UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

CHANNELIZED FACIES RECOVERY BASED ON WEIGHTED SPARSE REGULARIZATION

TESIS PARA OPTAR AL GRADO DE MAGISTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

HERNÁN ALBERTO CALDERÓN AMOR

PROFESOR GUÍA:
DR. JORGE SILVA SÁNCHEZ
DR. JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:
DR. MARCOS ORCHARD CONCHA
DR. PABLO IRARRÁZAVAL MENA

SANTIAGO DE CHILE
January, 2016

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGISTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA
POR: HERNÁN ALBERTO CALDERÓN AMOR
FECHA: JANUARY, 2016
PROFESOR GUÍA: DR. JORGE SILVA SÁNCHEZ, DR. JULIÁN ORTIZ CABRERA

CHANNELIZED FACIES RECOVERY BASED ON WEIGHTED SPARSE REGULARIZATION

Comprender los fenómenos de nuestro planeta es esencial en diversos problemas de estimación y predicción, tales como minería, hidrología y extracción de petróleo. El principal inconveniente para resolver problemas inversos en geoestadística es la falta de datos, lo que imposibilita la generación de modelos estadísticos confiables. Debido a esto, es necesario incorporar información adicional para estimar las variables de interés en locaciones no medidas, como por ejemplo utilizando imágenes de entrenamiento.

Esta Tesis aborda el problema de interpolación espacial de estructuras de canal basada en teorías de representación *sparse* de señales y estadísticos multipuntos. El trabajo se inspira en la teoría de *Compressed Sensing* (CS), la cual ofrece un nuevo paradigma de adquisición y reconstrucción de señales, y simulación multipunto (MPS), técnica que provee realizaciones realistas de diversas estructuras geológicas. Esta Tesis se motiva por estos dos enfoques, explorando la fusión de ambas fuentes de información, tanto geológica-estructural como la descomposición de dicha estructura en un dominio transformado.

La principal contribución de este trabajo es el uso de algoritmos MPS para incorporar información a priori al algoritmo de reconstrucción, convirtiendo información geológica en información de señal. El algoritmo MPS es utilizado para estimar el soporte de la estructura subyacente, identificando las posiciones de los coeficientes transformados significativos y generando un ranking para los elementos de la base DCT (Discrete cosine transform). Este ranking es usado para la creación de una matriz de pesos, la cual impone una particular estructura directamente en el algoritmo de reconstrucción. Esta metodología es validada mediante el estudio de tres modelos de canal.

Respecto a los resultados, primero se estudian diversas definiciones de la matriz de pesos para determinar la mejor configuración. Segundo, se estudia un enfoque multiescala de regularización *sparse* con el propósito de mejorar los desempeños clásicos de minimización en norma ℓ_1-ponderada. Con ello, se valida el uso de varias reconstrucciones a distintos niveles de escala para reducir los artefactos inducidos en la reconstrucción a imagen completa. Finalmente, el método es comparado con diversas técnicas de interpolación. De este análisis, se observa que el método propuesto supera a las técnicas convencionales de regularización en norma ℓ_1, tanto con pesos como sin ponderadores, al igual que el algoritmo multipunto utilizado. Esto valida la hipótesis sobre la complementariedad de las informaciones de patrones estadísticos y estructuras de señal. Para cada modelo, el método es capaz de inferir la estructura predominante de canal, incluso en un escenario inferior al 1% de datos adquiridos.

Finalmente, se han identificado algunas posibles áreas de investigación futura. Algunas de estas posibles aristas son: CS binario, dada la naturaleza binaria de los modelos estudiados; algoritmos *greedy* para la extensión al análisis 3D; y métodos adaptativos de CS para resolver simultáneamente el problema de localización de sondajes y reconstrucción de señales.

Summary

Understanding the phenomena of our planet is essential in several forecasting problems, such as mining, hydrology and oil industry. The main roadblock at solving ill-posed inverse problems in geosciences is the lack of data, which precludes the construction of meaningful models. Then, it is necessary to include additional information sources to estimate the variables of interest.

This Thesis addresses the problem of channelized two-facies image recovery based on sparse signal regularization techniques and multiple-points statistics. This work is inspired by the recent Compressed Sensing (CS) theory, which provides a new paradigm of data acquisition and signals reconstruction, and multiple-points simulation (MPS), which offers realistic realizations of several geological structures. The Thesis is motivated by these frameworks, exploring the fusion of both structural information: geological and signal information.

The major contribution of this work is the use of MPS algorithms to provide geological prior information, transforming the conventional multiple-points statistics in information for image recovery. The MPS algorithm is used to learn the support of the underlying structure, identifying the location of the relevant transform coefficients and generating a prior rank for the DCT (Discrete cosine transform) atoms. From this rank, a weighting matrix is performed to impose a particular geological structure directly in the sparse-promoting algorithm. This methodology was validated through the study of three different channelized models.

Regarding the empirical results, several algorithmic aspects were studied to validate the proposed method. First, the impact of different weighting matrix definitions is analyzed. Secondly, an average multiscale sparse regularization approach is adopted to improve the standard weighted ℓ_1-minimization. Furthermore, the use of several recovered images at different scales is validated to reduce non-channelized artefacts present in the full-image reconstruction. Finally, the proposed method is compared with others techniques in the literature, including the MPS algorithm. From this analysis, it is shown that this approach outperforms the standard unweighted and weighted ℓ_1-minimizers as well as the MPS algorithm. This new method is able to recover the dominant channel structure for the three analyzed models even when less than 1% of measurements are available.

As final remark, some directions are considered as attractive future works. First, it is interesting the development of techniques that impose binary values directly in the recovery algorithm, exploiting the categorical nature of these models. Secondly, the possibility of using greedy algorithms is attractive on the line of extending these approaches to the 3D analysis. Finally, the new adaptive CS framework provides an interesting research branch to accomplish simultaneously the problems of optimal well placement and image recovery.

Acknowledgements

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General background

Understanding the phenomena and structures of our planet is essential in several forecasting problems present in many disciplines, such as mining, hydrology and oil extraction. The knowledge of geological structures is often expressed in terms of a model, where estimating the parameters of this model turns out to be a fundamental problem in many earth exploration scenarios. Unfortunately, in several applications, it is not possible to establish an analytical expression for these physical or statistical models, making the forecasting process very challenging. In particular, the lack of data is presented as one of the main issues in geological inverse problems. Obtaining meaningful statistical models is extremely difficult and infeasible in several cases. This problem leads to many undesirable consequences, for instance statistical mismodeling, unrealistic solutions and forecasting results very far from actual reality.

To some extent, most conventional approaches that treat with the lack of information can be framed into two types of inference techniques: the minimization of a smooth objective function, typically variations of the mean square error (variogram-based models), or the simulation according to patterns presented in a training image (multiple-point statistics) [1–3]. Variograms-based estimation methods are usually used in the study of continuous variables and promote smooth solutions. Furthermore, variograms can be used in the simulation framework, providing uncertainty analyses and non-smooth solutions. However, these simulation methods result in the same main problem of traditional approaches: poor geological realism. On the other hand, multiple-points simulation is widely used in categorical variables, which provides several plausible realizations of a field by the extraction of high-order conditional statistics from training images, leading to more geologically realistic models. These methods have been used in many inverse earth problems, providing outstanding outcomes even in critical under-sampling environments. Nevertheless, all information sources used by both approaches proceed from spatial information; i.e., the estimated models utilized in traditional geostatistical methods capture only spatial features such as correlations, variograms and conditional statistical patterns. This motivates the use of image processing techniques,

where analysis is accomplished in an alternative transform domain. This research branch is particularly interesting because some features might be efficiently represented in a transform space, for instance complexity and orientation of the medium [4, 5]. Thus, spatial and transform signal information are presented as complementary information sources, and could be exploited jointly to provide a new paradigm in geostatistical framework.

1.2 Specific background

Nowadays, image recovery is an active research area in image processing. The aim of these methods is restoring the original signals based on models extracted from the image itself and signal prior information. Some example of image recovery approaches are: image denoising, deblurring, inpainting, zooming and super-resolution [6–8]. The geological field can be modeled as an image, and consequently, the problem of spatial interpolation can be posed as an image recovery instance.

Compressed Sensing (CS) has been introduced as a powerful theory of sampling and signal reconstruction [9–13]. In contrast to classical sampling approaches, CS provides a simultaneous acquisition and compression scheme based on a random sampling process. This theory is aimed at solving underdetermined linear systems supported over the idea that the signal to be recovered can be represented with a small fraction of coefficients with respect to the ambient dimension. Then, the sampling rate increases linearly as a function of the significant coefficients of the signal, instead of the signal dimension. This is instrumental in geological inverse problem, where measurements are expensive and an exhaustive sampling is not possible.

Despite the success of sparse regularization approaches in different geological applications [4, 14, 15], important limitations have been identified due to the critical sub-sampling rate. From first results of CS, the incorporation of prior information has been a priority in order to decrease the sampling ratio established by standard sparse signal recovery algorithms [16–18]. In multiple-point simulation framework, the manner to incorporate prior information is by using training images to learn statistical patterns and reproduce them in the simulation stage. This thesis is motivated by these ideas, exploring the fusion of both information sources: geological prior models, obtained from training images; and signal information, in terms of the redundancy of the underlying field in an alternative transform domain.

1.3 Hypotheses

This thesis is supported by the following hypotheses:

- *Geological images present compressible representations in some alternative domains.* This property is the main foundation of several image recovery strategies, and can be

exploited by ℓ_1-regularization algorithms. In the context of geological applications, two-facies channelized images exhibit features (piecewise nature, orientation and complexity) that might be efficiently described by some dictionaries, for instance Wavelet basis and Discrete cosine transform (DCT).

- *The location of the significant transform coefficients is persistent across several realizations of the geological field.* Geological images generated from a statistical or physical model present common spatial structure, such as orientation and complexity in the case of two-facies channelized models. This common structure in the spatial domain may result in a particular decomposition in the transform domain. Thus, it could be possible to infer the support of the actual underlying channel structure from the study of several realizations of the field.
- *It is possible to incorporate a prior geological model in the sparse promoting algorithm to enhance the standard ℓ_1 minimizer.* Conceptually, the main issue to be solved in inverse problems is the estimation of the support of the signal; i.e., the location of the significant coefficients. Then, a prior geological model might reduce the searching space, forcing a consistent support with the prior model and decreasing the number of measurements needed to achieve an acceptable distortion level.

1.4 Main objective

The main objective of this thesis is studying, developing and implementing the fusion of multiple-point simulation and ℓ_1-regularization technique to solve inverse problems in geosciences. Particularly, the thesis is focused on the reconstruction of channelized geological fields in the subsurface from pixel-based measurements.

1.5 Specific objectives

- Studying the state of the art of sparse regularization framework with prior information.
- Introducing a methodology to incorporate geological prior information in sparse signal recovery algorithms.
- Exploring the weighted sparse minimization in different channelized structures.
- Obtaining performance metrics, determining operating regimes, uncertainty and robustness of the method in terms of the complexity of the model and sampling ratio.
- Comparing the obtained results with other methods in the literature.

1.6 Structure of the thesis

The document is organized as follows:

- Chapter 2 provides the state of the art of the problem. This section presents the variogram-based method, multiple-point simulation techniques, and basics results on sparse regularization theory.
- Chapter 3 presents the methodology. This chapter includes the problem formulation, the study of the statistical support induced by multiple-points statistics, a methodology for incorporating the prior model in the recovery algorithm, the penalization matrix proposed in this thesis and the experimental settings.
- Chapter 4 provides the main outcomes and the analyses of the proposed method. In addition, other recovery algorithms are presented in order to illustrate the gain of the proposed method, including the widely used multiple-point simulation.
- Finally, chapter 5 provides final remarks and future work.

Chapter 2

State of the art

This chapter is aimed at providing the state-of-the-art of the problem addressed in this thesis. First, the spatial interpolation based on classical estimation theory is introduced as the conventional method used in the literature. Second, a brief review of multiple-points simulations (MPS) is presented. Specifically, the motivation, main ideas of MPS and the SNESIM algorithm are included. Finally, the sparse regularization theory is provided. In this section the sparse and compressible representation, basics of CS theory, weighed ℓ_1-minimization and related works are developed.

