m}^ε since the coordinates that originally resided in the kernel of \mathbf{F} are now informed by the reference model \mathbf{m}_V^{ref} . Also, the coordinates in the r first vectors of the \mathbf{V} base are linear combinations of the reference model coordinates $m_{V_k}^{ref}$ and the data d_{U_k} . The weights of the linear combination are respectively ε^2 and μ_k . Thus, the reference model serves to incorporate information from the kernel. The hyper-quadric in the regularized case is

$$\begin{aligned} & \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} m_{V_k} - \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \varepsilon^2 \sum_{k=r+1}^n (m_{V_k} - m_{V_k}^{ref})^2 = \\ & = tol^2 + \sum_{k=1}^r \frac{\varepsilon^2 m_{V_k}^{ref} + \mu_k d_{U_k}^{obs}}{\sqrt{\mu_k^2 + \varepsilon^2}} - \|\mathbf{d}_U^{obs}\|_2^2 - \varepsilon^2 \sum_{k=1}^n m_{V_k}^{ref 2} \end{aligned} \quad (15)$$

The regularization also deforms the topography of the cost function with respect to the unregularized case according to the expression

$$tol_i^* = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left(tol^2 - \sum_{k=r+1}^s d_{U_k}^{obs 2} \right) + \sum_{k=1}^r \frac{\varepsilon^2}{\mu_k^2 + \varepsilon^2} (d_{U_k}^{obs} - \mu_k m_{V_k}^{ref})^2}. \quad (16)$$

In this case the value tol_i^* depends on the axe-index i , that is, this deformation is different for each axe of the hyper-quadric. The regions with error tolerance $tol > tol_{ci}(\mu_i)$

$$tol_{ci}(\mu_i) = \sqrt{\sum_{k=r+1}^s d_{U_k}^{obs 2} - \mu_i^2 \sum_{k=1}^r \frac{1}{\mu_k^2 + \varepsilon^2} (d_{U_k}^{obs} - \mu_k m_{V_k}^{ref})^2}, \quad (17)$$

decrease in size due to the effect of the regularization. The values of $tol_{ci}(\mu_i)$ depend on each singular value, and thus, the axes of the hyper-quadric associated to the biggest singular values (the directions of smaller uncertainty) might keep unchanged by the regularization, since $tol_{ci}(\mu_i)$ might not exist, being a complex number. In conclusion, in linear inverse problems the regularization deforms homogeneously (it does not depend on the model that has been considered) and anisotropically (depends on the axes of the region of equivalence) the region of equivalence.

Fig. 1 (right) shows the effect of the regularization in the previous linear regression synthetic case. The regularization parameter was set to $\varepsilon^2 = 0.5$. The regularized least squares solution found was ($\alpha = 3.84, \beta = 2.23$). According to our analysis it is possible to observe that the 9% equivalent region decreases in size with respect to the unregularized case, mainly on the directions of maximum uncertainty that are associated to the smallest singular values of the forward operator. Using Eq. (16) this anisotropic deformation can be computed, obtaining the values $tol_{min}^* = 9.01\%$ and $tol_{max}^* = 9.46\%$.

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

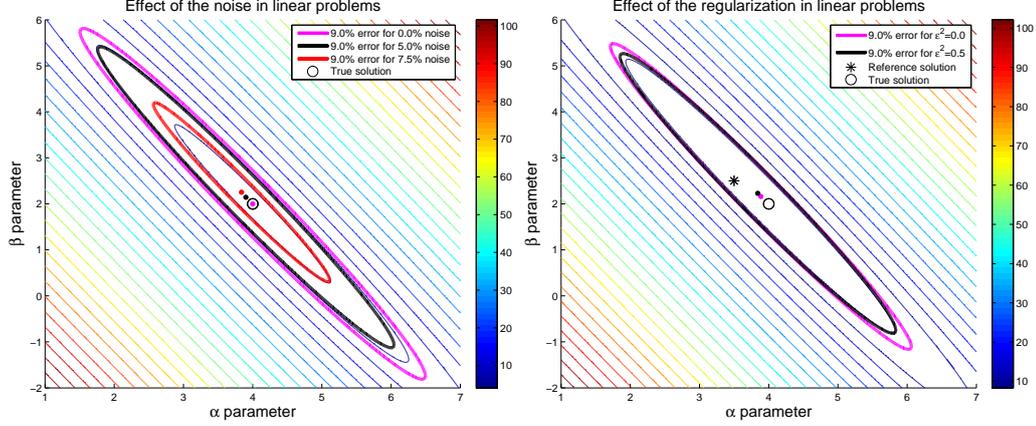


Fig. 1. Cost functions in linear problems. Left : effect of different levels of noise. Right : effect of the regularization parameter.

3. The effect of the noise in nonlinear inverse problems

To explore how the noise affects the cost function landscape, let us denote $c_p(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2$ the perturbed cost function under the presence of noise. We have

$$\begin{aligned} c_p(\mathbf{m}) &= \|\mathbf{F}(\mathbf{m}) - (\mathbf{d}^{true} - \delta\mathbf{d})\|_2^2 = \\ &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}\|_2^2 + \|\delta\mathbf{d}\|_2^2 - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] = \\ &= c(\mathbf{m}) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2. \end{aligned} \quad (18)$$

The necessary condition to find the solution model \mathbf{m}_{sol} where the minimum of $c_p(\mathbf{m})$ is achieved is $\nabla c_p(\mathbf{m}_{sol}) = 0$. Taking into account that $\nabla c_p(\mathbf{m}) = \nabla c(\mathbf{m}) - 2\mathbf{J}\mathbf{F}_{\mathbf{m}}^T \delta\mathbf{d}$, then, the necessary stationary condition implies $\nabla c(\mathbf{m}_{sol}) = 2\mathbf{J}\mathbf{F}_{\mathbf{m}_{sol}}^T \delta\mathbf{d}$, instead of $\nabla c(\mathbf{m}_{sol}) = 0$ for the free-noise case. This simple analysis shows that in presence of noise the minimum of the perturbed cost function will never coincide with the minimum in the free-noise case, which is the model that is supposed to have generated the observed data. Also, the linearized equivalent region of value tol in a model \mathbf{m}_0 is the hyper-quadric

$$\Delta\mathbf{m}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + 2(\Delta\mathbf{d} - \delta\mathbf{d})^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + (\Delta\mathbf{d} - \delta\mathbf{d})^T (\Delta\mathbf{d} - \delta\mathbf{d}) \leq tol^2, \quad (19)$$

with $\Delta\mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}$ and $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$. The center of the linearized hyper-quadric coincides with the Gauss-Newton solution of the nonlinear inverse problem in \mathbf{m}_0 ([1])

$$\Delta\mathbf{m}_{GN} = -\left(\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}\right)^\dagger \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T (\Delta\mathbf{d} - \delta\mathbf{d}) = \Delta\mathbf{m}_{GN}^{true} + \left(\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}\right)^\dagger \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \delta\mathbf{d}. \quad (20)$$

The true solution \mathbf{m}_{GN}^{true} found by the Gauss-Newton method for the free-noise case is perturbed by the noise $\delta\mathbf{d}$ through the Gauss-Newton pseudo inverse operator $\left(\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}\right)^\dagger \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T$. Now, considering the singular value decomposition of $\mathbf{J}\mathbf{F}_{\mathbf{m}_0} = \mathbf{U}\Sigma_{\mathbf{m}_0}\mathbf{V}^T$ we have

$$\left(\mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}\right)^\dagger \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \delta\mathbf{d} = \mathbf{V} (\Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0})^\dagger \Sigma_{\mathbf{m}_0}^T \mathbf{U}^T \delta\mathbf{d} = \sum_{k=1}^r \frac{\delta d_{Uk}}{\mu_k} \mathbf{v}_k, \quad (21)$$

and the noise in data is amplified due to the action of the smallest singular values μ_k of the Jacobian $\mathbf{J}\mathbf{F}_{\mathbf{m}_0}$. This effect is similar to the ill-conditioning effect shown for linear inverse problems, but in this case the

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

amplification depends on the model \mathbf{m}_0 that is considered to perform the linearization. As in the linear case, the noise does not affect the conditioning of the matrix of the linearized hyper-quadric, but deforms the topography of the cost function in a non-homogeneous way. To prove this fact, let us reasoning using the linearized region of equivalence calculated in the model \mathbf{m}_0 . Referred to the \mathbf{U} and \mathbf{V} bases provided by the singular value decomposition of the Jacobian $\mathbf{JF}_{\mathbf{m}_0}$, the linearized region of equivalence of value tol^* is

$$\Delta \mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta \mathbf{m}_V + 2 \Delta \mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta \mathbf{m}_V + \|\Delta \mathbf{b}_U\|_2^2 < tol^{*2}, \quad (22)$$

with $\Delta \mathbf{m}_V = \mathbf{V}^T \Delta \mathbf{m}$, $\Delta \mathbf{b}_U = \mathbf{U}^T (\Delta \mathbf{d} - \delta \mathbf{d})$. Then, the linearized equivalent region of value tol in the model \mathbf{m}_0 for the free-noise case is

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta d_{U_k})^2 = tol^2 - \sum_{k=r+1}^s \Delta d_{U_k}^2. \quad (23)$$

Reasoning in a similar way to the linear case, the tol equivalent region in the noise-free case will have the same size that the tol^* region under the effect of noise and the relation of both tolerances is

$$tol^*(\mathbf{m}_0; \delta \mathbf{d}) = \sqrt{tol^2 + \sum_{k=r+1}^s (\Delta b_{U_k}^2 - \Delta d_{U_k}^2)}. \quad (24)$$

Now, denoting $\Delta = \sum_{k=r+1}^s (\Delta b_{U_k}^2 - \Delta d_{U_k}^2)$, we have

1. If $\Delta > 0$ then $tol^*(\mathbf{m}_0; \delta \mathbf{d}) > tol$. Thus the models have to be found in a region with higher misfit. This case will generally apply for models with very small data misfits since $\Delta d_{U_k}^2 \rightarrow 0$. Relatively, the regions of lower misfits decrease in size with respect to the free of noise case.
2. If $\Delta < 0$ then $tol^*(\mathbf{m}_0; \delta \mathbf{d}) < tol$. This will happen for models with medium-high misfits if $\sum_{k=r+1}^s (\Delta b_{U_k}^2 - \Delta d_{U_k}^2) < 0$. Thus, in presence of noise the models of medium-high misfits will eventually increase in size.
3. This analysis depends on the model \mathbf{m}_0 that has been considered. These features will also explain why global search methods perform well under the presence of noise for inverse problems where the dimension is moderate and the forward problem is fast to be computed.

Finally, this analysis is only approximative since the higher order terms of the Taylor expansion have not been taken into account. These terms will cause that the deformation has even more non-homogeneous character depending how the noise project into these terms.

To show the effect of the noise in the nonlinear case we have analyzed a simple nonlinear regression problem $y = \alpha (1 - e^{-\beta x}) + \delta$, generating synthetically a dataset of 100 points. In addition to the true noise-free solution ($\alpha_t = 20, \beta_t = 0.1$), the problem was solved also adding a level of white noise of $\delta \rightarrow N(0, 0.05)$. The model ($\alpha_0 = 19.0, \beta_0 = 0.15$) was set as initial model for the nonlinear least squares optimization. Fig. 2 (left) shows the 9% linearized and full nonlinear contour lines of relative misfit for these two cases. The background of the figure shows the contour lines of the corresponding noise-free linearized problem. The nonlinear equivalence regions exhibit the croissant-shape which is typical from nonlinear inverse problems ([1]). The hyper-quadric represents the linearized region of equivalence around the solution that has been adopted. It is possible to observe that the linearized region of equivalence only spans locally the region of nonlinear uncertainty. It can be also observed that both regions of equivalence decrease in size when the level of noise increases.

4. The effect of the regularization in nonlinear inverse problems

Let us denote $c_p^r(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2$ the cost function with zero order regularization, where the indexes p and r stand respectively for perturbed and regularized cost function. We have

$$c_p^r(\mathbf{m}) = c^r(\mathbf{m}) - 2\delta \mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta \mathbf{d}\|_2^2, \quad (25)$$

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

where $c^r(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2$ is the regularized cost function for the free-noise case.

Imposing the stationary condition $\nabla c_p^r(\mathbf{m}_{sol}) = \nabla c^r(\mathbf{m}_{sol}) - 2\mathbf{J}\mathbf{F}_{\mathbf{m}_{sol}}^T \delta\mathbf{d} = 0$, that is, $\nabla c^r(\mathbf{m}_{sol}) = 2\mathbf{J}\mathbf{F}_{\mathbf{m}_{sol}}^T \delta\mathbf{d}$. Thus, the noise affect the solution found also in the regularized case.

The nonlinear equivalent region fulfills

$$c_p^r(\mathbf{m}) \leq tol^2 \Leftrightarrow c^r(\mathbf{m}) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{true}] + \|\delta\mathbf{d}\|_2^2 \leq tol^2. \quad (26)$$

Now, adopting a linearization of $c^r(\mathbf{m})$ in a model \mathbf{m}_0 located on the nonlinear equivalent region, it is possible to obtain the linearized region of equivalence for the regularized case

$$\begin{aligned} & \Delta\mathbf{m}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0}^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + 2(\Delta\mathbf{d} - \delta\mathbf{d})^T \mathbf{J}\mathbf{F}_{\mathbf{m}_0} \Delta\mathbf{m} + \\ & + \varepsilon^2 (\Delta\mathbf{m} - \Delta\mathbf{m}^{ref})^T (\Delta\mathbf{m} - \Delta\mathbf{m}^{ref}) + (\Delta\mathbf{d} - \delta\mathbf{d})^T (\Delta\mathbf{d} - \delta\mathbf{d}) < tol^2. \end{aligned} \quad (27)$$

Considering the singular value decomposition of the jacobian $\mathbf{J}\mathbf{F}_{\mathbf{m}_0} = \mathbf{U}\Sigma_{\mathbf{m}_0}\mathbf{V}^T$, the equation of the hyper-quadratic referred to the \mathbf{U} , \mathbf{V} bases is

$$\begin{aligned} & \Delta\mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + 2\Delta\mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + \\ & + \varepsilon^2 (\Delta\mathbf{m}_V - \Delta\mathbf{m}_V^{ref})^T (\Delta\mathbf{m}_V - \Delta\mathbf{m}_V^{ref}) + \|\Delta\mathbf{b}_U\|_2^2 < tol^2, \end{aligned} \quad (28)$$

that is

$$\begin{aligned} & \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} \Delta m_{V_k} - \frac{\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \varepsilon^2 \sum_{k=r+1}^n (\Delta m_{V_k} - \Delta m_{V_k}^{ref})^2 = \\ & = tol^2 + \sum_{k=1}^r \frac{(\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k})^2}{\mu_k^2 + \varepsilon^2} - \|\Delta\mathbf{b}_U\|_2^2 - \varepsilon^2 \sum_{k=1}^r \Delta m_{V_k}^{ref^2} \end{aligned} \quad (29)$$

Performing a similar analysis to the linear case, for the linearized equivalent regions with and without regularization we arrive at a similar relationship to the linear case

$$tol_i^*(\mathbf{m}_0) = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left(tol^2 - \sum_{k=r+1}^n \Delta b_{U_k}^2 \right) + \sum_{k=1}^r \frac{\varepsilon^2}{\mu_k^2 + \varepsilon^2} (\Delta b_{U_k} + \mu_k \Delta m_{V_k}^{ref})^2}, \quad (30)$$

where tol_i^* is the tolerance of the equivalent region in presence of regularization, and tol without it. In this case, $tol_i^*(\mathbf{m}_0)$ depends on the model \mathbf{m}_0 that has been considered to perform this analysis.

The conclusions of this analysis are the following

1. The center of the linearized hyper-quadratic provides the regularized solution of the linearized inverse problem in \mathbf{m}_0 .
2. The noise and the regularization deform nonlinearly the cost function topography in the neighborhood of \mathbf{m}_0 acting differently in each model component. Thus, the regularization and the noise in data have similar effects on the cost function landscape, although some differences exist, since the regularization also acts in an anisotropic fashion, deforming differently each axe of the linearized region of the equivalence.
3. The nonlinear and linearized equivalent regions are completely different. It has been analytically proved that the nonlinear equivalent region has a meandering valley shape. Also, there could be other low misfit basins in the cost function landscape if the forward operator is not injective ([1]). This is one of the main differences between the linear and nonlinear inverse problems. It can be easily understood that in the non-injective case the regularization might provide different solutions located in different basins of the nonlinear equivalent region, depending on the reference model that has been adopted. These solutions might not coincide with the model parameters that have generated the observed data.

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

Fig. 2 (right) shows the effect of the regularization in the previous nonlinear regression synthetic case. In this numerical example, the level of white noise was $\delta \rightarrow N(0, 0.05)$, the regularization parameter was set to $\varepsilon^2 = 0.5$, the reference model \mathbf{m}^{ref} was $(\alpha_r = 15, \beta_r = 0.07)$, and the initial model to perform the nonlinear least squares optimization was $(\alpha_0 = 19.0, \beta_0 = 0.15)$. The figure shows the contour lines of 9% relative misfit regions for the linearized and full nonlinear problems. It can be observed that the linearized regions of equivalence decrease in size anisotropically as the regularization parameter increases. The size of the nonlinear equivalent region also decreases in size less rapidly than the linearized equivalent region.

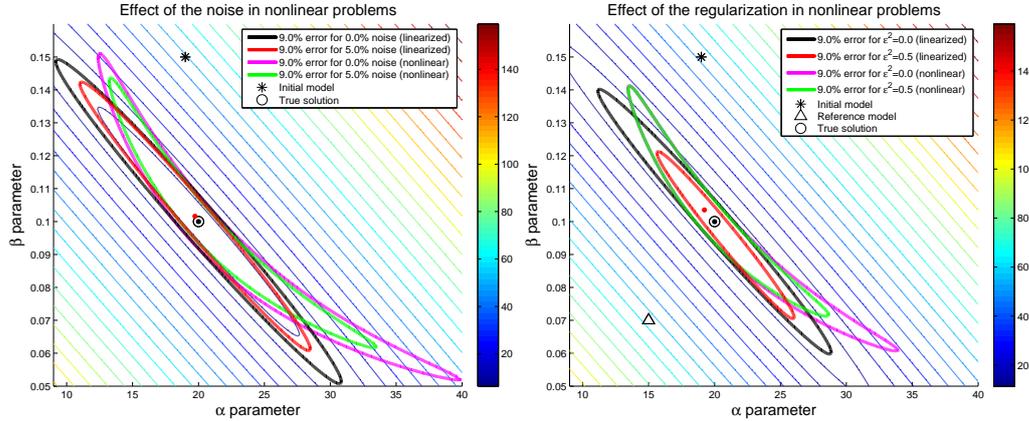


Fig. 2. Cost functions in nonlinear problems. Left : effect of noise. Right : effect of the regularization parameter.

5. Conclusions

Anyone working on inverse problems is aware of their ill-posed character. In the case of inverse problems this concept, proposed by J. Hadamard in 1902 ([4]), admits revision since it is somehow related to the use of local optimization methods. A more general definition would be to analyze the existence of families of model parameters that are compatible with the prior information that predict the observed data within the same error bounds. These models are called equivalent. Otherwise said, the ill-posed character of discrete inverse problems (ill-conditioning) originates that their solution is uncertain. Traditionally nonlinear inverse problems in discrete form have been solved via local optimization methods with Tikhonov's regularization, but linear analysis techniques failed to account for the uncertainty in the solution that has been adopted. As a result of this fact, uncertainty analysis in nonlinear inverse problems has been approached in a probabilistic framework (Bayesian approach), but these methods are hindered by the curse of dimensionality and by the high computational cost needed to solve the corresponding forward problems. Global optimization techniques are very attractive, but most of the times are heuristic and have the same limitations than Monte Carlo methods. After the discredit of deterministic methods and some initial years of Bayesian fever, now the pendulum seems to return back, because practitioners seem to be aware that the uncertainty analysis in high dimensional nonlinear inverse problems cannot (and should not be) solved via random sampling methodologies. The main reason is that the uncertainty *space* in nonlinear inverse problems has a certain geometrical structure that is embedded in the forward physics and also in the observed data. Thus, problems with structure should be approached with techniques coming from linear algebra and optimization theory. This research provides new insights to understand uncertainty from deterministic point of view, particularly the role of noise and that of the regularization in linear and nonlinear inverse problems. This understanding is a necessary step to design more efficient methods of sampling the uncertainty space taking into account the geometric structure of the space of solutions, especially for high dimensional nonlinear inverse problems with very costly forward problems, as in the geometric sampling approach ([5]).

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

References

1. J. L. Fernández Martínez, M. Z. Fernández Muñiz, and M. J. Tompkins, “On the topography of the cost functional in linear and nonlinear inverse problems,” *Geophysics*, vol. 77, no. 1, pp. W1–W15, 2012.
2. R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*. Oxford : Academic Press, 2012.
3. A. N. Tikhonov and V. Y. Arsenin, *Solution of ill-posed problems*. New York : Wiley and Sons, 1977.
4. J. Hadamard, “Sur les problèmes aux dérivées partielles et leur signification physique,” *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.
5. M. J. Tompkins, J. L. Fernández Martínez, D. L. Alumbaugh, and T. Mukerji, “Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling : Marine controlled-source electromagnetic examples,” *Geophysics*, vol. 76, no. 4, pp. F263–F281, 2011.

PUBLICATION 2

**The effect of the noise and the
regularization in inverse problems.
Geophysical implications**

Authors:

José Luis García Pallero
Juan Luis Fernández Martínez
María Zulima Fernández Muñiz
Luis Mariano Pedruelo González

15TH ANNUAL CONFERENCE OF THE INTERNATIONAL
ASSOCIATION FOR MATHEMATICAL GEOSCIENCES
(IAMG)
MADRID, SEPTEMBER 2–6, 2013
MATHEMATICS OF PLANET EARTH. LECTURE NOTES IN
EARTH SCIENCES
2014, PP. 695–698
[doi:10.1007/978-3-642-32408-6_150](https://doi.org/10.1007/978-3-642-32408-6_150)

The effect of the noise and the regularization in inverse problems. Geophysical implications

J. L. G. Pallero, J. L. Fernández-Martínez, Z. Fernández-Muñiz and L. M. Pedruelo-González

Abstract The solution of geophysical inverse problems has an intrinsic uncertainty that is mainly caused by noise in data, incomplete data sampling and simplified physics. This paper analyzes the roles of noise in data and that of the regularization for nonlinear inverse problems. We prove that noise deforms the topography of the cost function non-homogeneously, generally decreasing the regions of low misfits. As a result of this deformation, finding the global optimum by direct search methods becomes a more difficult task. Nevertheless, noise acts similarly to a regularization when local optimization methods are used. Tikhonov's regularization transforms the linearized hyper-quadric of equivalence from an elliptical cylinder to a very oblong ellipsoid in the directions that originally spanned the kernel of the linearized forward operator in absence of regularization, and also deforms anisotropically the regions of equivalence. Prior models in the regularization term serves to inform the components of the solution that locally belongs to the kernel of the Jacobian. Unfortunately regularization does not cause the disappearance of the nonlinear equivalent models. Thus a full nonlinear uncertainty analysis is still needed.

Key words: Inverse problems, noise, Tikhonov regularization, uncertainty analysis

J. L. G. Pallero

ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Madrid, Spain, e-mail: jlg.pallero@upm.es

J. L. Fernández-Martínez

Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain, e-mail: jlfm@uniovi.es

Z. Fernández-Muñiz

Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain, e-mail: zulima@uniovi.es

L. M. Pedruelo-González

Departamento de Matemáticas, Universidad de Oviedo, Oviedo, Spain, e-mail: mari-anopg@educastur.princast.es

1 Effect of the noise in nonlinear inverse problems

It has been shown in [2] that in the case of nonlinear inverse problems the nonlinear region of equivalence has a curvilinear valley shape, and depending on the injectivity of the forward operator several basins of low misfit can coexist in the cost function landscape. To explore in detail how the noise affects the cost function landscape for this kind of problems, let us adopt a linearization of the forward operator \mathbf{F} in a model \mathbf{m}_0 that belongs to the nonlinear equivalent region of value tol :

$$V_{tol} = \left\{ \mathbf{m} : c_p(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 \leq tol^2 \right\}, \quad (1)$$

where $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^s$, $\mathbf{m} \in \mathbb{R}^n$, $\mathbf{d}^{obs} = \mathbf{d}^{true} + \delta\mathbf{d} \in \mathbb{R}^s$, and $\delta\mathbf{d}$ represents the noise. We have

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_0) + \mathbf{JF}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0) + o(\|\mathbf{m} - \mathbf{m}_0\|_2). \quad (2)$$

Substituting Eq. (2) into Eq. (1) the linearized hyper-quadric L_{tol} of value tol in \mathbf{m}_0 is

$$\Delta\mathbf{m}^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} \Delta\mathbf{m} + 2(\Delta\mathbf{d} - \delta\mathbf{d})^T \mathbf{JF}_{\mathbf{m}_0} \Delta\mathbf{m} + \|\Delta\mathbf{d} - \delta\mathbf{d}\|_2^2 < tol^2, \quad (3)$$

where $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$, $\mathbf{JF}_{\mathbf{m}_0}$ is the Jacobian of \mathbf{F} in \mathbf{m}_0 , and $\Delta\mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true}$.

Taking into account ([1]) the singular value decomposition $\mathbf{JF}_{\mathbf{m}_0} = \mathbf{U}\Sigma_{\mathbf{m}_0}\mathbf{V}^T$, Eq. (3) can be written as

$$\Delta\mathbf{m}_V^T \Sigma_{\mathbf{m}_0}^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + 2\Delta\mathbf{b}_U^T \Sigma_{\mathbf{m}_0} \Delta\mathbf{m}_V + \|\Delta\mathbf{b}_U\|_2^2 < tol^2, \quad (4)$$

where $\Delta\mathbf{b}_U = \Delta\mathbf{d}_U - \delta\mathbf{d}_U = \mathbf{U}^T [\mathbf{F}(\mathbf{m}_0) - \mathbf{d}^{true} - \delta\mathbf{d}]$ and $\Delta\mathbf{m}_V = \mathbf{V}^T \Delta\mathbf{m}$.