2.1 Spatial estimation: Kriging interpolator

Spatial interpolation consists of estimating the value of a regionalized variable in a unmeasured location. Estimation theory is the most predominant approach used for spatial interpolation in geoscience. These approaches model a set of spatial variables as random variables, and use random function optimization theory to predict the feature of interest. On this line, *kriging* [1] is the most commonly used method in geostatistics to spatial estimation. From first results, several approaches based on this mathematical framework have been developed. In this section, basics of the standard kriging method are provided.

To formulate the problem, it is necessary to establish some desired condition and hypotheses. As previously stated, the variable is modeled as a random variable, and consequently, present a PDF at every location. Additionally, the variables are spatially correlated. In this stochastic scenario, stationarity is the strongest hypothesis considered by kriging-based methods. In geological inverse problems, there is a single measurement of the variable at each location. Therefore, without stationarity hypothesis, it is not possible to infer the statistics of the random variable, and consequently, any estimation process could not be addressed. In particular, stationarity in spatial problems stipulates that the mean-value of the variable is independent of the location and the correlation between variables depends only on the distance between the specific locations [1]. In other words, these methods model the variables

of interest as ergodic processes.

Regarding the estimator construction, kriging methods propose linear structures; i.e.,

$$\hat{\mathcal{X}}(r_0) = \lambda_0 + \sum_{r_i \in \mathcal{R}_0} \lambda_i X(r_i) \tag{2.1}$$

where $\hat{\mathcal{X}}(r_0)$ is the estimation of the variable $\mathcal{X}$ at the position r_0, $\{\lambda_j\}$ a set of weights, $X(r_i)$ the measurement at the location r_i and $\mathcal{R}_0$ a neighborhood around r_0. Then, the problem consists of determining the optimal set of weights $\{\lambda_j\}$.

Concerning the conditions on the estimator, the first restriction rely on the unbiasedness condition; i.e.,

$$\mathbb{E}\{\hat{\mathcal{X}}(r_0) - \mathcal{X}(r_0)\} = 0. \tag{2.2}$$

In other words, on average, the estimator will be equal to the actual value. Unfortunately, the expectation has to be calculated over the statistics of $\mathcal{X}$, which are not available. Nevertheless, the stationarity hypothesis allows addressing this problem. The expected value of $\mathcal{X}$ is independent of the location, and as a result, it is possible to estimate it from the samples.

A second condition consists of regularizing the solution and restricting the pursuit space. This condition is related to the optimality of the method. In particular, kriging loss function corresponds to the classical *mean square error* (MSE); i.e.,

$$\text{minimize} \;\; \mathbb{E}\{(\hat{\mathcal{X}}(r_0) - \mathcal{X}(r_0))^2\}. \tag{2.3}$$

MSE is one of the most common loss function and has been used in several engineering applications as well as in research development. This type of penalization term presents several advantages, including closed solutions in many cases, easy and fast evaluation, and it provides a practical and theoretical mathematical framework. As in the unbiasedness condition, ergodicity is essential to evaluate the loss function. Nevertheless, it is necessary an additional statistical model to derive the solution of kriging estimator.

In order to solve Eq. (2.3), first and second order statistics values of the variable $\mathcal{X}$ need to be specified. As stated above, this information is not available at any geographic location, and consequently, additional assumptions are needed. Typically, the variogram is the most commonly utilized model. Variogram is a two-point statistical model that provides dissimilarity measures between variables separated by a distance h. Mathematically, the variogram is defined by

$$\gamma(h) = \mathbb{E}\{[\mathcal{X}(r') - \mathcal{X}(r)]^2 \; : \; ||r' - r|| = h\} \tag{2.4}$$

where independence from r and r' is due to the ergodic hypothesis. Thus, the problem consists of estimating the variogram at any distance h. For this purpose, an experimental variogram is computed from the data using the method of moments estimator or utilizing training images. Then, this experimental variogram has to be modeled to provide statistics for distances not included in the database.

Despite the kriging methods provide a mathematically consistent framework, several practical problems have been identified in spatial estimation. Next, some drawbacks related to the

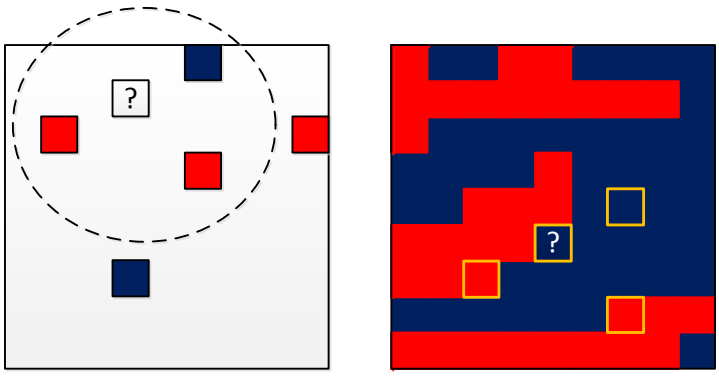
use of kriging method are discussed. In many applications, measurements are very limited, and as a result, estimating a meaningful variogram is not possible. The ergodic hypothesis might be extremely restrictive, and not valid in a large range of fields. Except for Gaussian models, the use of second order statistics is not sufficient to represent properly the underlying phenomenon. In other words, the studied phenomena may present higher order relevant moments than the variance and covariance, and consequently, variograms incur in a statistical mismodeling. On this line, two-point statistics models provide a poor representation of reality, and more complex structures might be used to represent subsurface structures of the earth. All these problems induce low credibility and poor realism in the solution provided by kriging, which is one of the main roadblocks that motivated the development of multiple-points simulation techniques.

2.2 Multiple-point Simulation

Multiple-point simulation (MPS) [2] has emerged as a new paradigm for geological modeling, due to the limitations provided by traditional approaches. Variogram-based methods present three main drawbacks. First, two-points second-order statistics are insufficient to characterize complex structures, and consequently, solutions provided by traditional methodologies present an important lack of physical realism. Secondly, estimating a meaningful variogram in critical under-sampling environments is infeasible, and as a result, the statistics promoted in the solution are not consistent with the actual field. Third, uncertainty and spatial variability are presented as key concepts in geostatistics. Conventional methods, such as kriging, provide unique solutions, and do not allow determining the uncertainty induce by the method or model. Variogram-based simulation is able to fix this issue, but still suffers of the first two problems. These issues motivated the development of MPS, where complex structures present in natural phenomena are characterized by high-order multiple-points statistics.

MPS techniques aim at reproducing and resembling geological reality. To attain this goal, MPS algorithms use training images (TIs), which symbolize a representation of the expected underlying phenomenon. Training images provide a statistical information source, from which MPS algorithms extract high-order conditional patterns statistics to reproduce them in the simulations. In other words, TIs contain expected patterns of the field of interest. This corresponds to the main contribution of MPS paradigm: the use of TIs to provide physical reality to the inference algorithm.

MPS philosophy departs from traditional approaches in several fundamental aspects. First, MPS is developed on a discrete domain (image), in contrast to variogram-based techniques. Secondly, kriging-based methods provide a unique solution, but MPS allows generate several plausible realization of the TI model. Finally, while conventional methods use first and second order two-point statistics, MPS extracts high-order conditional statistics from TIs using an arbitrary number of conditioning variables. The last characteristic provides MPS with the ability to replicate complex and realistic structures, which is the main problem identified in traditional methods.



(a) (b)

Figure 2.1: Illustration of pattern matching methodology of MPS. (a) Acquired/Simulated image and pattern selection, and (b) searching and matching of patterns in the TI.

Several MPS algorithms have been developed in recent years. Each algorithm presents different comparative advantages, and its use depends on the application, variables of interest, training images and computational considerations. Nevertheless, all these alternatives present a common fundamental set of stages that represent the foundations of the MPS framework. The first element is the simulation path. This stage corresponds to choosing a specific sequence of variables being simulated. The most commonly used approach is the random path. This approach consists of selecting randomly, typically using a uniform distribution, the location of the variable to be simulated. Thus, large-scale patterns will be simulated first, due to the under-sampling environments and the distances between each location of interest (hard and simulated data). This behaviour promotes desired features on the simulations in terms of the pattern reproduction at different scales. On one hand, small-scales patterns present in the TI should be reproduced as perfect as possible. On the other hand, large-scales patterns need to present a randomized reproduction, to provide variability to the simulations. Larger-scales features are simulated first, and consequently, they will present more variability than the subsequent simulations. The randomized reproduction is due to the random nature of the simulation path. To the degree that the simulation is performed, each simulated variable becomes hard datum, and as a result, the variability of small-scale features decreases.

A second common element of MPS algorithms is the estimation of the conditional probabilities and the simulation of regionalized variables. In a nutshell, the problem consists of estimating the conditional probabilities of each possible event (variable value) given a set of sampled and simulated data (template). To be more specific, let $\mathcal{X}(r_0)$ be the variable to be simulated at the location r_0. Note that r_0 represents a position within an image, in contrast to variogram-based methods, where r_0 represents a physical distance. Let $\{X(r_i)\}$, with $r_i \in \mathcal{R}_0$, be the measurements (or simulated variables) in a neighbourhood $\mathcal{R}_0$ of r_0. Then, the problem involves the estimation of p_c, given by:

$$p_c = \mathbb{P}\left\{\mathcal{X}(r_0) = c \,|\, \{X(r_i \in \mathcal{R}_0)\}\right\}, \tag{2.5}$$

where c represents a particular event or category. Thus, by computing p_c for each value

c, it is possible to obtain the conditional probability mass function of $\mathcal{X}$ at location r_0. Several approaches have been proposed to estimate p_c, but all of them consist of calculating conditional frequencies; i.e., the number of occurrences of a particular event in the TI. Fig. 2.1 illustrates the process to estimate the conditional frequencies. Finally, the obtained PMF is sampled and the simulated value is located at the position r_0.

2.2.1 SNESIM algorithm

SNESIM is a non-iterative algorithm for multiple-points geostatistics simulation of categorical variables [19]. SNESIM was motivated by the work of Guardiano and Srivastava [20], where a simple sequential simulation scheme was proposed. The problem presented in [20] was the multiple scans of the TI requires to the estimation of the PMF, and consequently, the MPS algorithm presents a high CPU cost. To solve this issue, Strebelle [19] proposed the use of a tree storage to represent all possible events given a particular template. This approach results in an extremely efficient event search process, which reduces the CPU costs with respect to the original algorithm; although it might requires high RAM demanding. For completeness, the pseudo-code of SNESIM is presented below. This algorithm was used in this thesis to generate the database in order to provide statistical information about the support of the models.

Algorithm 1: SNESIM pseudo-code [19]

Input: One categorical variable $\mathcal{X}$, simulation grid, categorical TI, conditioning data, local probabilities, parameters of the algorithm (minimum number of occurrences, proportion of categories, distribution updating parameters).
Output: Set of images with known variables at each location.
begin
    Incorporate conditioning data to the simulation grid.
    Define the template to be used.
    Construction of the searching tree from the TI.
    Discarding events that do not satisfy the minimum number of occurrences.
    Define the simulation path.
    while *there be unknown variables* **do**
        Defining a location using the simulation path.
        Defining the neighbourhood of the selected location.
        Estimating the conditional PMD of the variable using the search tree.
        Updating the local probability with the estimated PMD.
        Simulating the variable using the new distribution.
        Set the obtained value in the corresponding location.

2.3 Sparse Regularization

In signal processing, standard acquisition schemes are based on the classical Shannon-Nyquist sampling theorem [21, 22]. This theory stipulates that in order to preserve the information of the signal, the sampling rate must be at least two times the bandwidth of the signal. Unfortunately, most signals present high frequency components, and consequently, the sampling ratio tends to be large and unreachable in several cases. This results in many critical problems, such as: information loss, data storage requirements, data processing, high cost of measurements and devices design. These issues motivate the development of more complex strategies of sampling and raised the research of sparse representation in the context of signals recovery.