Calling $r = \text{rank}(\Sigma_{\mathbf{m}_0})$ and μ_k the corresponding singular values of $\Sigma_{\mathbf{m}_0}$, the linearized equivalent region of value tol in model \mathbf{m}_0 is then

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta b_{U_k})^2 = tol^2 - \sum_{k=r+1}^s \Delta b_{U_k}^2, \quad (5)$$

and for the free-noise case $\delta\mathbf{d} = \mathbf{0}$

$$\sum_{k=1}^r (\mu_k \Delta m_{V_k} + \Delta d_{U_k})^2 = tol^2 - \sum_{k=r+1}^s \Delta d_{U_k}^2. \quad (6)$$

The tol equivalent region in the noisy case, called now $tol^*(\mathbf{m}_0)$, will have the same size as the tol region for the free-noise case under the condition

$$tol^{*2}(\mathbf{m}_0) - \sum_{k=r+1}^s \Delta b_{U_k}^2 = tol^2 - \sum_{k=r+1}^s \Delta d_{U_k}^2, \quad (7)$$

then

$$tol^*(\mathbf{m}_0) = \sqrt{tol^2 + \sum_{k=r+1}^s (\delta d_{U_k}^2 - 2\Delta d_{U_k} \delta d_{U_k})}. \quad (8)$$

The effect of the noise and the regularization in inverse problems

3

This analysis depends on the level of the noise and on the model \mathbf{m}_0 that is considered. By continuity of the cost function there might exist a neutral line where $tol^*(\mathbf{m}_0; \delta\mathbf{d}) = tol(\mathbf{m}_0)$, so along it no deformation takes place. Inside the region limited by this line we will have $tol^*(\mathbf{m}_0; \delta\mathbf{d}) > tol(\mathbf{m}_0)$, and regions decrease in size; and outside $tol^*(\mathbf{m}_0; \delta\mathbf{d}) < tol(\mathbf{m}_0)$, that is, regions increase in size. In conclusion, noise deforms the topography of the cost function non-homogeneously, decreasing regions of low misfits and eventually increasing the regions of medium-high misfits. The higher order terms (particularly the Hessian) that have not been taken into account in Eq. (2) also induce a non-homogeneous deformation of the nonlinear equivalence region.

In the case of linear inverse problems the deformation due to noise is homogeneous, and tol^* does not depend on any model \mathbf{m}_0 . Eq. (8) becomes

$$tol^* = \sqrt{tol^2 + \sum_{k=r+1}^s (\delta d_{Uk}^2 + 2d_{Uk}^{obs} \delta d_{Uk})}, \quad (9)$$

2 Effect of the regularization in nonlinear inverse problems

The nonlinear equivalent region of value tol for the zero-order Tikhonov regularized inverse problem is

$$V_{tol} = \left\{ \mathbf{m} : c_p(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{obs}\|_2^2 + \varepsilon^2 \|\mathbf{m} - \mathbf{m}^{ref}\|_2^2 \leq tol^2 \right\}, \quad (10)$$

where ε^2 is the regularization parameter, and \mathbf{m}^{ref} is a reference model. Performing a similar analysis as in Sect. 1, we arrive to the expression for the linearized hyper-quadratic L_{tol}

$$\begin{aligned} & \sum_{k=1}^r \left(\sqrt{\mu_k^2 + \varepsilon^2} \Delta m_{V_k} - \frac{\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k}}{\sqrt{\mu_k^2 + \varepsilon^2}} \right)^2 + \varepsilon^2 \sum_{k=r+1}^s (\Delta m_{V_k} - \Delta m_{V_k}^{ref})^2 = \\ & = tol^2 + \sum_{k=1}^r \frac{(\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k})^2}{\mu_k^2 + \varepsilon^2} - \|\Delta \mathbf{b}_U\|_2^2 - \varepsilon^2 \sum_{k=1}^r \Delta m_{V_k}^{ref^2}. \end{aligned} \quad (11)$$

Taking into account that the hyper-quadratic for the non-regularized problem corresponds to Eq. (5), the tol equivalent region in the regularized case, called now $tol^*(\mathbf{m}_0)$, will have the same size than the tol region for the non-regularized case under the condition

$$\frac{\sqrt{tol^{*2}(\mathbf{m}_0) + \Delta}}{\sqrt{\mu_i^2 + \varepsilon^2}} = \frac{\sqrt{tol^2 - \sum_{k=r+1}^s \Delta b_{U_k}^2}}{\mu_i}, \quad (12)$$

4

Pallero et al.

where

$$\Delta = \sum_{k=1}^r \frac{(\varepsilon^2 \Delta m_{V_k}^{ref} - \mu_k \Delta b_{U_k})^2}{\mu_k^2 + \varepsilon^2} - \|\Delta \mathbf{b}_U\|_2^2 - \varepsilon^2 \sum_{k=1}^r \Delta m_{V_k}^{ref^2}. \quad (13)$$

Then we arrive to

$$tol^*(\mathbf{m}_0)_i = \sqrt{tol^2 + \frac{\varepsilon^2}{\mu_i^2} \left(tol^2 - \sum_{k=r+1}^s \Delta b_{U_k}^2 \right) + \sum_{k=1}^r \frac{\varepsilon^2}{\mu_k^2 + \varepsilon^2} \left(\Delta b_{U_k} + \mu_k \Delta m_{V_k}^{ref} \right)^2}, \quad (14)$$

showing that the deformation of the linearized region of equivalence for the regularized problem is non-homogeneous and anisotropic, depending on the values of each principal axis stated by μ_i , and also on the selected model \mathbf{m}_0 . Tikhonov's regularization might also be responsible for the introduction of local minima in the cost function landscape ([2]).

3 Implications in applied geophysics and conclusions

The typical situation in geophysical inversion is to deal with data that has finite dimension, is insufficient in number, and always inaccurate due to the effect of noise. These features together with the modelling and numerical errors cause the inverse solution to be uncertain, and the geophysical inverse problem has to be solved in two different stages: solution optimization/search and appraisal. The appraisal stage (or uncertainty analysis of the solution) is motivated by the fact that the observed data does not contain enough information to determine a geophysical plausible unique solution. Practitioners tend to underestimate the uncertainty of inverse problems solutions relying on the fact that: 1) Uncertainty has in any case a random structure impossible to be known. 2) Regularization techniques are the panacea and cause the disappearance of the equivalent solutions. 3) Linearization techniques can provide a precise idea about the solution that has been adopted. All these three assumptions are not very precise, motivating in some cases the decisions that are taken based on these solutions are also incorrect, inducing big economic loses.

This paper provides new insights to understand uncertainty in the solution of inverse problem from a deterministic point of view. Further research is needed to improve the understanding about the uncertainty of the solution in inverse problems.

References

1. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems. Academic Press, Oxford (2012)
2. J.L. Fernández Martínez, M.Z. Fernández Muñoz, M.J. Tompkins, On the topography of the cost functional in linear and nonlinear inverse problems. *Geophys.* **77-1**, W1-W15 (2012)

Conclusiones y líneas futuras

5.1 Conclusiones

Las principales conclusiones de esta Tesis Doctoral son:

1. Esta Tesis proporciona nuevos conocimientos que permiten la comprensión desde un punto de vista determinista de la incertidumbre, paso necesario para el diseño posterior de métodos eficientes de muestreo de las regiones de equivalencia en problemas inversos no lineales de alta dimensión. Se propone una definición general del concepto de mal condicionamiento, basada en la estructura de las regiones de equivalencia. En esencia, la incertidumbre en los problemas inversos no lineales no necesita ser tratada como un proceso aleatorio, ya que las regiones de equivalencia tienen una interpretación algebraica natural. Los métodos de optimización local pueden ser rediseñados para que realicen muestreo mientras optimizan, para lo cual su carácter explorativo debe ser optimizado y sostenido.
2. En los problemas inversos lineales el ruido desplaza la solución estimada por los métodos de mínimos cuadrados (solución que se corresponde con el centro de la hipercuádrica de equivalencia) y deforma de manera homogénea la topografía de la función de coste. Se proporciona también una interpretación geométrica del efecto de la regularización, la cual actúa como limitación al tamaño de los ejes de la hipercuádrica de equivalencia en las direcciones del núcleo del operador directo. Geométricamente, la hipercuádrica original, que es un hipercilindro elíptico, se transforma en un hiperelipsoide cuyo centro coincide con la solución regularizada del problema. Se ha evaluado cuantitativamente la deformación de la función de coste en ambos casos, debido al ruido y debido a la regularización, y se proporcionan las correspondientes ecuaciones para su cuantificación.
3. El ruido desplaza la solución estimada por los métodos clásicos de optimización no lineal y deforma la topografía de la función de coste de manera no homogénea, en el sentido de que dicha deformación depende del modelo inicial \mathbf{m}_0 elegido para la resolución del problema. Bajo ciertas condiciones, el ruido hace que las regiones de equivalencia de buen ajuste decrezcan en extensión y, al mismo tiempo, que las de ajuste medio aumenten en tamaño. Esto dificulta la solución del problema inverso para los métodos de búsqueda global, mientras que la detección de las zonas de ajuste medio se vuelve más fácil. El papel de la regularización en las regiones de equivalencia linealizadas es similar al caso de los problemas inversos lineales, deformándolas de manera no homogénea, actuando de forma diferente en cada componente del modelo. La diferencia estriba en que en el caso no lineal dicha deformación depende del modelo inicial \mathbf{m}_0 . La regularización no causa la desaparición de los modelos equivalentes, sino que sólo sirve para mejorar localmente el

mal condicionamiento del jacobiano del operador directo. Además, las regiones de equivalencia de los problemas no lineal original y linealizado son muy distintas. La linealización aproxima únicamente de forma local dichas regiones de equivalencia y, para un nivel de error dado, puede contener modelos con un error mayor que el especificado a la vez que ignora otros modelos que sí entrarían en tolerancia. Aunque la regularización sirve para estabilizar la inversión ante la presencia de ruido, la solución regularizada puede estar lejos de la real, ya que pueden existir otras soluciones equivalentes en otras cuencas de la función de coste. Debido a estos hechos, se puede afirmar que siempre es necesario un análisis no lineal completo del problema en geofísica aplicada a fin de evaluar los posibles riesgos en todo proceso de toma de decisiones.

4. Este análisis teórico de los efectos del ruido y la regularización abre nuevas posibilidades para el diseño de algoritmos más eficientes de optimización y muestreo.
5. Se ha diseñado una aplicación práctica del algoritmo Particle Swarm Optimization (PSO) para la resolución del problema gravimétrico inverso 2D y 3D en cuencas sedimentarias, en su modalidad no lineal de determinación de la interfase de separación sedimentos-basamento. Se detalla el método de diseño del espacio de búsqueda, así como la posibilidad de uso de constreñimientos y su comparación con sus equivalentes en los métodos de optimización local. Se han testeado los miembros GPSO, CC-PSO, CP-PSO, PP-PSO y RR-PSO de la familia PSO, mostrando que se obtienen resultados realistas en cuanto a la estimación del mejor modelo y el análisis de su incertidumbre, especialmente con el miembro PP-PSO. Se ha aplicado el método a dos entornos reales, uno en 2D en Atacama (Chile) y otro en un entorno 3D en la cuenca de Argelès-Gazost (Francia), obteniéndose resultados compatibles con los obtenidos por otros métodos, pero añadiendo además una estimación realista de su incertidumbre.

5.2 Líneas futuras de investigación

El desarrollo de esta Tesis Doctoral ha abierto varias líneas de trabajo que serán abordadas en el futuro, las principales de las cuales son:

1. En esta Tesis sólo se han analizado los efectos del ruido y la regularización, pero otra fuente importante del mal planteamiento de los problemas inversos es el hecho de que el operador \mathbf{F} , debido a los errores de modelado, es sólo una aproximación de la realidad. Una importante línea futura de investigación es el estudio detallado del efecto de los errores del operador \mathbf{F} en los problemas inversos lineales y no lineales. Por otro lado, otra línea de investigación consiste en el análisis del ruido y la regularización en problemas inversos considerando funciones obetivo basadas en normas distintas a la L_2 .
2. Una línea inmediata de trabajo es la creación de un paquete de software para la solución y el análisis de incertidumbre del problema gravimétrico inverso en cuencas sedimentarias en 2D y 3D usando el algoritmo Particle Swarm Optimization. Los artículos 4 y 5 han demostrado en la práctica, mediante ejemplos sintéticos y reales, que el método es una herramienta potente para la determinación de la interfase de separación sedimentos-basamento y el análisis de su incertidumbre. El objetivo consiste en crear un software que permita la utilización de los modelos 2D y 3D, proporcionando al usuario la posibilidad de selección de parámetros como la resolución espacial del modelo, el modelo de variación

vertical y horizontal de densidad a utilizar, funciones avanzadas de representación gráfica de resultados y otras características relacionadas. Esta línea ha sido sugerida por Sylvain Bonvalot (Laboratoire GET (Université de Toulouse, CNRS, IRD, CNES) - Bureau Gravimétrique International, Toulouse, France).

3. Una línea futura importante será la integración en el método gravimétrico propuesto (especialmente en 3D) de las técnicas de reducción de la dimensión desarrolladas en [Fernández Muñiz \(2012\)](#). Aunque para problemas de tamaño medio (el problema real resuelto en el artículo 5 tiene 300 incógnitas, lo cual es un número de parámetros bastante común) el tiempo de procesado es bajo, si la cuenca en estudio comprende un área muy grande las dimensiones del problema se incrementan de manera rápida (para un área y prismas cuadrados el número total de incógnitas es N^2), lo que podría dar lugar a un problema inabordable por métodos de búsqueda global si la capacidad de cómputo es limitada. La aceleración del tiempo de cómputo es uno de los beneficios de las técnicas de reducción de la dimensión. Por otro lado, se estudiará cómo las citadas técnicas mejoran el condicionamiento del problema gravimétrico inverso, así como el modo en que la solución se ve afectada por el efecto del ruido en los datos. [Fernández-Martínez \(2015\)](#) muestra avances preliminares en este sentido.
4. Otra futura línea de investigación importante es la aplicación del algoritmo Particle Swarm Optimization y de las técnicas de reducción de la dimensión a otros problemas geofísicos. Una técnica interesante es el problema magnetotelúrico inverso, método mediante el cual se puede determinar la estructura eléctrica y magnética del subsuelo a partir de mediciones en superficie de los campos geomagnéticos y geoelectrónicos naturales de la Tierra. Este problema presenta, desde el punto de vista de esta Tesis, tres características importantes: (i) un alto nivel de ruido en los datos observados, (ii) alto número de parámetros modelo, y (iii) alto coste computacional del problema directo. Estas particularidades convierten al método magnetotelúrico en un reto a la hora de ser abordado mediante el algoritmo PSO, pero los buenos resultados obtenidos en la inversión gravimétrica indican que el problema puede ser resuelto con garantías.
5. Finalmente, una línea por explorar es el diseño y análisis de otros algoritmos de muestreo y optimización, como los de evolución diferencial y su combinación con técnicas de reducción de la dimensión.