Sparse regularization framework is presented as a powerful alternative to the classical sampling scheme, providing lower sampling rates than the standard theory. As previously mentioned, the main idea is modeling the signal as an object that presents a simpler representation in some alternative transform domain. Then, the problem consist of finding the simplest signal, in this alternative domain, that is consistent with the samples. On this, Compressed Sensing (CS) [12, 23, 24] has contributed significantly in the sparse signal recovery framework, enabling large reduction of measurements and computational costs in several applications, including medical imaging, geological ill-posed inverse problems and interferometry [4,10,25]. In particular, the RIPless theory of CS [24] has provided several theoretical guarantees based on the complexity of the signal, in terms of its sparsity, and the coherence of the measurement scheme. This particular theory is fundamental in the context of geological inverse problems, where the sampling domain is fixed and highly restrictive [4].

To some extent, the differences between the CS paradigm and classical sampling approaches are threefold. First, CS is elaborated on a finite-dimensional signal reconstruction framework, in contrast to classical theory, which is focused on continuous-time (or spatial) signal recovery. Secondly, the acquisition processes differ in manner and meaning. As previously stated, the sampling theorem establishes a structured and exhaustive acquisition stage (pulses train) based on the bandwidth of the signal. On the contrary, randomness is one of the key concepts in the CS framework. Regarding the samples, while in standard sampling scheme each measurement represents the signal at a specific time (or location), in CS each sample is obtained as the inner product between a sampling basis and the original signal. Finally, the third difference is the synthesis equation; i.e., the procedure used to recover the signal from its samples. Sampling theorem recovers the original signal using the sinc interpolator, in contrast to sparse regularization techniques, which recover the signal solving a non-linear optimization problem.

2.3.1 Sparse and compressible representation

Sparse representation is an emergent research area in image recovery. Sparseness and compressibility refer to the signal ability to be represented with a small proportion of coeffi-

cients relative to its ambient dimension [12, 24, 26]. Generally, most signals are non-sparse in the canonical domain, but instead they present a sparse (or compact) decomposition in an appropriate space (or transform domain). This sparseness property is instrumental in many inverse problems frameworks, where additional assumptions are needed to solve underdetermined systems. In signal processing, there are emblematic examples of algorithm aimed at compressing the information given an acceptable distortion level, such as JPEG, JPEG2000 and MP3 standards.

To be more specific, let x be a n-dimensional vector and the linear decomposition of x in a basis or frame $\mathcal{U} = \{u_i\}_{i=1}^n$ given by

$$x = \sum_{i=1}^{n} z(i)u_i = Uz \in \mathbb{R}^n, \tag{2.6}$$

where $z = [z(1)\ z(2) \cdots z(n)]^T$ denotes the transform coefficients of x with respect to $\mathcal{U}$ and $U = [u_1\ u_2 \cdots u_n] \in \mathbb{R}^{n \times n}$ represents the collection of the atoms $\mathcal{U}$. If x can be perfectly represented by k elements of the basis $\mathcal{U}$; i.e., z has at most k nonzero entries, x is said to be k-sparse in the $\mathcal{U}$ domain.

Sparse signal models are widely used in several signals processing applications (signals reconstruction, speech recognition, features extraction, denoising, statistical learning theory, to name some examples). Compression is one of the classical issues where sparseness is instrumental. This discipline is aimed at representing a signal (such as video, audio or image) with the minimum information as possible given an acceptable distortion rate. To this end, compression algorithms exploit the redundancy of the signal in an appropriate space, shrinking to zero the irrelevant components. Emblematic examples in image compression are the JPEG standard [27], which uses the DCT basis to compress the image, and JPEG2000 [28], which uses Wavelet transform to generate sparse representations of images without losses. Additionally, promoting sparse or compressible structures results in some interesting and desired properties such as: sampling rate decreasing, computing resources reduction (storing data, model evaluation, optimization and matrices operation), simpler description of models and statistical robustness.

Unfortunately, most natural signals do not admit a pure sparse representation. Nevertheless, signals can be compressed with an appropriate transformation. Then, compressibility refers to the ability of signals to be well-approximated by sparse representations. In other words, if most of the signal energy is concentrated in few transform coefficients, then the signal is said to be compressible. Formally, x is said to be compressible in the domain $\mathcal{U}$ if there are constants $C_0 > 0$ and $q > 0$ such that

$$|z(n_i)| \leq C_0 i^{-q}, \tag{2.7}$$

where $x = Uz$ and $|z(n_{i-1})| > |z(n_i)|$. Then, as the magnitude of the transform coefficients rapidly decay, most of the energy is concentrated in few entries of z, and consequently, the energy of the approximation error decays almost at the same rate [29].

These notions raise the problem of finding the best sparse representation of a signal. This process is named *sparse approximation*, and it is one of the main foundations of CS theory.

Formally, let $(x, z, \mathcal{U})$ the vectors and the basis previously stated. It is defined the best k-terms approximation of x as

$$x_k = U\, z_k \tag{2.8}$$

where z_k is the vector that retains the k largest coefficients (in magnitude) of z and U is the transformation shown in Eq. (2.6).

On this context, an active research area in sparse representation is the development of overcomplete dictionaries, called frames, that can represent the signal with more coefficients than the original dimension. Frames can provide better representations than bases in several application due to their redundancy [30]. Redundant representations allow the extraction and separation of several features, such as texture, cartoon, edges, oscillatory patterns and noise; making frames more robust to error and information losses than bases. Nevertheless, this is not a dimension studied in this work, but it provides an attractive research area of future developments.

2.3.2 Basics of RIPless Compressed Sensing theory

CS is a well-elaborated theory to solve underdetermined linear systems based on sparseness and compressibility notions [12, 24]. CS has established that it is possible to recover the original signal perfectly, or obtain a small recovery error, imposing conditions over the acquisition process and the signal structure. This section presents the foundations of the RIPless theory of CS elaborated by Candés and Plan [24].

Regarding the sensing method, CS theory is elaborated over a randomized linear sensing process; i.e:

$$y = \tilde{A}x \in \mathbb{R}^m, \tag{2.9}$$

where $\tilde{A} \in \mathbb{R}^{m\times n}$, $m \ll n$, is a matrix generated by m realizations of a random vector $a(w)$, x is the signal of interest and y represents the measurements. Then, the problem consists of recovering x from the limited data y. All requirements for this sampling scheme are summarized in two properties: isotropy and coherence [24]. On the one hand, the isotropy property stipulates that each measurement has to be obtained independently; i.e,

$$\mathbb{E}\{a\, a^T\} = I_{n\times n}, \tag{2.10}$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation operator and $I_{n\times n}$ the $n \times n$ identity matrix. In other words, the elements (rows) of $\tilde{A}$ are uncorrelated and properly normalized to satisfy a unitary variance. As previously mentioned, randomness is one of the main differences between classical sampling strategies and CS paradigm. On the other hand, coherence represents a compromise between the sensing matrix $\tilde{A}$ and transform matrix U. This property quantifies the orthogonality between the sensing and transform spaces. The coherence is defined by

$$\mu(\tilde{A}, U) = \max_{(i,j)\in[0,n-1]^2} |\langle a_i, u_j\rangle|^2, \tag{2.11}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $\mathbb{R}^n$, a_i represents the transpose of the i-th row of $\tilde{A}$ and u_j the j-th column of U.

In order to provide an interpretation of the coherence (or orthogonality) between bases, consider the Fourier and canonical domains. Fourier basis is composed of pure complex sinusoids at frequencies $\frac{2\pi i}{n}$, $i = 0, \ldots, n-1$. In contrast, canonical basis is formed by deltas located at different time positions (or location in the context of image processing). It is possible to note that the projections of each atom of the canonical basis corresponds to the conjugate of the Fourier basis elements. In other words, the Fourier transform of a delta function is a pure complex sinusoid. Analogously to spikes in Fourier domain. Then, an element entirely concentrated in the canonical domain presents a dispersedly distributed representation in the Fourier domain. As a result, the coherence between both bases will be small. Indeed, the coherence between the Fourier and canonical bases takes the minimum possible value [24]. More generally, a pair of bases will be incoherent (small values of coherence) if the features that present a compressible representation in one of the basis are dispersedly distributed on the other.

Regarding the recovery algorithm, the main idea is to find the simplest (sparsest) vector z that explains the measurements. This idea justifies the use of the following optimization problem

$$\hat{z} = \underset{\tilde{z} \in \mathbb{R}^n}{\operatorname{argmin}} ||\tilde{z}||_0 \qquad s.t \;\; y = A\tilde{z}, \tag{2.12}$$

where $||z||_0$ is the number of nonzero entries of z. Unfortunately, this problem corresponds to a combinatorial optimization problem, and consequently is not tractable in practice.

A relaxed convex alternative for the problem in Eq. (2.12) is the ℓ_1-minimization in Eq. (2.13), which has shown outstanding performances in sparse signal recovery [12, 24, 31]. On this line, CS has established theoretical guarantees for the ℓ_1-minimization algorithm, which has motivated several new sensing-recovery schemes. In particular, the problem

$$\hat{z} = \underset{\tilde{z} \in \mathbb{R}^n}{\operatorname{argmin}} ||\tilde{z}||_1 \quad s.t \; y = A\tilde{z}, \tag{2.13}$$

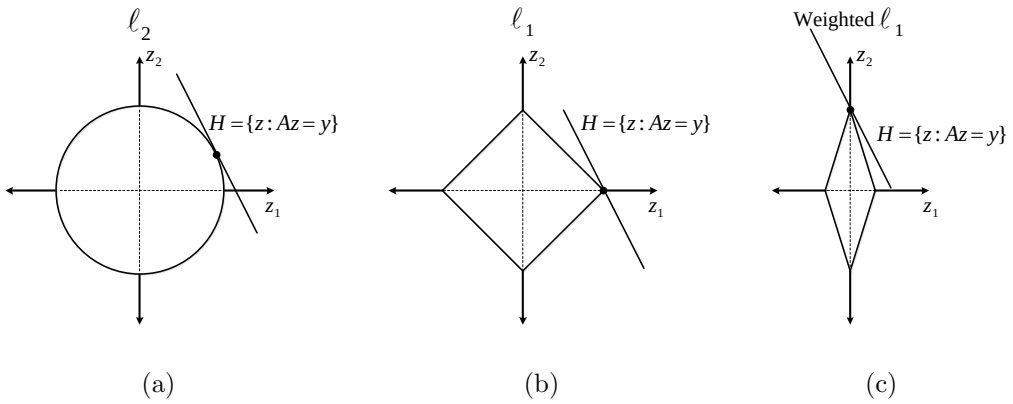
is termed basis pursuit (BP) [32] and it has shown good performances in several applications [4, 15, 33, 34]. In [23], Donoho demonstrates that under specific conditions, the solutions of Eq. (2.12) and Eq. (2.13) coincide, rising the development of several efficient algorithms. In fact, unlike the ℓ_0-minimizer, BP is a convex optimization, and consequently, can be posed as a linear programming problem [32].

The main result of CS theory states sufficient conditions under which the algorithm in Eq. (2.13) recovers the original signal x [24]. Formally, let x an n-dimensional signal, z the transform coefficients of x according to the basis U and A the sensing matrix in Eq. (2.13). Furthermore, let $\hat{z}$ the solution of the ℓ_1-minimization problem in Eq. (2.13). Then, if A satisfies the isotropy property in Eq. (2.10) and the number of measurements m accomplishes the following constraint

$$m \geq C_\beta \, \mu(A) \, k \, \log\left(\frac{n}{k}\right), \tag{2.14}$$

where n is the dimension of x, k the sparsity of z and $\mu(A)$ the coherence defined in Eq. (2.11); with probability $1 - 6/n - 6e^{-\beta}$, $\hat{z}$ satisfies that

$$||z - \hat{z}||_2^2 \leq \; C_1 \, \frac{||z - z_k||_1}{\sqrt{k}}, \tag{2.15}$$



(a) (b) (c)

Figure 2.2: Sparse regularization geometry. (a) ℓ_2-ball where sparsity is not promoted. (b) Standard ℓ_1-ball where no particular sparse structure is promoted. (c) Weighted ℓ_1-ball where sparse solutions with $z_1 = 0$ is promoted.

where z_k is the best k-term approximation of z and C_1 is a constant. If the basis $\mathcal{U}$ compresses the information (energy) of x in few transform coefficients, the right-hand-side (RHS) of Eq. (2.15) will be close to zero. In other words, if the transform domain $\mathcal{U}$ generates a representation of x that can be well-approximated by $k = \mathcal{O}\left(\frac{m}{\mu(U)\log n}\right)$ coefficients, the reconstruction error will be small. Indeed, if the signal is exactly k-sparse, the reconstruction is perfect ($z = z_k$).