Conclusions and future research

6.1 Conclusions

The main conclusions of this PhD Thesis are:

1. This PhD Thesis provides new insights to understand uncertainty from a deterministic point of view, which is a necessary step to design more efficient methods to sample the uncertainty region(s) of equivalent solutions in high dimensional nonlinear inverse problems. A more general definition of the ill-condition character of inverse problems, that is related to the structure of the regions of equivalence, has been proposed. In essence, uncertainty in nonlinear inverse problems needs not be treated as a random process, since the equivalences have a natural algebraic interpretation. Local optimization methods can be redesigned to perform sampling while optimizing. For that purpose, the explorative character of these algorithms should be improved and maintained.
2. For linear inverse problems, noise in data shifts the solution found by least squares methods, which is the center of the hyper-quadric of equivalence, and deforms homogeneously the topography of the cost function, in the sense that all the regions of equivalence either increase or decrease in size for different values of the error tolerance. It has been also provided a geometrical interpretation of the role of the regularization. The regularization serves to limit the sizes of the axes of the hyper-quadric in the directions corresponding to the null space of the forward operator. Geometrically, the original hyper-quadric, which is an elliptical hyper-cylinder, becomes a hyper-ellipsoid, whose center coincides with the regularized solution of the problem. The deformation of the cost function in both cases, due to noise and due to Tikhonov regularization, has been evaluated and the corresponding equations are provided.
3. Noise shifts the solution found by the typical nonlinear optimization methods, and deforms the topography of the cost function non-homogeneously in the sense that the deformation depends on the initial model \mathbf{m}_0 selected to solve the problem. Under certain conditions, the noise might decrease the size of the regions of very low misfits and at the same time increase the size of the regions of medium misfits. The inverse problem solution becomes very difficult to find using search methods, but locating the regions of medium misfits is easier. The role of the regularization in the linearized region of equivalence is similar to that shown for the linear case, deforming the linearized regions of equivalence in a non-homogeneous way, acting different in each model component. The difference is that in the nonlinear case the deformation depends on the initial model \mathbf{m}_0 . The regularization does not cause the disappearance of the nonlinear equivalent models: regularization only serves

at improving locally the ill-conditioning of the Jacobian of the forward operator. Besides, the linearized and the nonlinear problem have very different equivalence regions. The linearization only approximates locally the equivalence regions of the original nonlinear problem: for a given error level, the linearized equivalence region contain models with a higher error level than the specified level, while other models which are in the tolerance level remain outside. Although the regularization serves to stabilize the inversion under the presence of noise, the regularized solution can be relatively far from the real one since other equivalent solutions might exist in different basins of the cost function topography. Due to these facts, it can be affirmed that a full-nonlinear uncertainty analysis is always needed in applied geophysics in order to quantify risk in decision-making approaches.

4. The theoretical analysis of the effect of noise and that of the regularization opens the possibility of designing more efficient optimization and sampling algorithms.
5. A Particle Swarm Optimization application has been designed for the basement relief determination in sedimentary basins via a nonlinear gravity inversion problem schema in 2D and 3D. The search space determination was detailed, as well as the possibility of using constraints and its comparison with their counterparts in local optimization techniques. The GPSO, CC-PSO, CP-PSO, PP-PSO and RR-PSO family members were employed, showing that the PSO method gives realistic results (best model estimation and uncertainty analysis) for the proposed problem, especially the PP-PSO. Two real problems (2D in Atacama, Chile, and 3D in the Argelès-Gazost basin, France) have been solved, showing that the PSO algorithm produces compatible results with regard to the best model estimation, and, additionally, gives realistic uncertainty analysis based on the cost function sampling generated in the PSO process.

6.2 Future research

This PhD Thesis has opened several lines of work and research that will be addressed in the near future. The main ones are:

1. In this PhD Thesis only the effect of noise in data and regularization was analyzed, but another important source of ill-conditioning in inverse problems is the fact that the operator \mathbf{F} is only an approximation of the reality due to modelling errors. An important future line of research will be the detailed study of the errors effect of \mathbf{F} in linear and nonlinear inverse problem solutions. On the other hand, another future research line consists in analyzing the effects of noise and regularization in inverse problems using other norms than the Euclidean for the cost function definition.
2. An immediate line of work consists in creating a distributable software package for the solution and uncertainty analysis of the 2D and 3D gravity inversion problem in sedimentary basins using the Particle Swarm Optimization algorithm. Articles 4 and 5 have demonstrated in practice, using synthetic as well as real examples, that the method is a powerful tool for the basement relief determination and its uncertainty analysis. The goal is to create a software package that integrates the 2D and 3D models, providing the user the possibility of selection parameters such as model resolution, horizontal and vertical density variation models, advanced plotting capabilities and other related options in an easy way. This solution has been suggested by Sylvain Bonvalot (Laboratoire GET (Université de Toulouse, CNRS, IRD, CNES) - Bureau Gravimétrique International, Toulouse, France).

3. An important research line will be the integration of the dimension reduction techniques developed in [Fernández Muñiz \(2012\)](#) in the proposed gravity inversion method, specially for the 3D version. Although for medium-sized problems (in the article 5 the real problem has 300 unknowns, which is a common number of parameters) the processing time is low, if the basin under study covers a large area the dimensions of the problem will increase rapidly (for a square area and prisms, the total number of unknowns is N^2), which could lead to an unapproachable inversion using global search algorithms if the computing power at disposal is not too high. So the speed-up of the problem solution is one of the benefits that the dimension reduction techniques can provide. On the other hand, it will be studied how the dimension reduction techniques serve to improve the conditioning of the gravity inverse problem, as well as how the solution is affected by the effect of noise in data. A preliminary approach has been given in [Fernández-Martínez \(2015\)](#).
4. Other important line of research is the application of the Particle Swarm Optimization algorithm and dimension reduction techniques to other higher dimensional geophysical inverse problems solution and its uncertainty analysis is proposed. An interesting problem to address is the magnetotelluric inverse problem, which is a method for inferring the Earth's subsurface electrical structure from natural geomagnetic and geoelectric measurements at the Earth's surface. This problem has three main issues from the point of view of this PhD Thesis: (i) high level of noise in data, (ii) high number of model parameters, and (iii) high computational cost of the forward problem. These issues make the problem a challenge in order to be treated with the Particle Swarm Optimization algorithm, but the promising results obtained with the inverse gravity problem indicate that the problem might be successfully solved.
5. Finally, the design and analysis of other sampling while optimizing algorithms, such as differential evolution combined with model reduction techniques.

Appendix A

Reuse permissions and impact factor reports

In this appendix, the Elsevier and Springer reuse permissions for including the published articles and congress contributions that conform the core of this PhD Thesis are included. For space optimization, only the first page of the documents correspondent to articles 1, 2 and 3, which is the one that contains the particular document details, is shown. The complete text of the license (common to all Elsevier articles) is shown in the license document correspondent to article 4. As article 5 is under revision at the time of this document composition, no license is included for it.

In application of rules of the Universidad de Oviedo referred to the modality of Thesis as Compendium of Publications, reports about the impact factor of the journals where the articles were published (*Journal of Applied Geophysics*) are also included. This information was obtained through the web from <http://www.isiknowledge.com>, section *Journal Citation Reports*.

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Feb 18, 2016

This is a License Agreement between JLG Pallero ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	JLG Pallero
Customer address	ETSI en Topog., Geodesia y Cartografía Madrid, 28031
License number	3811820948762
License date	Feb 18, 2016
Licensed content publisher	Elsevier
Licensed content publication	Journal of Applied Geophysics
Licensed content title	From Bayes to Tarantola: New insights to understand uncertainty in inverse problems
Licensed content author	J.L. Fernández-Martínez, Z. Fernández-Muñiz, J.L.G. Pallero, L.M. Pedruelo-González
Licensed content date	November 2013
Licensed content volume number	98
Licensed content issue number	n/a
Number of pages	11
Start Page	62
End Page	72
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Análisis de la incertidumbre en problemas inversos geofísicos no lineales de alta dimensión/Uncertainty analysis in nonlinear, high-dimensional geophysical inverse problems

Expected completion date	Jul 2016
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 EUR
VAT/Local Sales Tax	0.00 EUR / 0.00 GBP
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Feb 18, 2016

This is a License Agreement between JLG Pallero ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	JLG Pallero
Customer address	ETSI en Topog., Geodesia y Cartografía Madrid, 28031
License number	3811821158649
License date	Feb 18, 2016
Licensed content publisher	Elsevier
Licensed content publication	Journal of Applied Geophysics
Licensed content title	The effect of noise and Tikhonov's regularization in inverse problems. Part I: The linear case
Licensed content author	J.L. Fernández-Martínez, J.L.G. Pallero, Z. Fernández-Muñiz, L.M. Pedruelo-González
Licensed content date	September 2014
Licensed content volume number	108
Licensed content issue number	n/a
Number of pages	10
Start Page	176
End Page	185
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No

Title of your thesis/dissertation	Análisis de la incertidumbre en problemas inversos geofísicos no lineales de alta dimensión/Uncertainty analysis in nonlinear, high-dimensional geophysical inverse problems
Expected completion date	Jul 2016
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 EUR
VAT/Local Sales Tax	0.00 EUR / 0.00 GBP
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Feb 18, 2016

This is a License Agreement between JLG Pallero ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	JLG Pallero
Customer address	ETSI en Topog., Geodesia y Cartografía Madrid, 28031
License number	3811821464983
License date	Feb 18, 2016
Licensed content publisher	Elsevier
Licensed content publication	Journal of Applied Geophysics
Licensed content title	The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case
Licensed content author	J.L. Fernández-Martínez, J.L.G. Pallero, Z. Fernández-Muñiz, L.M. Pedruelo-González
Licensed content date	September 2014
Licensed content volume number	108
Licensed content issue number	n/a
Number of pages	8
Start Page	186
End Page	193
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No

Title of your thesis/dissertation	Análisis de la incertidumbre en problemas inversos geofísicos no lineales de alta dimensión/Uncertainty analysis in nonlinear, high-dimensional geophysical inverse problems
Expected completion date	Jul 2016
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 EUR
VAT/Local Sales Tax	0.00 EUR / 0.00 GBP
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Feb 18, 2016

This is a License Agreement between JLG Pallero ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	JLG Pallero
Customer address	ETSI en Topog., Geodesia y Cartografía Madrid, 28031
License number	3811830062210
License date	Feb 18, 2016
Licensed content publisher	Elsevier
Licensed content publication	Journal of Applied Geophysics
Licensed content title	Gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization
Licensed content author	J.L.G. Pallero, J.L. Fernández-Martínez, S. Bonvalot, O. Fudym
Licensed content date	May 2015
Licensed content volume number	116
Licensed content issue number	n/a
Number of pages	12
Start Page	180
End Page	191
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Análisis de la incertidumbre en problemas inversos geofísicos no lineales de alta dimensión/Uncertainty analysis in nonlinear, high-dimensional geophysical inverse problems

Expected completion date	Jul 2016
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 EUR
VAT/Local Sales Tax	0.00 EUR / 0.00 GBP
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and

conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In

no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the

preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
 - o via their non-commercial person homepage or blog
 - o by updating a preprint in arXiv or RePEc with the accepted manuscript
 - o via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
 - o directly by providing copies to their students or to research collaborators for their personal use
 - o for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
 - o via non-commercial hosting platforms such as their institutional repository
 - o via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your

article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a [CrossMark logo](#), the end user license, and a DOI link to the formal publication on ScienceDirect. Please refer to Elsevier's [posting policy](#) for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our [open access license policy](#) for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way

as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee. Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. Other Conditions:

v1.8

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Subject: Re: Permiso para reproducir el artículo enviado al IPDO-2013
From: Olivier Fudym <olivier.fudym@cnrs-dir.fr>
Date: 12/02/16 11:39
To: José Luis García Pallero <jlg.pallero@upm.es>

Prezado José Luis García Pallero,

Como director de publicacion de IPDO-2013, le autorizo la publicacion integral en un documento academico del articulo siguiente :

Juan Luis Fernández-Martínez, J.L.G. Pallero, M. Zulima Fernández-Muñiz, and Luis Mariano Pedruelo-González, (2013), Noise, regularization and uncertainty: new insights for linear and nonlinear inverse problems. 4th Inverse Problems, Design and Optimization Symposium (IPDO-2013). Albi, France, June 26-28, 2013

Le correponde incluir la mencion "con permiso del editor" en su texto.

Saludos cordiales,

Olivier Fudym

--

Olivier Fudym
Diretor CNRS Rio

Escritorio CNRS Brasil & Cone Sul
Bureau CNRS Brésil & Cône Sud
Oficina CNRS Brasil & Cono Sur

Avenida Presidente Antônio Carlos, 58 / Sala 416
20020-010 Rio de Janeiro - RJ
Tel: +55 21 3974 6685
Cel: +55 21 99557 4621
E-mail: contato@cnrs-brasil.org
Web: www.cnrs-brasil.org

**SPRINGER LICENSE
TERMS AND CONDITIONS**

Feb 18, 2016

This is a License Agreement between JLG Pallero ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3811870501254
License date	Feb 18, 2016
Licensed content publisher	Springer
Licensed content publication	Springer eBook
Licensed content title	The Effect of the Noise and the Regularization in Inverse Problems: Geophysical Implications
Licensed content author	José Luis García Pallero
Licensed content date	Jan 1, 2014
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are the sole author of the new work
Order reference number	None
Title of your thesis / dissertation	Análisis de la incertidumbre en problemas inversos geofísicos no lineales de alta dimensión/Uncertainty analysis in nonlinear, high-dimensional geophysical inverse problems
Expected completion date	Jul 2016
Estimated size(pages)	150
Total	0.00 EUR
Terms and Conditions	

Introduction

The publisher for this copyrighted material is Springer. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

Limited License

With reference to your request to reuse material on which Springer

controls the copyright, permission is granted for the use indicated in your enquiry under the following conditions:

- Licenses are for one-time use only with a maximum distribution equal to the number stated in your request.
 - Springer material represents original material which does not carry references to other sources. If the material in question appears with a credit to another source, this permission is not valid and authorization has to be obtained from the original copyright holder.
 - This permission
 - is non-exclusive
 - is only valid if no personal rights, trademarks, or competitive products are infringed.
 - explicitly excludes the right for derivatives.
 - Springer does not supply original artwork or content.
 - According to the format which you have selected, the following conditions apply accordingly:
 - **Print and Electronic:** This License include use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
 - **Print:** This License excludes use in electronic form.
 - **Electronic:** This License only pertains to use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
- For any electronic use not mentioned, please contact Springer at permissions.springer@spi-global.com.
- Although Springer controls the copyright to the material and is entitled to negotiate on rights, this license is only valid subject to courtesy information to the author (address is given in the article/chapter).
 - If you are an STM Signatory or your work will be published by an STM Signatory and you are requesting to reuse figures/tables/illustrations or single text extracts, permission is granted according to STM Permissions Guidelines: <http://www.stm-assoc.org/permissions-guidelines/>
- For any electronic use not mentioned in the Guidelines, please contact Springer at permissions.springer@spi-global.com. If you request to reuse more content than stipulated in the STM Permissions Guidelines, you will be charged a permission fee for the excess content. Permission is valid upon payment of the fee as indicated in the licensing process. If permission is granted free of charge on this occasion, that does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.
- If your request is for reuse in a Thesis, permission is granted free of charge under the following conditions:
This license is valid for one-time use only for the purpose of defending your thesis and with a maximum of 100 extra copies in paper. If the thesis is going to be published, permission needs to be reobtained.
 - includes use in an electronic form, provided it is an author-created

version of the thesis on his/her own website and his/her university's repository, including UMI (according to the definition on the Sherpa website: <http://www.sherpa.ac.uk/romeo/>);

- is subject to courtesy information to the co-author or corresponding author.

Geographic Rights: Scope

Licenses may be exercised anywhere in the world.

Altering/Modifying Material: Not Permitted

Figures, tables, and illustrations may be altered minimally to serve your work. You may not alter or modify text in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s).

Reservation of Rights

Springer reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction and (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

License Contingent on Payment

While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Springer or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received by the date due, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Springer reserves the right to take any and all action to protect its copyright in the materials.

Copyright Notice: Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material:

"Springer book/journal title, chapter/article title, volume, year of publication, page, name(s) of author(s), (original copyright notice as given in the publication in which the material was originally published)

"With permission of Springer"

In case of use of a graph or illustration, the caption of the graph or illustration must be included, as it is indicated in the original publication.

Warranties: None

Springer makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and

disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you without Springer's written permission.

No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer, by CCC on Springer's behalf).

Objection to Contrary Terms

Springer hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

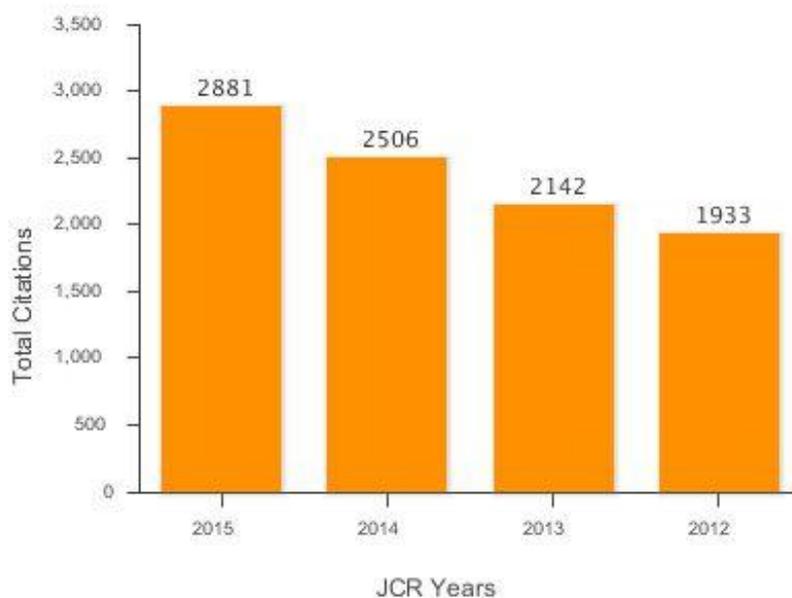
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

V 12AUG2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Journal Profile: JOURNAL OF APPLIED GEOPHYSICS

Essential Science Indicators : Total Citations Graph



Journal Citation Report : Impact factor

JCR Year	GEOSCIENCES, MULTIDISCIPLINARY			MINING & MINERAL PROCESSING			GEOSCIENCES, INTERDISCIPLINARY		
	Rank	Quartile	JIF Percentile	Rank	Quartile	JIF Percentile	Rank	Quartile	JIF Percentile
2015	116/184	Q3	37.228	10/21	Q2	54.762	NA	NA	NA
2014	96/175	Q3	45.429	6/20	Q2	72.500	NA	NA	NA
2013	106/174	Q3	39.368	7/21	Q2	69.048	NA	NA	NA
2012	94/172	Q3	45.640	3/20	Q1	87.500	NA	NA	NA
2011	77/170	Q2	55.000	3/23	Q1	89.130	NA	NA	NA
2010	91/167	Q3	45.808	5/23	Q1	80.435	NA	NA	NA
2009	77/155	Q2	50.645	4/24	Q1	85.417	NA	NA	NA
2008	66/144	Q2	54.514	5/16	Q2	71.875	NA	NA	NA
2007	80/137	Q3	41.971	7/16	Q2	59.375	NA	NA	NA
2006	64/131	Q2	51.527	2/15	Q1	90.000	NA	NA	NA
2005	82/129	Q3	36.822	4/16	Q1	78.125	NA	NA	NA
2004	102/128	Q4	20.703	6/17	Q2	67.647	NA	NA	NA
2003	94/128	Q3	26.953	6/20	Q2	72.500	NA	NA	NA
2002	66/122	Q3	46.311	6/21	Q2	73.810	NA	NA	NA
2001	97/117	Q4	17.521	8/19	Q2	60.526	NA	NA	NA
2000	NA	NA	NA	6/18	Q2	69.444	79/117	Q3	32.906
1999	NA	NA	NA	8/19	Q2	60.526	95/114	Q4	17.105
1998	NA	NA	NA	13/19	Q3	34.211	106/111	Q4	4.955

Journal Profile: JOURNAL OF APPLIED GEOPHYSICS

JCR Year	GEOSCIENCES, MULTIDISCIPLINARY			MINING & MINERAL PROCESSING			GEOSCIENCES, INTERDISCIPLINARY		
	Rank	Quartile	JIF Percentile	Rank	Quartile	JIF Percentile	Rank	Quartile	JIF Percentile
1997	NA	NA	NA	10/20	Q2	52.500	79/104	Q4	24.519

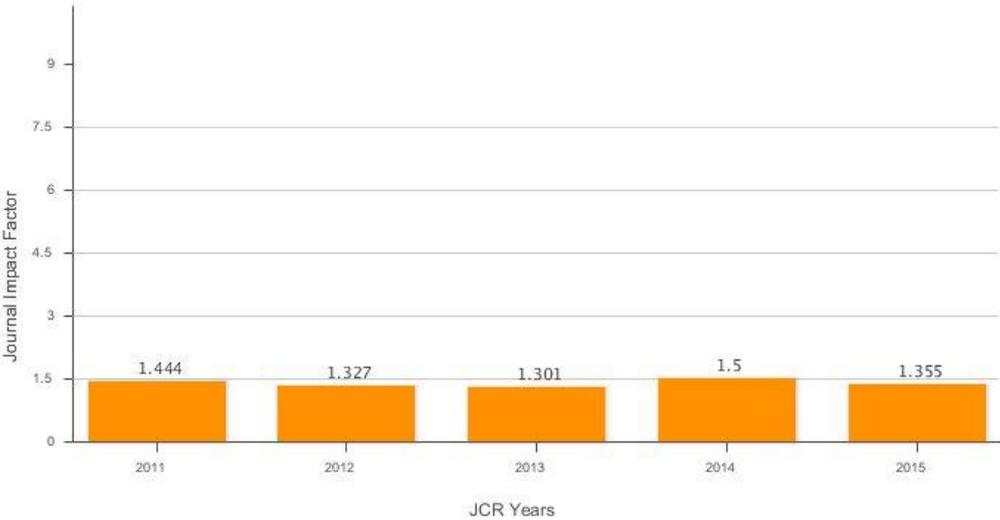
Essential Science Indicators : Total Citations

JCR Year	GEOSCIENCES
2014	135/393-Q2
2015	135/403-Q2
2012	131/385-Q2
2013	135/388-Q2

InCites™ Journal Citation Reports® 

Journal Profile: JOURNAL OF APPLIED GEOPHYSICS

Metric Trend: Journal Impact Factor



Bibliography

- Adema, Guy W.; Breckenridge, Roy M. and Sprende, Kenneth F. (2007). «Gravity, Morphology, and Bedrock Depth of the Rathdrum Prairie, Idaho». *Technical Report*, Idaho Geological Survey, University of Idaho, Moscow, Idaho 83844-3014.
[www.idahogeology.org/PDF/Technical_Reports_\(T\)/PDF/TechReport07-2_B.pdf](http://www.idahogeology.org/PDF/Technical_Reports_(T)/PDF/TechReport07-2_B.pdf)
- Al-Chalabi, Mahboud (1971). «Some studies relating to non uniqueness in gravity and magnetic inverse problems». *Geophysics*, **36(5)**, pp. 835–855. doi:[10.1190/1.1440219](https://doi.org/10.1190/1.1440219).
- (1972). «Interpretation of gravity anomalies by non-linear optimisation». *Geophysical Prospecting*, **20(1)**, pp. 1–16. doi:[10.1111/j.1365-2478.1972.tb00616.x](https://doi.org/10.1111/j.1365-2478.1972.tb00616.x).
- Alimen, Marie-Henriette (1964). *Le Quaternaire des Pyrénées de la Bigorre*. Mémoires pour servir à l'explication de la carte géologique de la France. Paris. Imprimerie nationale. ISBN 2-7159-0030-9.
- Aster, Richard C.; Borchers, Brian and Thurber, Clifford H. (2012). *Parameter Estimation and Inverse Problems*. Academic Press, New York, 2 edition. ISBN 978-0-12-385048-5.
- Backus, G. E. and Gilbert, J. F. (1967). «Numerical Applications of a Formalism for Geophysical Inverse Problems». *Geophysical Journal of the Royal Astronomical Society*, **13(1-3)**, pp. 247–276. doi:[10.1111/j.1365-246X.1967.tb02159.x](https://doi.org/10.1111/j.1365-246X.1967.tb02159.x).
- (1968). «The Resolving Power of Gross Earth Data». *Geophysical Journal of the Royal Astronomical Society*, **16(2)**, pp. 169–205. doi:[10.1111/j.1365-246X.1968.tb00216.x](https://doi.org/10.1111/j.1365-246X.1968.tb00216.x).
- (1970). «Uniqueness in the Inversion of Inaccurate Gross Earth Data». *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **266(1173)**, pp. 123–192. doi:[10.1098/rsta.1970.0005](https://doi.org/10.1098/rsta.1970.0005).
- Banks, H. T. and Groselj, Irene (1998). «A Comparison of Noise Generation Techniques and the Effects on Inverse Problem Calculations». *Technical Report*, Center for Research in Scientific Computation. North Carolina State University, Raleigh, NC 27695-8208.
<https://www.ncsu.edu/crsc/reports/reports98.html>
- Barbosa, Valéria C. F. and Silva, João B. C. (1994). «Generalized compact gravity inversion». *Geophysics*, **95(1)**, pp. 57–68. doi:[10.1190/1.1443534](https://doi.org/10.1190/1.1443534).
- Barbosa, Valéria C. F.; Silva, João B. C. and Medeiros, Walter E. (1997). «Gravity inversion of basement relief using approximate equality constraints on depths». *Geophysics*, **62(6)**, pp. 1745–1757. doi:[10.1190/1.1444275](https://doi.org/10.1190/1.1444275).