2.3.3 Sparse regularization with support prior knowledge

The classical CS theory does not incorporate any prior information about the signal, except by the sparseness or compressibility hypothesis. Nevertheless, in most cases additional information is available, such as: prior information about the support structure, transform coefficient values, historical data or a prior statistical model. These information sources can be used to enhance the standard basis pursuit performance in Eq. 2.13, reducing the number of measurements needed to recover the original signal [17, 18, 35]. On this line, there are some theories and algorithms that incorporate additional structural information in the recovery stage, for instance: structural dependencies between elements within the support of the signal [36], block-sparse signal recovery [35], and Bayesian CS [37].

A special attention have to be done on the Bayesian approach, due to some misconceptions developed in the sparse recovery literature. The Bayesian approach has been presented as an interesting research branch due to the fact that, under some hypothesis, BP problem in Eq. (2.13) can be stated as a particular case of the *maximum a posteriori* (MAP) rule. Indeed, if the distribution of the observations is Normal and the prior distribution of the transform coefficients is Laplacian, then, the problems of BP and MAP are equivalent. From this bridge, several works have developed consistent mathematical approaches motivated by traditional Bayesian techniques. Nevertheless, Laplacian hypothesis incurs in two mismodeling errors. First, this approach considers the elements of the transform domain independent, which

generally is not true in image processing. Secondly, several articles have demonstrated that Laplacian distribution does not generate sparse realizations [38–41], and consequently, this hypothesis is not valid.

In the context of ill-posed problems in geoscience, geological fields present specific spatial statistics such as: spatial correlations, connectivity, orientation, continuity, complexity and piecewise structures; and consequently, a persistent sparse structure in the transform domain is expected across different realizations of the field. This knowledge can be incorporated in the inverse process, imposing a solution consistent with the expected geological structure.

Motivated by ℓ_1-regularization, an attractive alternative to incorporate prior information is penalizing irrelevant components in the transform domain that are not consistent with the prior model. This penalization can be included through a set of weights $\{w_i\}_{i=1}^n$ in the regularization term; i.e.,

$$\hat{z} = \underset{\tilde{z}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{n} w_i|\tilde{z}_i| = \underset{\tilde{z}\in\mathbb{R}^n}{\operatorname{argmin}} ||W\,\tilde{z}||_1 \quad s.t\ y = A\tilde{z}, \tag{2.16}$$

where $W = diag(w_1, w_2 \ldots, w_n)$, A is the sensing matrix and z is the transform coefficients of the original signal x. Then, the recovered signal is computed by $\hat{x} = U\hat{z}$. In general, w_i is inversely related with the expected magnitude of z_i [16]. Thus, components with small expected values (irrelevant coefficients) will present a high penalization term in Eq. (2.16), and consequently, the algorithm will shrink them to zero. Geometrically, this approach deforms the standard ℓ_1-ball privileging some particular sparse structures, as it is illustrated in Fig. 2.2. Then, the incorporation of the penalization matrix W induces a reduction of the searching space[1], discarding solutions that fit the measurements, are k-sparse but do not honour the prior model imposed by W.

Weighted ℓ_1-minimization is an active research area in signal processing [17, 42–45]. Empirical results has shown outstanding performance in sparse signals recovery, outperforming significantly the classical unweighted approach [17, 44]. On this line, the OSCAR algorithm [42] is one of the first works that explores the idea of incorporating weights in the sparse regularization term, in the context of variable selection and clustering. While OSCAR was not formulated as a weighted ℓ_1-minimizer, subsequent works have posed the algorithm as a sorted ℓ_1-minimization and atomic norm formulation [44]. This algorithm belongs to a new family of sparse regularizers termed ordered weighted ℓ_1 minimizers [43–45]. A more general framework is proposed in [43], which has extended the main ideas developed in [42]. On the theoretical side, there have been some contributions, but there are still lot of work ahead. Authors in [17] have proposed a method to compute an optimal set of weights when the entries of the signal fall into two possible sets, each one with different known probability of being nonzero. Recently, Figueiredo and Nowak [45] studied the use of ordered weighted ℓ_1 regularization for sparse estimation with correlated variables, which establishes some theoretical upper-bounds for estimation error, convergence conditions and proves sufficient conditions for clustering.

On the other hand, but strongly related with weighted ℓ_1-minimization framework, is the reweighted ℓ_1-regularization. In fact, reweighted ℓ_1 minimizers consist of solving a sequence

[1]In standard ℓ_1-minimization, the searching space has a cardinality $\binom{n}{k}$.

of weighted ℓ_1-problems, where the solution of one iteration is used to define the penalization matrix W for the next iteration. The advantage of the reweighted algorithms is that they define the matrix W adaptively and do not need prior information, as in the weighted approach. The algorithm presented by Candès, Wakin and Boyd [16] was among the first works done on this line, where a simple update criterion was proposed. From this novel idea, several theoretical analyses have been made to understand the role of the matrix W and the updating rule. In [46] it is shown that the error bounds associated to the reweighted algorithms can be smaller than the unweighted minimizer. In [47] it is proven that by assuming the restricted isometry property (RIP) [48] and the null space property (NSP) [49], the solution of the algorithm presented in [16] converges to a stationary point. In [50] the authors extend the idea developed in [16] to a unified framework of reweighted algorithms, proving some convergence features and studying several weights definitions.

2.3.4 Compressed sensing in geoscience

The fundamental idea of CS has been applied in several inverse problems in geoscience. The works of Jafarpour et al. [5,14,51] have explored the idea of sparse modeling for estimating spatially-distributed reservoir properties. In these works, the authors propose the used of basis pursuit [32], LASSO [52] and least square error in the context of history matching, studying the log-permeability recovery and oil saturation profiles reproduction. The authors use the DCT and Wavelets as sparse-promoting transforms and variation of the algorithms. One of the main contributions of these works is proposing a modification of the standard recovery algorithm in order to incorporate nonlinear dynamic flow measurements. Nevertheless, the analysis is focused on a particular single-channel structure and a fixed sampling ratio higher than 1% of data.

Calderón et al. [4] adopt some ideas from [14] and extend the results in several dimensions. First, the authors focus the analysis on a more complex multichannel model and study the performance of the method in a large range of sampling ratios. Second, they elaborate a formal connection between CS framework and practical aspects in spatial interpolation problem. From this, the authors identify the DCT domain as the basis that offers the best compromise between compressibility and coherence for spatial random samples. Finally, they propose a multi-scale averaging reconstruction and adaptive hard-thresholding in order to improve the standard sparse regularization performance. Nevertheless, clear limitations are identified below 2% of measurements, where artefacts such as discontinuities and non-channelized structures deteriorate the recovered field.

The ideas of sparse regularization have been also applied to seismic exploration [15,53]. In this context, the success of the inversion relies on the ability to solve large systems of equations. In consequence, "*the curse of the dimensionality*" is presented as the most significant roadblock, and the efforts have been focused on decreasing the number of source experiments (measurements) to solve the medium. On this, the randomized dimensionality reduction provided by CS is adopted in seismic exploration with the purpose of reducing computational costs. As sparsifying domain, the authors explored the use of curvelet, which compresses curved edges, and wave atoms, which provides sparse representation for oscillatory patterns,

both features typically present in seismic data.

This thesis adopts previous developed ideas based on sparse regularization in geoscience to extend the state of the art in the context of solving geological ill-posed channelized problem. In particular, this work explores the fusion of multiple-point information and weighted sparse regularization.

Chapter 3

Methodology

3.1 Problem Statement

In this thesis, the channelized facies recovery is addressed as an image recovery problem. The main hypothesis is based on the idea that geological images present a persistent compressible representation in a transform domain. In particular, channelized permeable fields have a compressible decomposition in some alternative domains, such as discrete cosine transform (DCT)[1] [4, 14]; Wavelet[2] [51, 54]; wave atoms and curvelet [15, 53], which are overcomplete dictionaries and provide sparser representations than several bases [55].

Formally, consider a regular n-grid blocks of a continuous channelized permeable field, which corresponds to a discretization of a binary permeability distribution map. These types of models are widely used in geological modeling, due to their ability to mimic geological reality and provide acceptable flow-forecasting values. Let $x \in \mathbb{R}^n$ the vector that represents the original discrete channelized field; i.e, each entry of x corresponds to the permeability indicator in a specific location of the grid (first column of Fig. 3.1). Then, the problem consists of recovering the original field from a critical number of pixel-based random measurements $y \in \mathbb{R}^m$, as is shown in the third column of Fig. 3.1.

In several applications, the sparsifying domain is a design variable relying only on the ability of the basis to compress the signal. Unfortunately, the spatial sensing scheme (pixel-based measurements) represents a strong constraint, and consequently, sparseness is not the only figure of merit to define the transform domain. The sampling scheme can be modeled considering the sensing matrix $\tilde{A}$ as the identity matrix $I_{n\times n}$ with removed rows. In [4], the authors study the trade-off between compressibility and coherence of different transform domains in the context of channelized two-facies inverse problem. They conclude that the DCT basis offers the best compromise between compressibility and coherence, providing the minimum sampling ratio needed to achieve an acceptable distortion level in the reconstruction

[1] DCT is used in the JPEG algorithm [27]

[2] A particular wavelet is used in the JPEG2000 algorithm [28]

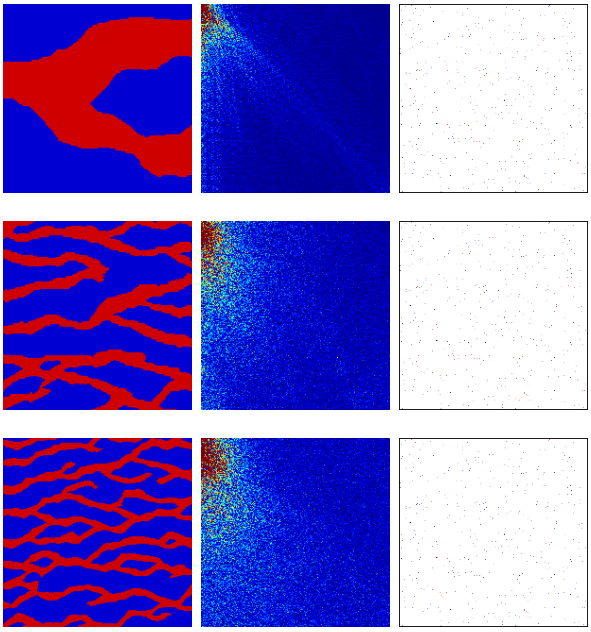


Figure 3.1: Example of different channelized models. From up to bottom by row: Singlechannel model, Multichannel model (multichannel 1) and a more complex multichannel structure (multichannel 2). From left to right by column: original discretized permeable map, DCT transform of the permeable field and measurements with 1% of sampling ratio.

of the fields. As a result, this thesis adopts the DCT basis as the sparse-promoting domain. Second column of Fig. 3.1 shows the transform coefficients $z \in \mathbb{R}^n$ for three channelized facies distributions.

3.2 Statistical support and weighting set definition

As indicated in Section 2.3.3, the analysis is focused on the family of weighted sparse promoting algorithms and the role of the penalization matrix W in Eq. (2.16). Thereby, the problem consists of how to use a geological prior model to identify relevant elements on the DCT domain. In other words, how to transform geological information in signal information. The hypothesis is that this additional knowledge might reduce the pursuit space and force a consistent support with geological structures. This prior information plays a key role in this context, due to the fact that the sampling ratio is far below the critical number of measurements stated by CS theory.

In this section, a statistical model of channelized facies fields is considered. The idea is to capture average patterns of the image in the transform representation (DCT domain) and to incorporate them within the sparse regularization term in Eq. (2.16). Formally, the image of interest is modeled as a random vector $X \in \mathbb{R}^n$ that follows a distribution γ. From this stochastic notion, it is possible to define the transform random vector $Z = U^\dagger X$. The hypothesis of this thesis is that common structures in the spatial domain (X) are reflected

Table 3.1: List of weighting matrix definitions implemented in this work. Γ_p has been introduced as the set that contains the p most significant coefficients of the statistical support.