- (1999). «Gravity inversion of a discontinuous relief stabilized by weighted smoothness constraints on depth». *Geophysics*, **64**(5), pp. 1429–1437. doi:[10.1190/1.1444647](https://doi.org/10.1190/1.1444647).
- Barhen, J.; Berryman, J. G.; Borcea, L.; Dennis, J.; de Groot-Hedlin, C.; Gilbert, F.; Gill, P.; Heinkenschloss, M.; Johnson, L.; McEvelly, T.; More, J.; Newman, G.; Oldenburg, D.; Parker, P.; Porto, B.; Sen, M.; Torczon, V.; Vasco, D. and Woodward, N. B. (2000). «Optimization and Geophysical Inverse Problems». *Technical Report LBNL-46959*, University of California at Berkeley. Ernest Orlando Lawrence Berkeley National Laboratory. Earth Sciences Division. doi:[10.2172/939130](https://doi.org/10.2172/939130).
<http://www.osti.gov/scitech/biblio/939130-4TWaxg/>
- Bayes, Thomas (1763). «An Essay towards Solving a Problem in the Doctrine of Chances». *Philosophical Transactions of the Royal Society of London*, **53**, pp. 370–418. doi:[10.1098/rstl.1763.0044](https://doi.org/10.1098/rstl.1763.0044).
- Beltrão, J. F.; Silva, J. B. C. and Costa, J. C. (1991). «Robust polynomial fitting method for regional gravity estimation». *Geophysics*, **56**(1), pp. 80–89. doi:[10.1190/1.1442960](https://doi.org/10.1190/1.1442960).
- Blackman, R. B. and Tukey, J. W. (1958a). «The Measurement of Power Spectra from the Point of View of Communications Engineering – Part I». *Bell System Technical Journal*, **37**(1), pp. 185–282. doi:[10.1002/j.1538-7305.1958.tb03874.x](https://doi.org/10.1002/j.1538-7305.1958.tb03874.x).
- (1958b). «The Measurement of Power Spectra from the Point of View of Communications Engineering – Part II». *Bell System Technical Journal*, **37**(2), pp. 485–569. doi:[10.1002/j.1538-7305.1958.tb03874.x](https://doi.org/10.1002/j.1538-7305.1958.tb03874.x).
- Blakely, Richard J. (1995). *Potential Theory in Gravity and Magnetic Applications*. Cambridge University Press. ISBN 0-521-57547-8.
- Bohidar, Rabi N.; Sullivan, Jeffrey P. and Hermance, John F. (2001). «Delineating Depth to Bedrock Beneath Shallow Unconfined Aquifers: A Gravity Transect Across the Palmer River Basin». *Groundwater*, **39**(5), pp. 729–736. doi:[10.1111/j.1745-6584.2001.tb02363.x](https://doi.org/10.1111/j.1745-6584.2001.tb02363.x).
- Boschetti, Fabio; Dentith, Mike and Lis, Ron (1995). «Genetic Algorithms Incorporating A Pseudo-Subspace Method». In: *IEEE International Conference on Evolutionary Computation (vol. 2)*, pp. 557–560. Perth, WA, 24 November–1 December 1995. ISBN 0-7803-2759-4. doi:[10.1109/ICEC.1995.487444](https://doi.org/10.1109/ICEC.1995.487444).
- Boschetti, Fabio; Dentith, Mike and List, Ron (1997). «Inversion of potential field data by genetic algorithms». *Geophysical Prospecting*, **45**(3), pp. 461–478. doi:[10.1046/j.1365-2478.1997.3430267.x](https://doi.org/10.1046/j.1365-2478.1997.3430267.x).
- Bott, M. H. P. (1960). «The use of rapid digital computing methods for direct gravity interpretation of sedimentary basins». *Geophysical Journal of the Royal Astronomical Society*, **3**, pp. 63–67. doi:[10.1111/j.1365-246X.1960.tb00065.x](https://doi.org/10.1111/j.1365-246X.1960.tb00065.x).
- Cady, John W. (1980). «Calculation of gravity and magnetic anomalies of finite length-right polygonal prisms». *Geophysics*, **45**(10), pp. 1507–1512. doi:[10.1190/1.1441045](https://doi.org/10.1190/1.1441045).
- Camacho, A. G.; Fernández, J.; Arjona, A.; Pallero, J. L. G.; García-Jerez, A. and Luzón, F. (2012). *Homenaje a la profesora María Dolores Romacho Romero*. Chapter Determinación

- de la estructura 3D de la cuenca de Zafarraya mediante observación gravimétrica, pp. 51–64. Universidad de Almería. ISBN 978-84-15487-24-1.
- Camacho, Antonio G.; Carmona, Enrique; García-Jerez, Antonio; Sánchez-Martos, Francisco; Prieto, Juan F.; Fernández, José and Luzón, Francisco (2015). «Structure of Alluvial Valleys from 3-D Gravity Inversion: The Low Andarax Valley (Almería, Spain) Test Case». *Pure and Applied Geophysics*, **172**(11), pp. 3107–3121. doi:[10.1007/s00024-014-0914-8](https://doi.org/10.1007/s00024-014-0914-8).
- Camacho, Antonio G.; Fernández, José and Gottsmann, Joachim (2011). «The 3-D gravity inversion package GROWTH2.0 and its application to Tenerife Island, Spain». *Computers and Geosciences*, **37**(4), pp. 621–633. doi:[10.1016/j.cageo.2010.12.003](https://doi.org/10.1016/j.cageo.2010.12.003).
- Camacho, Antonio G.; Montesinos, Fuensanta G. and Vieira, Ricardo (2002). «A 3-D gravity inversion tool based on exploration of model possibilities». *Computers and Geosciences*, **28**(2), pp. 191–204. doi:[10.1016/S0098-3004\(01\)00039-5](https://doi.org/10.1016/S0098-3004(01)00039-5).
- Camacho, Antonio G.; Nunes, J. Carlos; Ortiz, Esther; França, Zilda and Vieira, R. (2007). «Gravimetric determination of an intrusive complex under the Island of Faial (Azores): some methodological improvements». *Geophysical Journal International*, **171**(1), pp. 478–494. doi:[10.1111/j.1365-246X.2007.03539.x](https://doi.org/10.1111/j.1365-246X.2007.03539.x).
- Carlisle, Anthony and Dozier, Gerry (2001). «An Off-The-Shelf PSO». In: *Proceedings of the Particle Swarm Optimization Workshop*, Indianapolis, USA. <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.589.485>
- Chakravarthi, V. (1995). «Gravity Interpretation of Nonoutcropping Sedimentary Basins in which the Density Contrast Decreases Parabolically with Depth». *Pure and Applied Geophysics*, **145**(2), pp. 327–335. doi:[10.1007/BF00880274](https://doi.org/10.1007/BF00880274).
- Chakravarthi, V.; Raghuram, H. M. and Singh, S. B. (2002). «3-D forward gravity modeling of basement interfaces above which the density contrast varies continuously with depth». *Computers and Geosciences*, **28**(1), pp. 53–57. doi:[10.1016/S0098-3004\(01\)00080-2](https://doi.org/10.1016/S0098-3004(01)00080-2).
- Chakravarthi, V.; Ramamma, B. and Reddy, T. Venkat (2013). «Gravity anomaly modeling of sedimentary basins by means of multiple structures and exponential density contrast-depth variations: A space domain approach». *Journal of the Geological Society of India*, **82**(5), pp. 561–569. doi:[10.1007/s12594-013-0188-2](https://doi.org/10.1007/s12594-013-0188-2).
- Chakravarthi, V.; Singh, S. B. and Babu, G. A. (2001). «INVER2DBASE—A program to compute basement depths of density interfaces above which the density contrast varies with depth». *Computers and Geosciences*, **27**(10), pp. 1127–1133. doi:[10.1016/S0098-3004\(01\)00035-8](https://doi.org/10.1016/S0098-3004(01)00035-8).
- Chakravarthi, V. and Sundararajan, N. (2004). «Automatic 3-D gravity modeling of sedimentary basins with density contrast varying parabolically with depth». *Computers and Geosciences*, **30**(6), pp. 601–607. doi:[10.1016/j.cageo.2004.03.014](https://doi.org/10.1016/j.cageo.2004.03.014).
- (2006). «Gravity Anomalies of 2.5-D Multiple Prismatic Structures with Variable Density: A Marquardt Inversion». *Pure and Applied Geophysics*, **163**(1), pp. 229–242. doi:[10.1007/s00024-005-0008-8](https://doi.org/10.1007/s00024-005-0008-8).

- (2007). «3D gravity inversion of basement relief—A depth-dependent density approach». *Geophysics*, **72**(2), pp. 123–132. doi:[10.1190/1.2431634](https://doi.org/10.1190/1.2431634).
- Chapell, Alex and Kusznir, Nick (2008). «An algorithm to calculate the gravity anomaly of sedimentary basins with exponential density-depth relationships». *Geophysical Prospecting*, **56**(2), pp. 249–258. doi:[10.1111/j.1365-2478.2007.00674.x](https://doi.org/10.1111/j.1365-2478.2007.00674.x).
- Chen, Zhaoxi and Zhang, Xiaohong Meng Sheng (2015). «3D gravity interface inversion constrained by a few points and its GPU acceleration». *Computers and Geosciences*, **84**, pp. 20–28. doi:[10.1016/j.cageo.2015.08.002](https://doi.org/10.1016/j.cageo.2015.08.002).
- Clerc, M. and Kennedy, J. (2002). «The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space». *IEEE Transactions on Evolutionary Computation*, **6**(1), pp. 58–73. doi:[10.1109/4235.985692](https://doi.org/10.1109/4235.985692).
- Constable, Steven C.; Parker, Robert L. and Constable, Catherine G. (1987). «Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data». *Geophysics*, **52**(3), pp. 289–300. doi:[10.1190/1.1442303](https://doi.org/10.1190/1.1442303).
- Cordell, L. (1973). «Gravity analysis using an exponential density-depth function—San Jacinto Graben, California». *Geophysics*, **38**, pp. 684–690. doi:[10.1190/1.1440367](https://doi.org/10.1190/1.1440367).
- Curtis, Andrew and Lomax, Anthony (2001). «Prior information, sampling distributions, and the curse of dimensionality». *Geophysics*, **66**(2), pp. 372–378. doi:[10.1190/1.1444928](https://doi.org/10.1190/1.1444928).
- Dobrin, Milton B. (1960). *Introduction to geophysical prospecting*. McGraw-Hill, 2 edition.
- Engl, Heinz W.; Hanke, Martin and Neubauer, Andreas (1996). *Regularization of Inverse Problems*. Springer. ISBN 978-0-7923-4157-4.
- Fariás, Marcelo; Charrier, Reynaldo; Carretier, Sébastien; Martinod, Joseph; Fock, Andrés; Campbell, David; Cáceres, Joaquín and Comte, Diana (2008). «Late Miocene high and rapid surface uplift and its erosional response in the Andes of central Chile (33° – 35°S)». *Tectonics*, **27**(1). doi:[10.1029/2006TC002046](https://doi.org/10.1029/2006TC002046).
- Fernández-Álvarez, J. P.; Fernández-Martínez, J. L.; García-Gonzalo, E. and Menéndez-Pérez, C. O. (2006). «Application of the particle swarm optimization algorithm to the solution and appraisal of the vertical electrical sounding inverse problem». In: *10th Annual Conference of the International Association of Mathematical Geology (IAMG '06)*, Liège, Belgium.
- Fernández-Álvarez, J. P.; Fernández-Martínez, J. L. and Menéndez-Pérez, C. O. (2008). «Feasibility Analysis of the Use of Binary Genetic Algorithms as Importance Samplers Application to a 1-D DC Resistivity Inverse Problem». *Mathematical Geosciences*, **40**(4), pp. 375–408. doi:[10.1007/s11004-008-9151-y](https://doi.org/10.1007/s11004-008-9151-y).
- Fernández-Martínez, J. L. (2015). «Model reduction and uncertainty analysis in inverse problems». *The leading edge*, **34**(9), pp. 1006–1016. doi:[10.1190/tle34091006.1](https://doi.org/10.1190/tle34091006.1).
- Fernández-Martínez, J. L.; Fernández-Muñiz, M. Z. and Tompkins, M. J. (2012a). «On the topography of the cost functional in linear and nonlinear inverse problems». *Geophysics*, **77**(1), pp. W1–W15. doi:[10.1190/geo2011-0341.1](https://doi.org/10.1190/geo2011-0341.1).