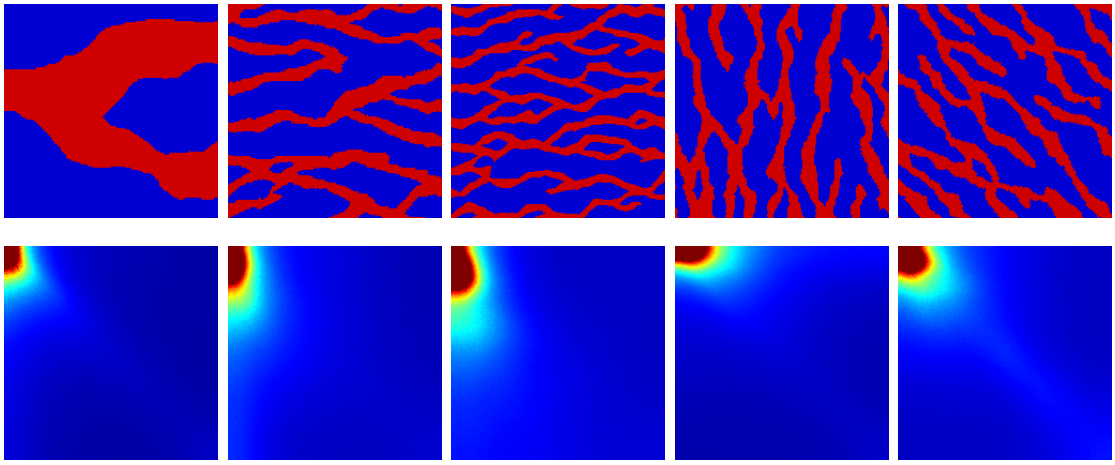
Name	Definition
Definition 1	$\frac{1}{\xi_i^{1-p}}, \quad 0 < p < 1$
Definition 2	$\frac{p + \xi_i^{1-p}}{\xi_i^{1-p}(\xi_i + \xi_i^p)}, \quad 0 < p < 1$
Definition 3	$\frac{p + \xi_i^{1-p}}{\xi_i^{1-p}(\xi_i + \xi_i^p)^{1-p}}, \quad 0 < p < 1$
Definition 4	$\frac{(1 + 2\xi_i)}{(\xi_i + \xi_i^2)^{1-p}}, \quad 0 < p < 1$
Definition 5	$\frac{1 + \xi_i^p}{\xi_i^{1+p}}, \quad 0 < p < 1$
Definition 6	$\begin{matrix} 1 & \text{if } \xi_i \in \Gamma_p \\ 0.001 & \text{otherwise} \end{matrix}, \quad 1 \leq p \leq n$
Definition 7	$\frac{1}{\mathbb{P}(z_i \in \Gamma_p)}, \quad 1 \leq p \leq n$

in a persistent sparse decomposition in the transform space (Z) across several realizations of the field. Motivated by recent weighted and reweighted sparse promoting algorithms, the magnitude of each coefficient of Z is used to generate a prior rank for atoms in the DCT basis. Thus, the average magnitude of the i-th transform coefficient is computed by

$$\xi_i = \mathbb{E}\{|Z_i|\} = \mathbb{E}_{X \sim \gamma}\{U_{\cdot i}^{\dagger} X\}, \ i \in \{1, \ldots, n\}. \tag{3.1}$$

The idea is to select $w_i = g(\xi_i)$, where $g : \mathbb{R} \to \mathbb{R}$ is a non-increasing function. As previously mentioned, several definitions for $g(\cdot)$ have been proposed in the reweighted ℓ_1-minimization framework. This work adopts some definitions from [50], and explores additional alternatives to take advantage of the database. Table 3.1 summarizes the weighting set definitions explored in this work.

In practice, it is not possible to access to the actual distribution γ. This thesis exploits the ability of MPS to generate realistic realizations of different geological structures, providing an empirical characterization of γ. MPS techniques use a training image to extract high order conditional statistics and simulate the unknown data. If there is no mismodeling, all simulated images honour the measurements and prior information (training image). Patterns and statistics present in the prior model are reproduced on the simulations, and consequently, these present common spatial structures, such as orientation and complexity in the case of channelized two-facies images.



(a) (b) (c) (d) (e)

Figure 3.2: Support trend for different models. Each column represents a different model. First row presents example of elements in the database. Second row presents the relevance indicator ξ of the respective model across one thousand images.

Formally, let $x^{(l)} \in \mathbb{R}^n$ be the l-th example of a particular channel structure database. Additionally, consider $z^{(l)} \in \mathbb{R}^n$ the DCT transform of $x^{(l)}$. Then, the estimator of the ξ_i associated with the atom $u_i \in \mathcal{U}$ is computed as

$$\hat{\xi}_i = \frac{1}{L} \sum_{l=1}^{L} |z_i^{(l)}|, \tag{3.2}$$

where $z_i^{(l)}$ is the i-th entry of $z^{(l)}$ and L the number of examples in the database. Is we assume that simulated samples $\{x^{(i)}\}$ are independently and identically distributed realizations of the random vector X, the law of large numbers stipulates that $\hat{\xi}$ converges to ξ as L tends to infinity. In particular, one hundred images are simulated for each model in order to provide statistical fidelity to the estimated support. Thus, the multiple-point statistical information can be transformed into signal information by analysing the trend of the transform coefficients across different realizations of the image. As a result, it is possible to incorporate a geological prior model in the recovery algorithm of Eq. (2.16).

Fig. 3.2 illustrates the average statistical support in the DCT space for different models. It is clear that most of the energy is concentrated at low frequency components (upper left corner of the image) due to the piecewise nature of the studied models. While high frequencies describe facies transitions, most of the channel structure is defined by low frequencies information. Therefore, in contrast to high-frequency component, low-frequency elements present a high probability of being significant coefficients and non-zero.

In addition, this statistical support provides some lights about the orientation and complexity of the field. If the channel structure presents an horizontal orientation, facies transitions occur in vertical direction, and consequently, vertical frequencies appear as relevant elements of the signal support (see Fig. 3.2.a-c). Analogously, the behavior of vertical structures is shown in Fig. 3.2.d. In the case of fields that have diagonal orientations, the support has no preferential directions and tends to be more symmetric (see Fig. 3.2.e). On the other

hand, channel complexity is directly related with the number of facies transitions. The higher the facies transitions number, the larger the high frequency components in the signal support. Thus, more complex structures, in terms of geological features, present higher sparsity levels. This explain why the spectrum illustrated in Fig. 3.2.c contains more high frequency components than the support in Fig. 3.2.a-b.

3.3 Experimental setting

The goal of this section is to state the experimental setting and analysis scenarios developed in this work. Regarding the sensing modality, samples are taken randomly with a uniform distribution and the analysis is focused on the range of $0.1 - 2\%$ of measurements. The DCT basis is used to provide sparse representations of the channelized two-facies images. The use of random sampling and DCT basis, are necessary to satisfy some theoretical conditions in sparse signal recovery framework, as was stated in the context of ill-posed geological problem in [4]. In addition, three different model complexity scenarios are studied in order to provide a more complete analysis of the proposed method.

Regarding the solver, most algorithms do not incorporate the weighting matrix W and only solve the standard ℓ_1-minimization problem in Eq. (2.13). In order to provide prior information to the algorithm, the incorporation of W is accomplished through the modification of the sensing matrix A in Eq. (2.13), with the purpose of using standard optimization codes in the literature. Specifically, by defining $\alpha = Wz$, the problem in Eq. (2.16) can be posed as

$$\hat{\alpha} = \underset{\tilde{\alpha}\in\mathbb{R}^n}{\operatorname{argmin}} ||\alpha||_1 \quad s.t \;\; y = AW^{-1}\alpha. \tag{3.3}$$

Then, the recovered field is computed by $\hat{x} = U\,W^{-1}\hat{\alpha}$. Thus, weighted ℓ_1-minimization can be posed as a classical ℓ_1-regularization instance [16].

As previously stated, multipoint simulation algorithm SNESIM is used to generate realizations of different channelized two-facies models. Particularly, one thousand images of 200×200 pixels are created for each model considered in this work. Examples of the database are presented in first row of Fig. 3.2. The database is used to define the relevance indicator ξ described in Eq. (3.2) and estimate the support of the actual field. This index is used to create the weighting matrix W and incorporate the model information in the solver of Eq. (2.16).

To begin, the variability of the method as a function of the weighting matrix W is analyzed. From this analysis, the best definition for the penalization term depending on the model complexity and sampling ratio is provided. In a second study, the role of the model complexity is analyzed by comparing the performance of the method for the three models. This comparison relates model complexity, in terms of geological features, with the performance of the recovery method, based on sparse notions. Third, a multi-scale averaging reconstruction [4] using the weighted ℓ_1-minimizer is adopted. In a nutshell, this approach consists of decomposing the image in non-overlapping blocks and applying the recovery algorithm inde-

pendently in each sub-image. The size of the sub-blocks is called scale decomposition. This procedure is performed across different block-sizes (scales), and consequently, a collection of partial recovered images are obtained. Then, the recovered field is computed averaging the independent block-size reconstructions. In addition, a two-level hard thresholding as a final post-processing stage is analyzed, due to the binary nature of the studied models. As in [4], the impact of scale decomposition is studied to illustrate the performance boost of the multi-scale reconstruction. Finally, a comparison with other signals processing and multiple-points simulation methods is performed.

The performances are presented in terms of signal-to-noise ratio (SNR) and similarity structural indexes [56] (SSIM), which is a perceptual indicator. Additionally, main results are presented over horizontal structures, duo to the orienteering-invariant of the sparse recovery theory in this context.

Chapter 4

Experimental results

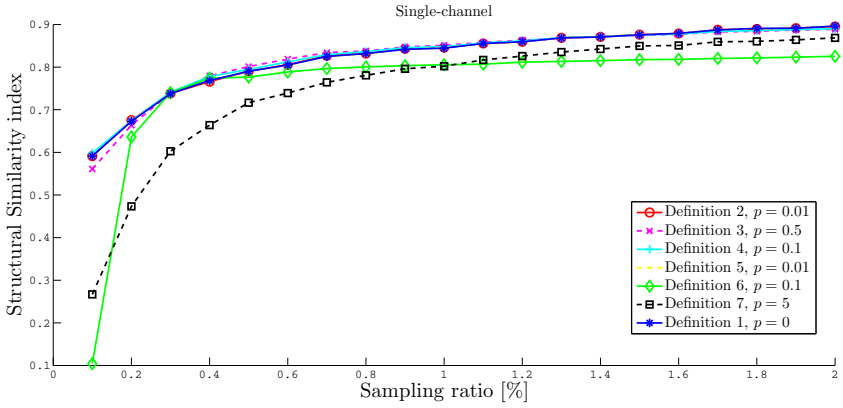
4.1 Weighting matrix variability

Selecting a proper set of weights is fundamental for several algorithm features, such as: convergence, time convergence and performance. Therefore, it is necessary to study the performance of different penalization matrices W in order to identify the best weights definition in this particular context. Some definitions from [50] are adopted and additional alternatives are considered.

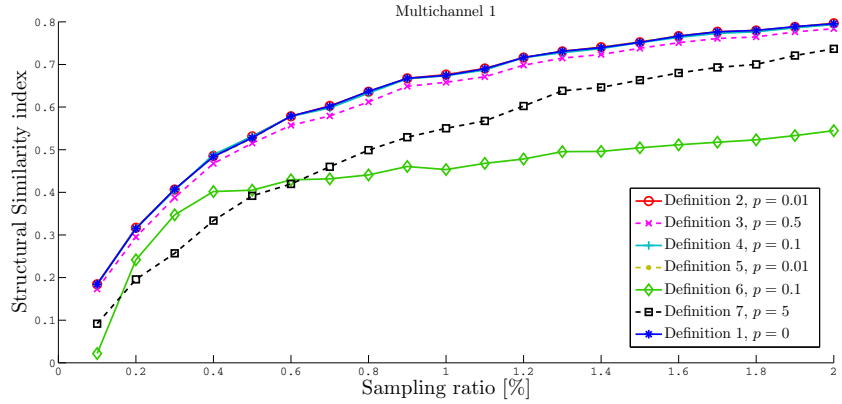
Regarding the experimental setting, the solver in Eq. (3.3) is applied for several definitions of matrix W. Table 3.1 summarizes the weighting matrices studied in this thesis. Each weighting set is properly normalized to a unitary minimum weight in order to provide stability to the algorithm. Concerning the definition 6, it has been introduced Γ_p as the set of variables that contains the p most significant coefficients of the statistical support. Additionally, in definition 7 the weighting set is computed using the functional $\mathbb{P}(z_i \in \Gamma_p)$, which represents the probability of $z_i \in \Gamma_p$. This probability is estimated from the database using a frequentist approach.

Fig. 4.1 shows the results obtained for different W matrices. It is important to mention that the results presented correspond to the best value of parameter p for each particular definition. As illustrated, there is no dominant definition for matrix W. While there are some alternatives that present poor performances, most studied cases provide similar results. This is due to the fact that most weights present similar power law decay curves and belong to a unified family of algorithms [50].

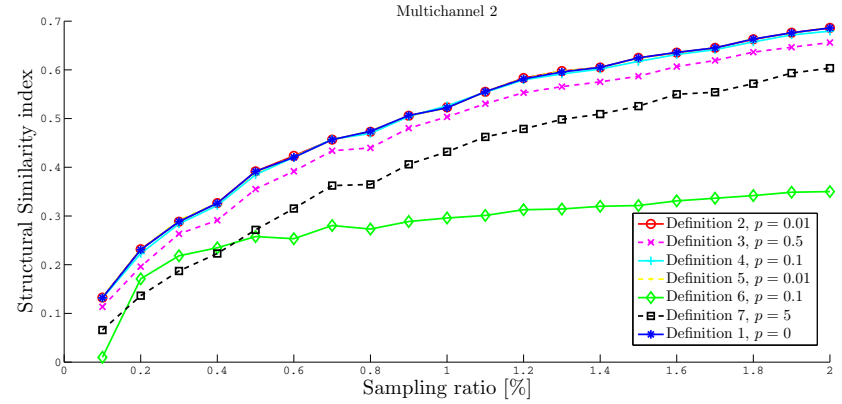
An interesting result is that the behaviour of different weighting matrices is persistent across the models. In other words, the comparative results between different definitions is independent of the model complexity. Weighting sets that offer poorer or better performance in a particular model present identical behaviour on the others. In consequence, the definition that provides better performance is the same for the three structures. Finally, the definition 1 with $p = 0$ provides the best recovery performances and is the one used to run the rest of



(a)



(b)



(c)

Figure 4.1: Variability of the method as a function of the weighting matrix W. Results are presented in terms of SSIM index. (a) Performance for Singlechannel, (b) performance for multichannel 1 and (c) performance for multichannel 2 model. Recovery images are performed from 25 realizations of the sensing matrix A. Each curve represents a particular definition, and is performed using the best parameter p.

the experiments.

4.2 Analysis of model complexity

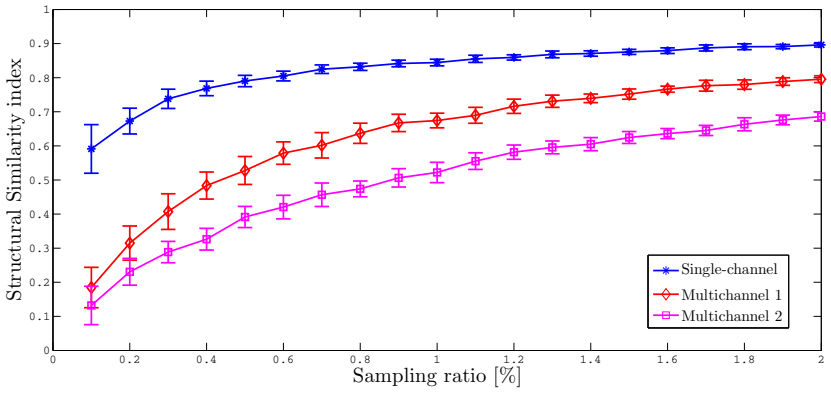
As stated in Section 3.3, it is important to analize the robustness of the method to the model complexity. To this end, three channelized structures −which might represent different scale levels of the same phenomenon− are analyzed in the range of $0.1-2\%$ of measurements. The weighting matrix W is performed using the definitions 1 with $p = 0$, which has shown better recovery performance as previously illustrated in Fig. 4.1.

First, as the sparsity is one of the most important concept in this work, it is important to analize the variables that impact the sparseness of the model in order to understand the results shown in Fig. 4.2. The complexity of the binary permeability map is directly related to the width of channels, which defines the number of facies transitions given a fixed image size. As previously stated, facies transitions define the sparsity of the image, and consequently, a more complex model, or higher resolution, in term of geological structure leads to a high sparsity level. On this, CS predicts that the smaller the sparsity, or complexity in the context of geological structures, the fewer the measurements needed to reconstruct the signal. This is not necessarily true in the context of patterns reproduction, where more complex models are richer in terms of the structures and provide a better scenario to statistics estimation of the model.

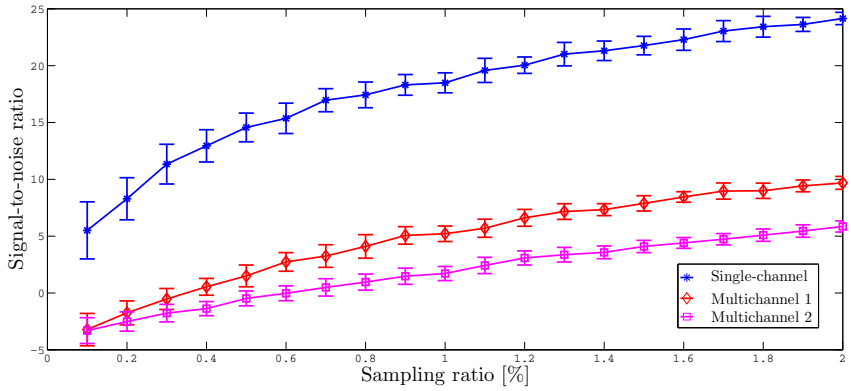
Fig. 4.2 shows the performance of the method for the three models in Fig. 3.2.a-c in terms of SNR and SSIM. As was stated, curves are ordered according to its sparseness in terms of SNR. The single-channel model presents sparser representations than the multichannel structures (see Fig. 3.2) due to the lower number of facies transitions, and consequently, it is possible to recover the field with fewer samples than the multichannel models. Additionally, this behaviour is enhanced by the weighting set definition. The support promoted by W in single-channel models presents a more compact distribution than the multichannel model case, and as a result, sparser decompositions are promoted in single channel solutions. Thus, better performances in single-channel models can be explained by the nature of the image itself (sparsity) and the penalization matrix W (statistical support). Analogous the analysis between the multichannel 1 and multichannel 2.

Regarding the SSIM index, results depend on the scale at which the comparison is performed. SSIM extracts only first and second order statistics to compare both images, pretending emulate the human perception, which analyzes (compares) two images in term of their structures at different scale levels. If images are compared very locally, SSIM extracts more structural information in more complex models, and consequently, multichannel structures might present better recovery performances in some cases. To mitigate this problem, the SSIM index is performed across several scales, which range from 4% to 100% of the image. Thus, SSIM index incorporates both local and global information, making the index more robust.

The problem of comparing two images in terms of their structures is not inherent to patterns reproduction problem. To illustrate this behaviour, consider the stationary nature of the field studied in this thesis. The stationarity hypothesis is used by multiple-points simulation to extract conditional statistics. Multichannel models present stationary structures, and



(a)



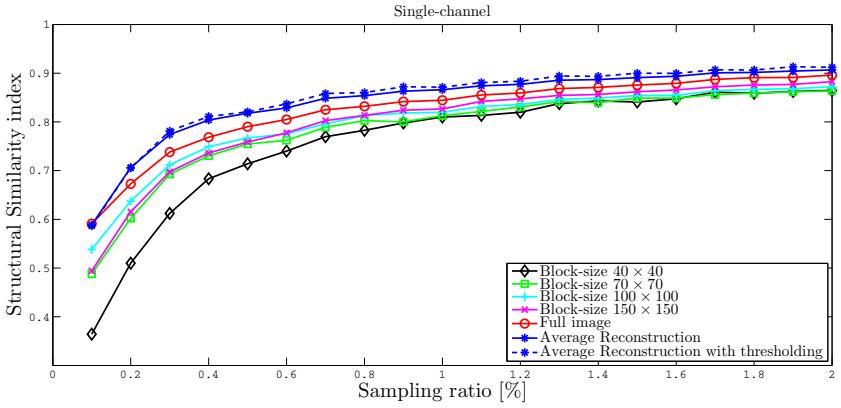
(b)

Figure 4.2: Analysis of the method as a function of the model complexity. Performance is obtained from 25 independently realizations of the sensing matrix. Results are presented in terms of (a) SNR and (b) SSIM index. Each curve represent a particular model: single-channel (blue line), multichannel 1 (red line) and multichannel 2 (magenta line).

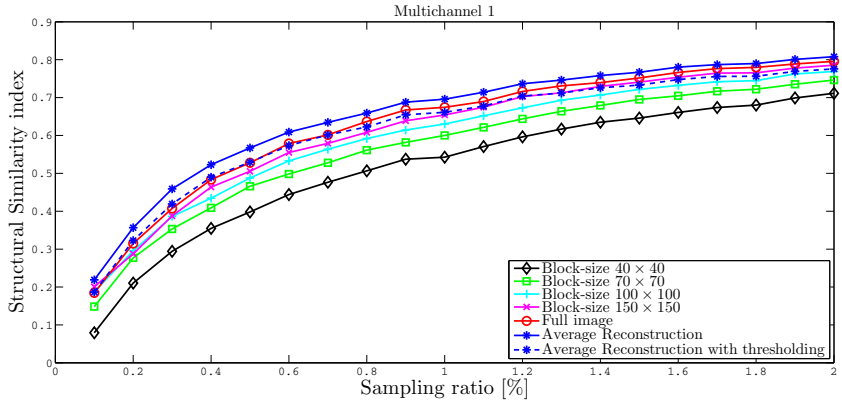
consequently, statistics can be estimated in any location of the image. On the other hand, it is clear that single-channel is not a stationary phenomenon. Then, there is a mismodeling in the estimation process, and as a result, patterns are not necessary reproduced in the simulations. Nevertheless, MPS has shown good performance reproducing this type of models due to their simplicity. Therefore, single-channel and multichannel problems might be addressed differently depending on the figure of merit. Moreover, in several cases both structures are analyzed with different strategies due to their distinct statistical features. In consequence, this motivates the use of different approaches as post-processing stage depending on the particular model.