- Fernández-Martínez, J. L.; Fernández-Muñiz, Z.; Pallero, J. L. G. and Pedruelo-González, L. M. (2013a). «From Bayes to Tarantola: New insights to understand uncertainty in inverse problems». *Journal of Applied Geophysics*, **98**, pp. 62–72. doi:[10.1016/j.jappgeo.2013.07.005](https://doi.org/10.1016/j.jappgeo.2013.07.005).
- Fernández-Martínez, J. L. and García-Gonzalo, E. (2008). «The Generalized PSO: A New Door for PSO Evolution». *Journal of Artificial Evolution and Applications*, **2008**, p. 15 pages. doi:[10.1155/2008/861275](https://doi.org/10.1155/2008/861275).
- (2009). «The PSO family: deduction, stochastic analysis and comparison». *Swarm Intelligence*, **3**, pp. 245–273. doi:[10.1007/s11721-009-0034-8](https://doi.org/10.1007/s11721-009-0034-8).
- (2011a). *Handbook of Swarm Intelligence: Concepts, Principles and Applications*. Chapter What Makes Particle Swarm Optimization a Very Interesting and Powerful Algorithm. Springer. ISBN 978-3-642-17390-5. doi:[10.1007/978-3-642-17390-5_2](https://doi.org/10.1007/978-3-642-17390-5_2).
- (2011b). «Stochastic Stability Analysis of the Linear Continuous and Discrete PSO Models». *IEEE Transactions on Evolutionary Computation*, **15(3)**, pp. 405–423. doi:[10.1109/TEVC.2010.2053935](https://doi.org/10.1109/TEVC.2010.2053935).
- (2012). «Stochastic Stability and Numerical Analysis of Two Novel Algorithms of PSO Family: PP-PSO and RR-PSO». *International Journal on Artificial Intelligence Tools*, **21(3)**, p. 20 pages. doi:[10.1142/S0218213012400118](https://doi.org/10.1142/S0218213012400118).
- Fernández-Martínez, J. L.; García-Gonzalo, E.; Fernández Álvarez, José P.; Kuzma, Heidi A. and Menéndez Pérez, César O. (2010a). «PSO: A powerful algorithm to solve geophysical inverse problems. Application to a 1D-DC resistivity case». *Journal of Applied Geophysics*, **71**, pp. 13–25. doi:[10.1016/j.jappgeo.2010.02.001](https://doi.org/10.1016/j.jappgeo.2010.02.001).
- Fernández-Martínez, J. L.; García-Gonzalo, E. and Naudet, Véronique (2010b). «Particle swarm optimization applied to solving and appraising the streaming-potential inverse problem». *Geophysics*, **75(4)**, pp. WA3–WA15. doi:[10.1190/1.3460842](https://doi.org/10.1190/1.3460842).
- Fernández-Martínez, J. L.; Mukerji, T.; García-Gonzalo, E. and Fernández-Muñiz, Z. (2011). «Uncertainty assessment for inverse problems in high dimensional spaces using particle swarm optimization and model reduction techniques». *Mathematical and Computer Modelling*, **54(11–12)**, pp. 2889–2899. doi:[10.1016/j.mcm.2011.07.009](https://doi.org/10.1016/j.mcm.2011.07.009).
- Fernández-Martínez, J. L.; Mukerji, Tapan; García-Gonzalo, E. and Suman, Amit (2012b). «Reservoir characterization and inversion uncertainty via a family of particle swarm optimizers». *Geophysics*, **77(1)**, pp. M1–M16. doi:[10.1190/geo2011-0041.1](https://doi.org/10.1190/geo2011-0041.1).
- Fernández-Martínez, J. L.; Pallero, J. L. G.; Fernández-Muñiz, M. Z. and Pedruelo-González, L. M. (2013b). «Noise, regularization and uncertainty: new insights for linear and nonlinear inverse problems». In: *6th Inverse Problems, Design and Optimization Symposium (IPDO-2013)*, Albi, France, 26–28 June 2013. ISBN 979-10-91526-01-2.
- Fernández-Martínez, J. L.; Pallero, J. L. G.; Fernández-Muñiz, Z. and Pedruelo-González, L. M. (2014a). «The effect of noise and Tikhonov’s regularization in inverse problems. Part I: The linear case». *Journal of Applied Geophysics*, **108**, pp. 176–185. doi:[10.1016/j.jappgeo.2014.05.006](https://doi.org/10.1016/j.jappgeo.2014.05.006).

- (2014b). «The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case». *Journal of Applied Geophysics*, **108**, pp. 186–193. doi:10.1016/j.jappgeo.2014.05.005.
- Fernández Muñiz, María Zulima (2012). *Problemas inversos en espacios de alta dimensión y técnicas de reducción de parámetros*. Ph.D. thesis, Universidad de Oviedo. Departamento de Matemáticas.
<http://hdl.handle.net/10651/13494>
- Fullea, J.; Fernández, M. and Zeyen, H. (2008). «FA2BOUG—A FORTRAN 90 code to compute Bouguer gravity anomalies from gridded free-air anomalies: Application to the Atlantic-Mediterranean transition zone». *Computers and Geosciences*, **34**, pp. 1665–1681. doi:10.1016/j.cageo.2008.02.018.
- Gabalda, G.; Bonvalot, S. and Hipkin, R. (2003). «CG3TOOL: An interactive computer program for Scintrex CG-3M gravity data processing». *Computers & Geosciences*, **29(2)**, pp. 155–171. doi:10.1016/S0098-3004(02)00114-0.
- Gabalda, Germinal; Nalpas, Thierry and Bonvalot, Sylvain (2005). «The Base of the Atacama Gravels Formation (26°S, Northern Chile): First results from gravity data». In: *6th International Symposium on Andean Geodynamics (ISAG 2005, Barcelona)*. *Extended Abstracts*, pp. 286–289. Paris, IRD. ISBN 2-7099-1575-8.
- García González, María Esperanza (2011). *Avances en el análisis teórico del algoritmo Particle Swarm Optimization. Aplicación a la resolución de problemas inversos*. Ph.D. thesis, Universidad de Oviedo. Departamento de Matemáticas.
<http://hdl.handle.net/10651/12971>
- García-Gonzalo, E. and Fernández-Martínez, J. L. (2014). «Convergence and stochastic stability analysis of particle swarm optimization variants with generic parameter distributions». *Applied Mathematics and Computation*, **249**, pp. 286–302. doi:10.1016/j.amc.2014.10.066.
- García-Gonzalo, E.; Fernández-Martínez, J. L. and Cernea, Ana (2014). «Four-Points Particle Swarm Optimization Algorithms». *Journal of Multiple-Valued Logic and Soft Computing*, **22(3)**, pp. 239–266.
- Goldberg, David E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, 1 edition. ISBN 978-0-201-15767-3.
- González Montesinos, Fuensanta (1999). *Inversión gravimétrica 3D por técnicas de evolución. Aplicación a la isla de Fuerteventura*. Ph.D. thesis, Universidad Complutense de Madrid. Facultad de Ciencias Matemáticas. Sección departamental de Astronomía y Geodesia. Existe edición por parte del Servicio de Publicaciones del Cabildo de Fuerteventura, 2002, ISBN 84-87461-93-X.
- Gouveia, Wences P. and Scales, John A. (1998). «Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis». *Journal of Geophysical Research*, **130(B2)**, pp. 2759–2779. doi:10.1029/97jb02933.
- Granser, H. (1987). «Three-dimensional interpretation of gravity data from sedimentary basins using an exponential density-depth function». *Geophysical Prospecting*, **35(9)**, pp. 1030–1041. doi:10.1111/j.1365-2478.1987.tb00858.x.

- Hadamard, Jacques (1902). «Sur les problèmes aux dérivées partielles et leur signification physique». *Princeton University Bulletin*, **13**, pp. 49–52.
- Hinze, W. J.; von Frese, R. R. B. and Saad, A. H. (2013). *Gravity and Magnetic Exploration. Principles, Practices and Applications*. Cambridge University Press, 1 edition. ISBN 978-0-521-87101-3.
- Holland, John H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence (Complex Adaptive Systems)*. MIT University Press, 1 edition. ISBN 978-0-262-08213-6. La edición original del texto data de 1975.
- Işık, Mahir and Şenel, Hakki (2009). «3D gravity modeling of Büyük Menderes basin in Western Anatolia using parabolic density function». *Journal of Asian Earth Sciences*, **34**, pp. 317–325. doi:[10.1029/JB093iB04p03211](https://doi.org/10.1029/JB093iB04p03211).
- Isacks, B. L. (1988). «Uplift of the Central Andean Plateau and bending of the Bolivian Orocline». *Journal of Geophysical Research*, **93(B4)**, pp. 3211–3231. doi:[10.1029/JB093iB04p03211](https://doi.org/10.1029/JB093iB04p03211).
- Jackson, D. D. (1972). «Interpretation of Inaccurate, Insufficient and Inconsistent Data». *Geophysical Journal of the Royal Astronomical Society*, **28(2)**, pp. 97–109. doi:[10.1111/j.1365-246X.1972.tb06115.x](https://doi.org/10.1111/j.1365-246X.1972.tb06115.x).
- (1979). «The use of *a priori* data to resolve non-uniqueness in linear inversion». *Geophysical Journal of the Royal Astronomical Society*, **57(1)**, pp. 137–157. doi:[10.1111/j.1365-246X.1979.tb03777.x](https://doi.org/10.1111/j.1365-246X.1979.tb03777.x).
- Kaltenbacher, B.; Neubauer, A. and Scherzer, O. (2008). *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. de Gruyter. ISBN 978-3-11-020827-6.
- Kennedy, J. and Eberhart, R. (1995). «Particle swarm optimization». In: *IEEE International Conference on Neural Networks (vol. 4)*, pp. 1942–1948. Perth, WA, 27 November–1 December 1995. ISBN 0-7803-2768-3. doi:[10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968).
- Krimmel, Robert M. (1970). «Gravimetric Ice Thickness Determination, South Cascade Glacier, Washington». *Northwest Science*, **44(3)**, pp. 147–153.
- Li, Junlun; Andrews-Hanna, Jeffrey C.; Sun, Youshun; Phillips, Roger J.; Plaut, Jeffrey J. and Zuber, Maria T. (2012). «Density variations within the south polar layered deposits of Mars». *Journal of Geophysical Research*, **117(E4)**, pp. 1–13. doi:[10.1029/2011JE003937](https://doi.org/10.1029/2011JE003937).
- Litinsky, V. A. (1989). «Concept of effective density: Key to gravity determination for sedimentary basins». *Geophysics*, **54**, pp. 1474–1482. doi:[10.1190/1.1442611](https://doi.org/10.1190/1.1442611).
- MacMillan, W. D. (1958). *The theory of the potential*. Dover Publications, New York, 1 edition.
- Mantlík, František; Matias, Manuel; Lourenço, Jose; Grangeia, Carlos and Tareco, Hélder (2009). «The use of gravity methods in the internal characterization of landfills—a case study». *Journal of Geophysics and Engineering*, **6**, pp. 357–364. doi:[10.1088/1742-2132/6/4/003](https://doi.org/10.1088/1742-2132/6/4/003).

- Marquardt, Donald W. (1963). «An Algorithm for Least-Squares Estimation of Nonlinear Parameters». *Journal of the Society for Industrial and Applied Mathematics*, **11(2)**, pp. 431–441. doi:[10.1137/0111030](https://doi.org/10.1137/0111030).
- Martín Atienza, Beatriz (2001). *Modelado e inversión en 2D y 3D de anomalías gravimétricas producidas por cuerpos cuya geometría y densidad de masa se describen utilizando funciones polinómicas: aplicaciones a datos gravimétricos de Canadá y México*. Ph.D. thesis, Universidad Complutense de Madrid. Facultad de Ciencias Físicas. Departamento de Física de la Tierra, Astronomía y Astrofísica I. <http://eprints.ucm.es/4471/>
- Menke, William (2012). *Geophysical Data Analysis: Discrete Inverse Theory: MATLAB Edition*. Elsevier, 3 edition. ISBN 978-0-12-397160-9.
- Mickus, K. L.; Aiken, C. L. V. and Kennedy, W. D. (1991). «Regional-residual gravity anomaly separation using the minimum-curvature technique». *Geophysics*, **56(2)**, pp. 279–283. doi:[10.1190/1.1443041](https://doi.org/10.1190/1.1443041).
- Montesinos, F. G.; Arnosó, J. and Vieira, R. (2005). «Using a genetic algorithm for 3-D inversion of gravity data in Fuerteventura (Canary Islands)». *International Journal of the Earth Sciences*, **94(2)**, pp. 301–316. doi:[10.1007/s00531-005-0471-6](https://doi.org/10.1007/s00531-005-0471-6).
- Mortimer, Cedrig (1973). «The Cenozoic history of the southern Atacama Desert, Chile». *Journal of the Geological Society*, **129(5)**, pp. 505–526. doi:[10.1144/gsjgs.129.5.0505](https://doi.org/10.1144/gsjgs.129.5.0505).
- Mosegaard, Klaus and Tarantola, Albert (1995). «Monte Carlo sampling of solutions to inverse problems». *Journal of Geophysical Research*, **100(B7)**, pp. 12431–12447. doi:[10.1029/94JB03097](https://doi.org/10.1029/94JB03097).
- Moussirou, Bérangé (2013). *Quantification du remplissage quaternaire des vallées glaciaires des Pyrénées par la méthode gravimétrique*. Master's thesis, Université Toulouse III–Paul Sabatier.
- Murthy, I. V. Radhakrishna and Rao, D. Bhaskara (1979). «Gravity anomalies of two-dimensional bodies of irregular cross-section with density contrast varying with depth». *Geophysics*, **44(9)**, pp. 1525–1530. doi:[10.1190/1.1441023](https://doi.org/10.1190/1.1441023).
- Nabighian, M. N.; Ander, M. E.; Grauch, V. J. S.; Hansen, R. O.; LaFehr, T. R.; Li, Y.; Peirce, J. W.; Philips, J. D. and Ruder, M. E. (2005). «Historical development of the gravity method in exploration». *Geophysics*, **70(6)**, pp. 63ND–89ND. doi:[10.1190/1.2133785](https://doi.org/10.1190/1.2133785).
- Nagy, D.; Papp, G. and Benedek, J. (2000). «The gravitational potential and its derivatives for the prism». *Journal of Geodesy*, **74**, pp. 552–560. doi:[10.1007/s001900000116](https://doi.org/10.1007/s001900000116).
- Nettleton, L. L. (1976). *Gravity and magnetism in oil prospecting*. McGraw-Hill. ISBN 0-07-046303-4.
- Nocedal, Jorge and Wright, Stephen J. (1999). *Numerical Optimization*. Springer, New York, 1 edition. ISBN 978-0-387-40065-5.