4.3 Scale variability and thresholding stage

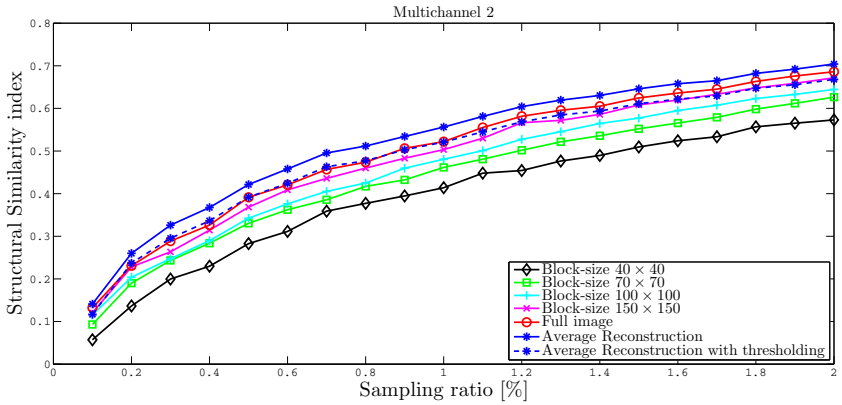
In order to outperform the standard weighted ℓ_1-minimization, this thesis extends main



(a)



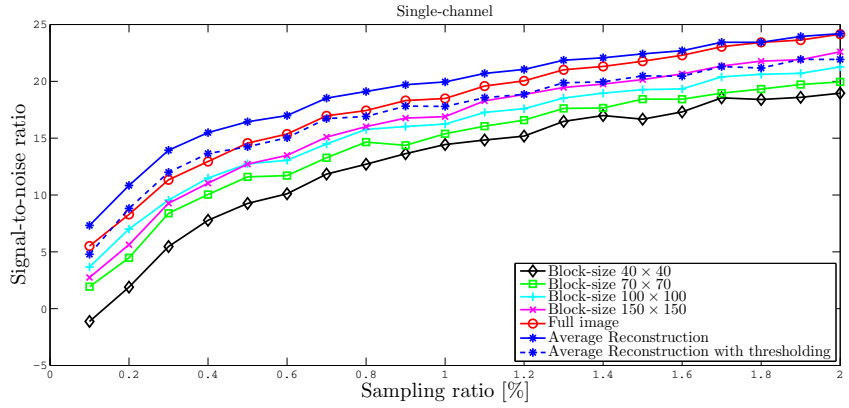
(b)



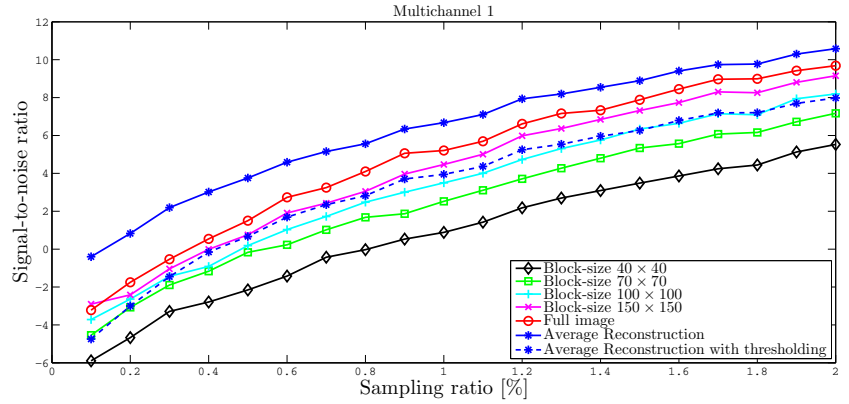
(c)

Figure 4.3: Variability of the method as a function of the scale decomposition. Results are presented in terms of SSIM index. Each graphics contains the performance curves of different block-sizes, the reconstruction with full image and the average reconstruction with and without thresholding. (a) Performance for Singlechannel, (b) performance for multichannel 1 and (c) performance for multichannel 2 model. Recovery images are performed from 25 realization of the sensing matrix A.

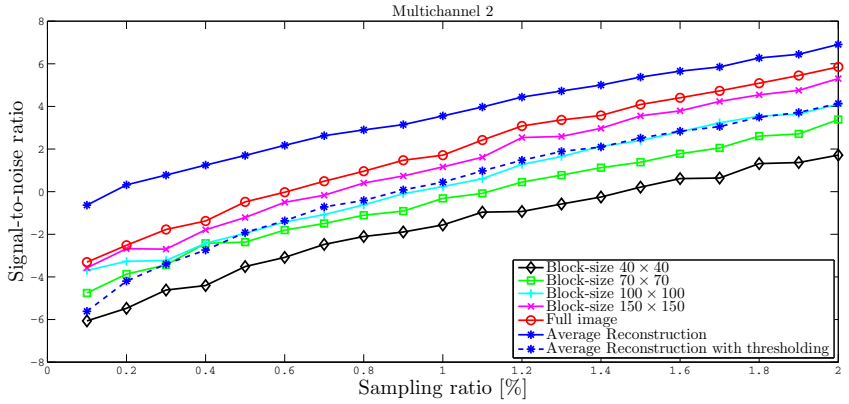
ideas from [4] adopting a multiscale averaging reconstruction scheme. Then, it is analized the role of decomposing the problem at different scales (block-sizes) as well as an adaptive



(a)



(b)



(c)

Figure 4.4: Same caption as Fig. 4.3 but using SNR index.

hard-thresholding as final post-processing stage. For more details, [4] elaborates a complete analysis of this method for a particular multichannel structure.

Fig. 4.3 and 4.4 show the performance of the multiscale method in terms of SSIM and SNR, respectively, for different decomposition scales. Additionally, each graphic contains the average reconstruction with and without thresholding. Results show that the smaller the

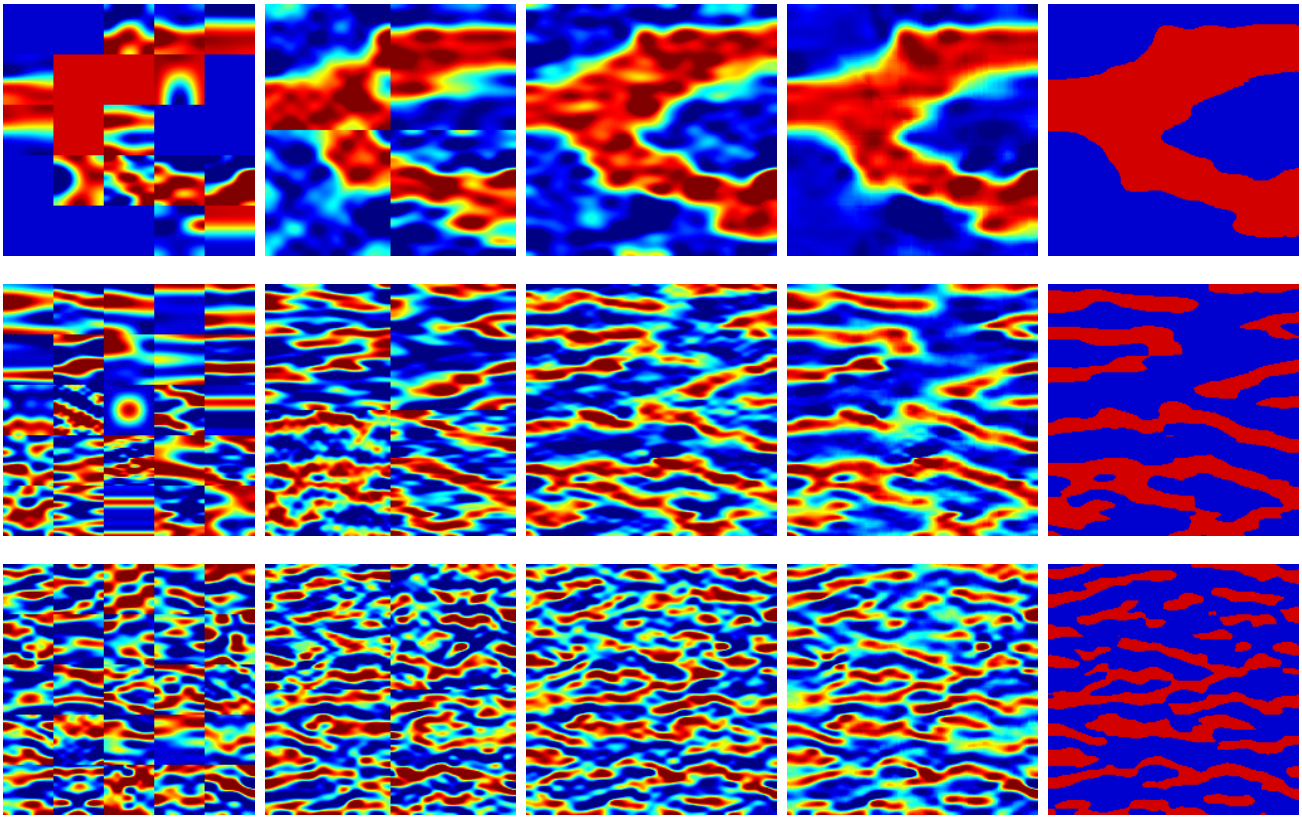


Figure 4.5: Illustrations of the variability of the method as a function of block-size. From left to right by column: $40 \times 40pp$, $100 \times 100pp$, full image, average recovered image and average reconstruction after thresholding. From up to bottom by row: single-channel recovered fields from 0.4% of measurements, multichannel 1 reconstructions from 0.7% of measurements and multichannel 2 recovered images from 1% of measurements.

decomposition scale (larger blocks), the better the quality of the recovered field. This trend is consistent with the original work [4], where this behaviour is explained by the reduction of the sparsity as the block-size increases [4, Section 4]. Concerning the averaging process, the integration of several recovered images at different scales provides a performance boost with respect to the standard weighted ℓ_1-minimizer. This gain can be attributed to the fact that persistent features across different scales prevail, for instance orientation, complexity and dominant channel structures. Thus, these common features are accentuated while the artefacts of the particular recovered fields are mitigated (high/low permeable zones without channelized structures and distortions induced by the decomposition of the image).

Regarding the hard-thresholding stage, it is not possible to provide a unified conclusion respect to the three models. As previously stated, a separated analysis is needed for single-channel and multichannel models, which is consistent with the standard in the literature. In the case of single-channel structures, the hard-thresholding process leads to an outstanding increase in the quality of the recovered images, unlike the case of multichannel fields. This behaviour is attributed to the hard-thresholding procedure and the nature of the models itself. The hard-thresholding honours the proportion of low and high permeability values in the samples, and does not impose any channel features, such as connectivity, orientation

and complexity. Therefore, in the multichannel case, the continuous and interconnected structure is not preserved after the hard-thresholding process, and consequently, the quality of the image is deteriorated. SSIM index compares three features in the images: luminosity, contrast and structures [57]. While the luminosity presents an enhancement after the hard-thresholding, the factors associated to the contrast and the structure are deteriorated, and as a results, the quality in terms of the SSIM index diminishes as a whole.

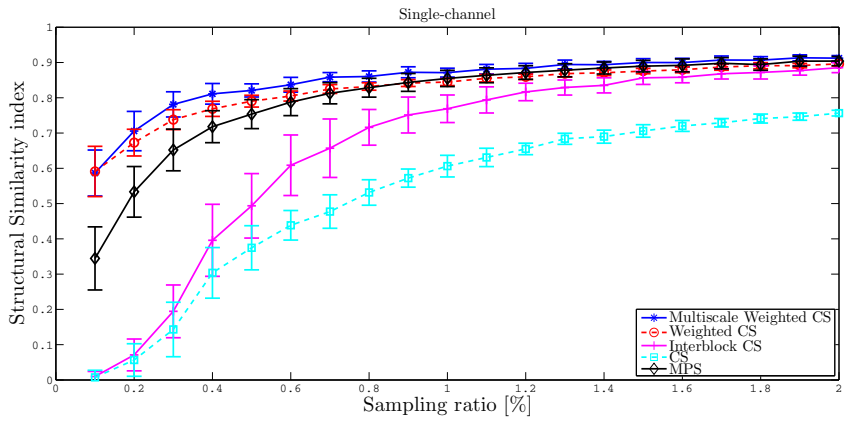
On the other hand, single-channel structures is a simpler model, hence, it presents fewer facies transitions and details. The weighted CS recovery algorithm is able to estimate the dominant channel structure (low frequency information), due to the distribution of the energy in the weighting matrix W (see Fig. 3.2). This constitutes an important fraction of the recovered field in the case of single-channel structures. Therefore, by applying the hard-thresholding, the interconnected structure obtained by the algorithm is preserved and, as a result, the quality of the recovered image increases due to the binary nature of single-channel models. Fig. 4.5 shows some examples of recovered fields in order to illustrate the impact of the decomposition scale and hard-thresholding for the three channelized models studied in this work.