- Osyrov, K.; Nichols, D.; Woodward, M.; Zdraveva, O. and Yarman, C. E. (2008). «Uncertainty and resolution analysis for anisotropic tomography using iterative eigendecomposition». In: *SEG Technical Program Expanded Abstracts*, pp. 3244–3249. Las Vegas, USA. doi:[10.1190/1.3064019](https://doi.org/10.1190/1.3064019).
- Pallero, J. L. G.; Fernández-Martínez, J. L.; Bonvalot, S. and Fudym, O. (2015). «Gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization». *Journal of Applied Geophysics*, **116**, pp. 180–191. doi:[10.1016/j.jappgeo.2015.03.008](https://doi.org/10.1016/j.jappgeo.2015.03.008).
- Pallero, J. L. G.; Fernández-Martínez, J. L.; Fernández-Muñiz, Z. and Pedruelo-González, L. M. (2014). «The Effect of the Noise and the Regularization in Inverse Problems: Geophysical Implications». In: Eulogio Pardo-Igúzquiza; Carolina Guardiola-Albert; Javier Heredia; Luis Moreno-Merino; Juan José Durán and José Antonio Vargas-Guzmán (Eds.), *Mathematics of Planet Earth. Proceedings of the 15th Annual Conference of the International Association for Mathematical Geosciences*, pp. 695–698. Springer, Madrid, Spain, 2–6 September 2013. ISBN 978-3-642-32407-9. doi:[10.1007/978-3-642-32408-6_150](https://doi.org/10.1007/978-3-642-32408-6_150).
- Parker, Paul B. (1999). *Genetic Algorithms and Their Use in Geophysical Problems*. Ph.D. thesis, University of California at Berkeley. Ernest Orlando Lawrence Berkeley National Laboratory. Earth Sciences Division. doi:[10.2172/8770](https://doi.org/10.2172/8770).
<http://www.osti.gov/scitech/biblio/8770/>
- Parker, R. L. (1977). «Understanding Inverse Theory». *Annual Review of Earth and Planetary Sciences*, **5**, pp. 35–64. doi:[10.1146/annurev.ea.05.050177.000343](https://doi.org/10.1146/annurev.ea.05.050177.000343).
- (1994). *Geophysical Inverse Theory*. Princeton University Press. ISBN 978-0-691-03634-9.
- Perrouy, Stéphane (2008). *Mesures géophysiques du remplissage sédimentaire Quaternaire dans les vallées de Bagnères de Bigorre et d'Argelès-Gazost (Hautes-Pyrénées)*. Master's thesis, Université Toulouse III–Paul Sabatier.
- Pick, Miloš; Pícha, Jan and Vyskočil, Vincenc (1973). *Theory of the Earth's Gravity Field*. Elsevier. ISBN 0-444-40939-4.
- Popper, Karl (2005). *The Logic of Scientific Discovery*. Routledge Classics. ISBN 0-203-99462-0.
- Pujol, Jose M. (2013). «The Backus-Gilbert method and their minimum-norm solution». *Geophysics*, **78(3)**, pp. W9–W30. doi:[10.1190/geo2012-0264.1](https://doi.org/10.1190/geo2012-0264.1).
- Rao, C. Visweswara; Chakravarthi, V. and Raju, M. L. (1993). «Parabolic density function in sedimentary basin modelling». *Pure and Applied Geophysics*, **140(3)**, pp. 493–501. doi:[10.1007/bf00876967](https://doi.org/10.1007/bf00876967).
- (1994). «Forward modeling: gravity nomalies of two-dimensional bodies of arbitrary shape with hyperbolic and parabolic density functions». *Computers and Geosciences*, **20(5)**, pp. 873–880. doi:[10.1016/0098-3004\(94\)90118-X](https://doi.org/10.1016/0098-3004(94)90118-X).
- Rao, C. Visweswara; Raju, M. L. and Chakravarthi, V. (1995). «Gravity modelling of an interface above which the density contrast decreases hyperbolically with depth». *Journal of Applied Geophysics*, **34(1)**, pp. 63–67. doi:[10.1016/0926-9851\(94\)00057-U](https://doi.org/10.1016/0926-9851(94)00057-U).

- Rao, D. Bhaskara (1986). «Modelling of sedimentary basins from gravity anomalies with variable density contrast». *Geophysical Journal of the Royal Astronomical Society*, **84**, pp. 207–212. doi:[10.1111/j.1365-246X.1986.tb04353.x](https://doi.org/10.1111/j.1365-246X.1986.tb04353.x).
- (1990). «Analysis of gravity anomalies of sedimentary basins by an asymmetrical trapezoidal model with quadratic density function». *Geophysics*, **55(2)**, pp. 226–231. doi:[10.1190/1.1442830](https://doi.org/10.1190/1.1442830).
- René, R. M. (1986). «Gravity inversion using open, reject, and “shape-of-anomaly” fill criteria». *Geophysics*, **51(4)**, pp. 988–994. doi:[10.1190/1.1442157](https://doi.org/10.1190/1.1442157).
- Sanyi, Yuan; Shangxu, Wang and Nan, Tian (2009). «Swarm intelligence optimization and its application in geophysical data inversion». *Applied Geophysics*, **6(2)**, pp. 166–174. doi:[10.1007/s11770-009-0018-x](https://doi.org/10.1007/s11770-009-0018-x).
- Scales, John A. and Snieder, Roel (1997). «To Bayes or not to Bayes?» *Geophysics*, **62(4)**, pp. 1045–1046. doi:[10.1190/1.6241045.1](https://doi.org/10.1190/1.6241045.1).
- (2000). «The Anatomy of Inverse Problems». *Geophysics*, **65(6)**, pp. 1708–1710. doi:[10.1190/geo2000-0001.1](https://doi.org/10.1190/geo2000-0001.1).
- Scales, John A. and Tenorio, Luis (2001). «Prior information and uncertainty in inverse problems». *Geophysics*, **66(2)**, pp. 389–397. doi:[10.1190/1.1444930](https://doi.org/10.1190/1.1444930).
- Sen, Mrinal K. and Stoffa, Paul L. (2013). *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press, 2 edition. ISBN 978-1-107-01190-8.
- Shaw, Ranjit and Srivastava, Shalivahan (2007). «Particle swarm optimization: A new tool to invert geophysical data». *Geophysics*, **72**, pp. F75–F83. doi:[10.1190/1.2432481](https://doi.org/10.1190/1.2432481).
- Shumskiy, P. A. (1960). «Density of glacier ice». *Journal of Glaciology*, **3(27)**, pp. 568–573.
- Silva, João B. C.; Costa, Denis C. L. and Barbosa, Valéria C. F. (2006). «Gravity inversion of basement relief and estimation of density contrast variation with depth». *Geophysics*, **71(5)**, pp. J51–J58. doi:[10.1190/1.2236383](https://doi.org/10.1190/1.2236383).
- Silva, João B. C.; Oliveira, Alexandre S. and Barbosa, Valéria C. F. (2010). «Gravity inversion of 2D basement relief using entropic regularization». *Geophysics*, **75(3)**, pp. I29–I35. doi:[10.1190/1.3374358](https://doi.org/10.1190/1.3374358).
- Silva, João B. C.; Teixeira, Wlamir A. and Barbosa, Valéria C. F. (2009). «Gravity data as a tool for landfill study». *Environmental Geology*, **57**, pp. 749–757. doi:[10.1007/s00254-008-1353-6](https://doi.org/10.1007/s00254-008-1353-6).
- Skeels, D. C. (1947). «Ambiguity in gravity interpretation». *Geophysics*, **12(1)**, pp. 43–56. doi:[10.1190/1.1437295](https://doi.org/10.1190/1.1437295).
- Snieder, R. (1991). «An extension of Backus-Gilbert theory to nonlinear inverse problems». *Inverse Problems*, **7(3)**, pp. 409–433. <http://iopscience.iop.org/article/10.1088/0266-5611/7/3/008>
- Stern, Timothy A. (1978). «Gravity Survey of the Taylor Glacier, Victoria Land, Antarctica». *Technical Report 8*, Geology Department, Victoria University of Wellington. www.victoria.ac.nz/antarctic/pdf/ADS05.pdf

- Sweilam, N. H.; Gobashy, M. M. and Hashem, T. (2008). «Using Particle Swarm Optimization with Function Stretching (SPSO) For Integrating Gravity Data: A Visibility Study». *Proceedings of the Mathematical and Physical Society of Egypt*, **86(2)**, pp. 259–281.
- Tarantola, Albert (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia. ISBN 978-0-89871-572-9.
<http://www.ipgp.fr/~tarantola/Files/Professional/Books/index.html>
- (2006). «Popper, Bayes and the inverse problem». *Nature Physics*, **2**, pp. 492–494. doi:10.1038/nphys375.
<http://www.ipgp.fr/~tarantola/>
- Tarantola, Albert and Valette, Bernard (1982a). «Generalized nonlinear inverse problems solved using the least squares criterion». *Reviews of Geophysics*, **20(2)**, pp. 219–232. doi:10.1029/RG020i002p00219.
<http://www.ipgp.fr/~tarantola/>
- (1982b). «Inverse Problems = Quest for Information». *Journal of Geophysics*, **50**, pp. 159–170.
<http://www.ipgp.fr/~tarantola/>
- Tekkeli, Anisya B. and Bal, Oya Tarhan (2012). «Modeling of Gravity Data Using Particle Swarm Optimization». In: *International Geophysical Conference and Oil & Gas Exhibition*, pp. 1–4. Istanbul, Turkey, 17–19 September 2012. doi:10.1190/IST092012-001.106.
- Telford, W. M.; Geldart, L. P.; Sheriff, R. E. and Keys, D. A. (1976). *Applied Geophysics*. Cambridge University Press. ISBN 978-0-5213-3938-4.
- Tiberi, Christel; Ebinger, Cynthia; Ballu, Valéria; Stuart, Graham and Oluma, Befekadu (2005). «Inverse models of gravity data from the Red Sea–Aden–East African rifts triple junction zone». *Geophysical Journal International*, **163**, pp. 775–787. doi:10.1111/j.1365-246X.2005.02736.x.
- Tibshirani, Robert (1996). «Regression Shrinkage and Selection via the Lasso». *Journal of the Royal Statistical Society. Series B (Methodological)*, **58(1)**, pp. 267–288.
<https://www.jstor.org/stable/2346178>
- Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. V. H. Winston and Sons, Washington, DC. ISBN 978-0-4709-9124-4.
- Tompkins, Michael J.; Fernández Martínez, Juan L.; Alumbaugh, David L. and Mukerji, Tapan (2011a). «Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: Marine controlled-source electromagnetic examples». *Geophysics*, **76(4)**, pp. F263–F281. doi:10.1190/1.3581355.
- Tompkins, Michael J.; Fernández Martínez, Juan Luis and Fernández Muñoz, Zulima (2011b). «Marine electromagnetic inverse solution appraisal and uncertainty using model-derived basis functions and sparse geometric sampling». *Geophysical Prospecting*, **59(5)**, pp. 947–965. doi:10.1111/j.1365-2478.2011.00955.x.

- (2013). «Comparison of sparse-grid geometric and random sampling methods in non-linear inverse solution uncertainty estimation». *Geophysical Prospecting*, **61**(1), pp. 28–41. doi:[10.1111/j.1365-2478.2012.01057.x](https://doi.org/10.1111/j.1365-2478.2012.01057.x).
- Toushmalani, Reza (2013a). «Comparison result of inversion of gravity data of a fault by particle swarm optimization and Levenberg-Marquardt methods». *SpringerPlus*, **2**, p. 462. doi:[10.1186/2193-1801-2-462](https://doi.org/10.1186/2193-1801-2-462).
- (2013b). «Gravity inversion of a fault by Particle Swarm Optimization (PSO)». *SpringerPlus*, **2**, p. 315. doi:[10.1186/2193-1801-2-315](https://doi.org/10.1186/2193-1801-2-315).
- Venteris, Erik and Miller, Maynard (1993). «Gravitational profiles on the Taku glacier system». *Open File Report*, Glaciological and Arctic Sciences Institute, University of Idaho.
- Vernon, Antoine; Nalpas, Thierry; Dabard, Marie-Pierre; Mpodozis, Constantino; Riquelme, Rodrigo and Hérail, Gérard (2005). «Preservation of the Miocene Atacama Gravels: Climatic/depositional-erosional balance in the El Salvador area, North Chilean Andes». In: *6th International Symposium on Andean Geodynamics (ISAG 2005, Barcelona)*. *Extended Abstracts*, pp. 783–786. Barcelona, Spain. ISBN 2-7099-1575-8.
- Vugrin, Kay E. (2005). *On the Effects of Noise on Parameter Identification Optimization Problems*. Ph.D. thesis, Faculty of the Virginia Polytechnic Institute and State University. <http://vtechworks.lib.vt.edu/handle/10919/27515/>
- Wolpert, David H. and Macready, William G. (1997). «No Free Lunch Theorems for Optimization». *IEEE Transactions on Evolutionary Computation*, **1**(1), pp. 67–82. doi:[10.1109/4235.585893](https://doi.org/10.1109/4235.585893).
- Wziontek, H.; Wilmes, H. and Bonvalot, S. (2012). «AGrav: An International Database for Absolute Gravity Measurements». In: Steve Kenyon; Christina Maria Pacino and Urs Marti (Eds.), *Geodesy for Planet Earth: Proceedings of the 2009 IAG Symposium*, pp. 1037–1042. Buenos Aires, Argentina, 31 August–4 September 2009. ISBN 978-3-642-20338-1. doi:[10.1007/978-3-642-20338-1_130](https://doi.org/10.1007/978-3-642-20338-1_130).
- Xu, Ya; Hao, Tianyao; Li, Zhiwei; Duan, Qiuliang and Zhang, Lili (2009). «Regional gravity anomaly separation using wavelet transform and spectrum analysis». *Journal of Geophysics and Engineering*, **6**(3), pp. 279–287. doi:[10.1088/1742-2132/6/3/007](https://doi.org/10.1088/1742-2132/6/3/007).
- Zhdanov, Michael S. (2015). *Inverse Theory and Applications in Geophysics*. Elsevier, 2 edition. ISBN 978-0-444-62674-5.