4.4 Multiscale averaging reconstruction and comparison with other methods

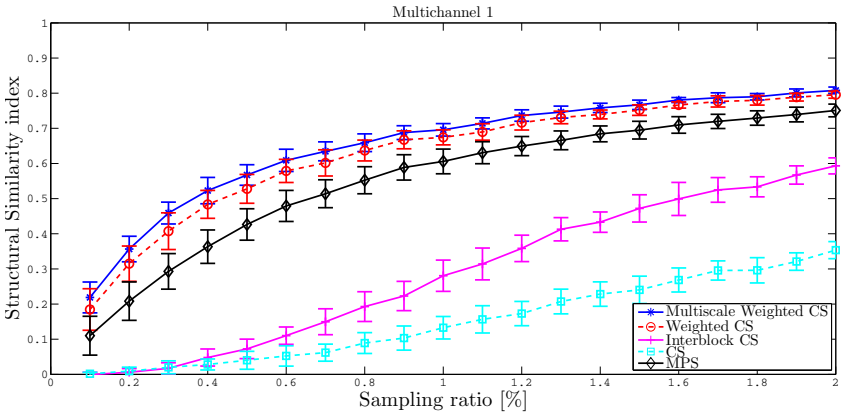
In order to compare the performance of the weighted multi-scale reconstruction method proposed in this thesis, additional interpolation methods are explored, such as: standard weighted ℓ_1-minimizer (WBP) and basis pursuit (BP) using the DCT basis; inter-block CS method developed in [4] (IBCS); and multi-point algorithm SNESIM [58]. Despite that MPS was not elaborated to reconstruct the actual field, it has been incorporated in this analysis to illustrate that the proposed method can also achieve good performances even in the more challenging problem of recovering the underlying reality.

Fig. 4.6 and 4.7 show the performance of all implemented methods for the three models analized in this work in term of SNR and SSIM. Each figure contain the performance with (solid line) and without (dashed line) thresholding. It is remarkable the ability of the method to recover the single channel model, achieving a good performance even in the range below 0.5% of measurements (see Fig. 4.8). Single channel images, being a simple model, can be well estimated by only low frequency information, where the support imposed by W is concentrated. On the other hand, the recovery method for multichannel structures shows clear limitations in the range below 0.7% of samples for the case of the multichannel 1 and 1% for multichannel 2. In these multichannel models, high frequency components are relevant in order to solve the facies transitions into the image. Unfortunately, the algorithm shrink to zero high frequency coefficients and, as a result, the recovered multichannel structures present several artefacts in channel edges, which deteriorate the quality of the solution.

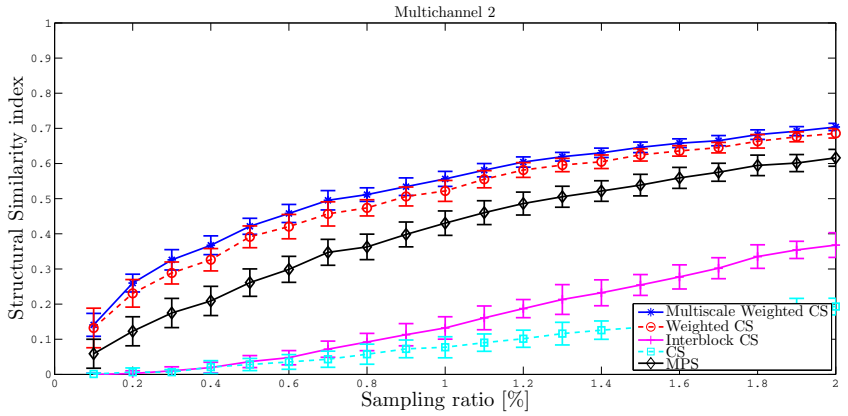
Concerning the comparative results, the proposed weighted CS-based scheme outperforms all other interpolation techniques in a large range of sampling rates. Remarkably, the incorpo-



(a)



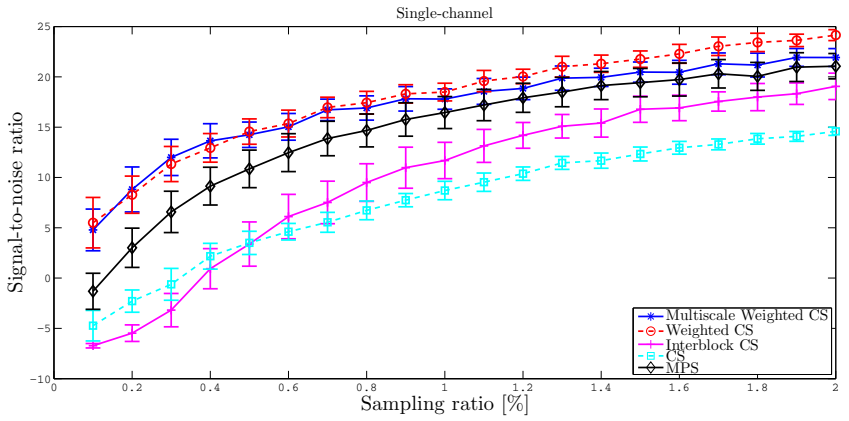
(b)



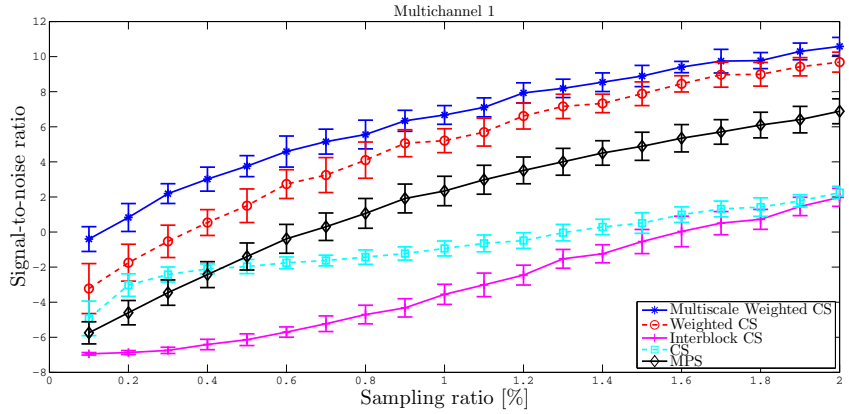
(c)

Figure 4.6: Average performance and standard deviation of the proposed weighted multi-scale CS method, standard weighted ℓ_1-minimization, inter-block CS method, standard ℓ_1-regularization and multiple-points simulation. Results are presented in terms of SSIM index and performed across 25 realization of sensing matrix A. (a) Single-channel, (b) multichannel 1 and (c) multichannel 2 model.

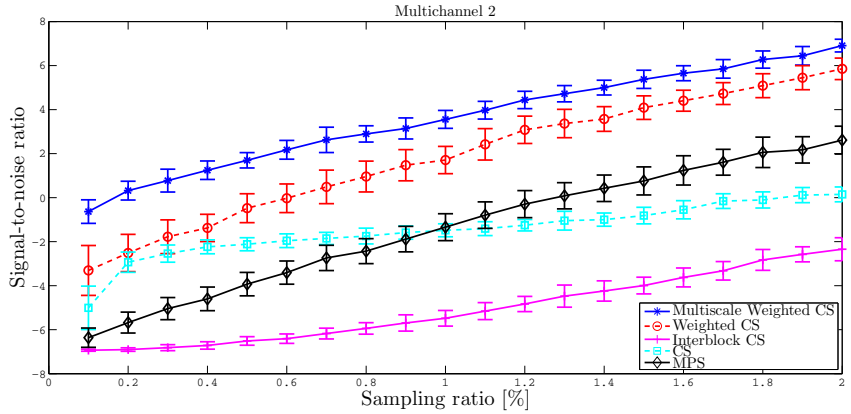
ration of the weighting penalization matrix W shows outstanding improvements with respect to the standard unweighted ℓ_1 minimization [4, 14], allowing the recovery of acceptable fields



(a)



(b)



(c)

Figure 4.7: Same caption as Fig. 4.6 but using SNR index.

even in the range below 1% of the data, especially in the single-channel model. This is clearly illustrated in Figs. 4.8, 4.9 and 4.10, where concrete examples of reconstructed fields are shown. Additionally, note that the variability of the method is reduced too, due to the particular structure forced by the weighting matrix. Thus, MPS proves to be an excellent source of information to estimate the support of the actual field, and consequently, it can be used to define an appropriate weighting matrix W.

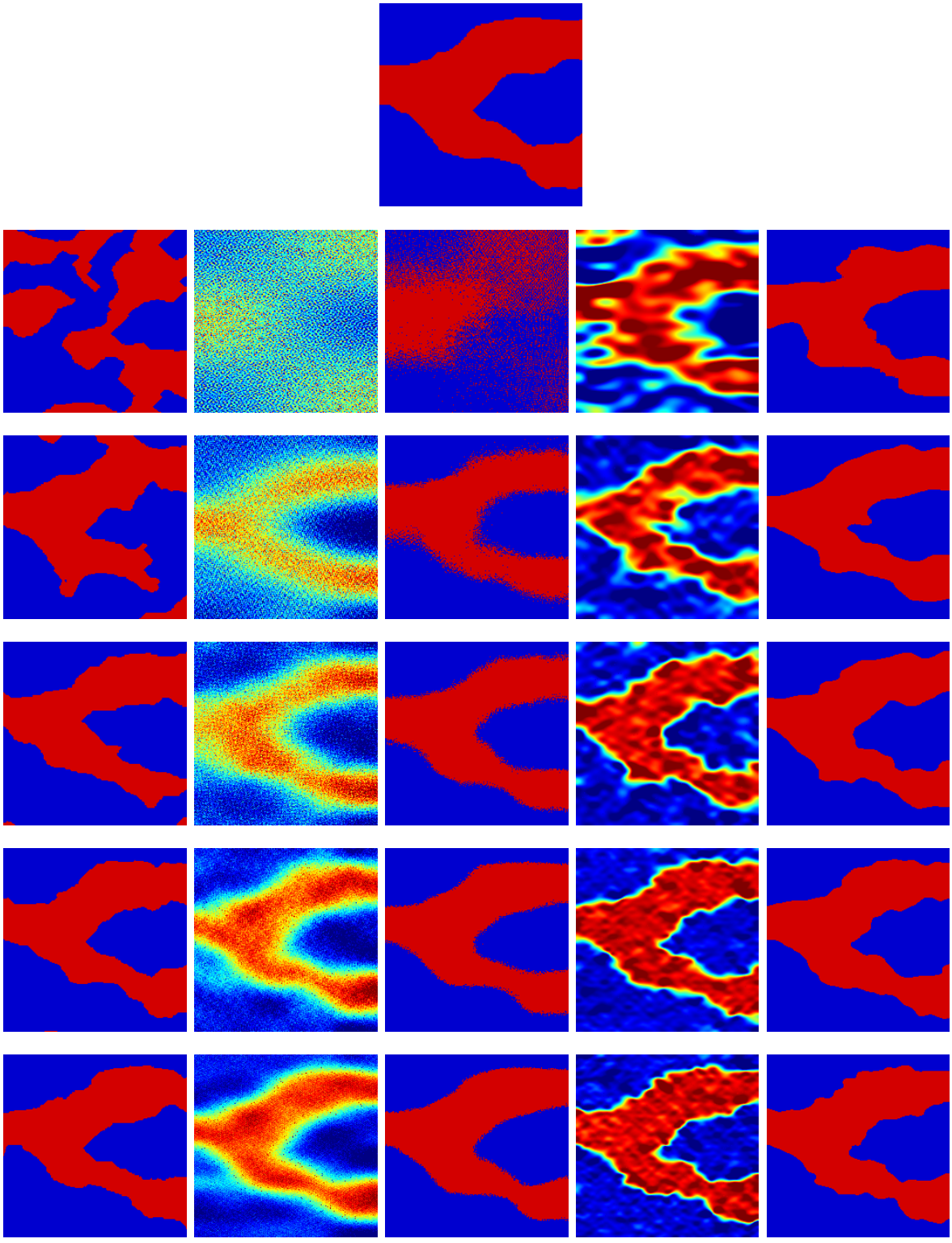


Figure 4.8: Comparative illustration with different sampling rates. Each column represent a particular recovery method and each row a sampling ratio. From left to right by column: multiple-points simulation, unweighted ℓ_1-minimization, unweighted inter-block CS method [4], weighted ℓ_1 regularization and weighted multiscale reconstruction method. From up to bottom by row: Original image, reconstruction from 0.2%, 0.5%, 1%, 1.5% and 2% of measurements.

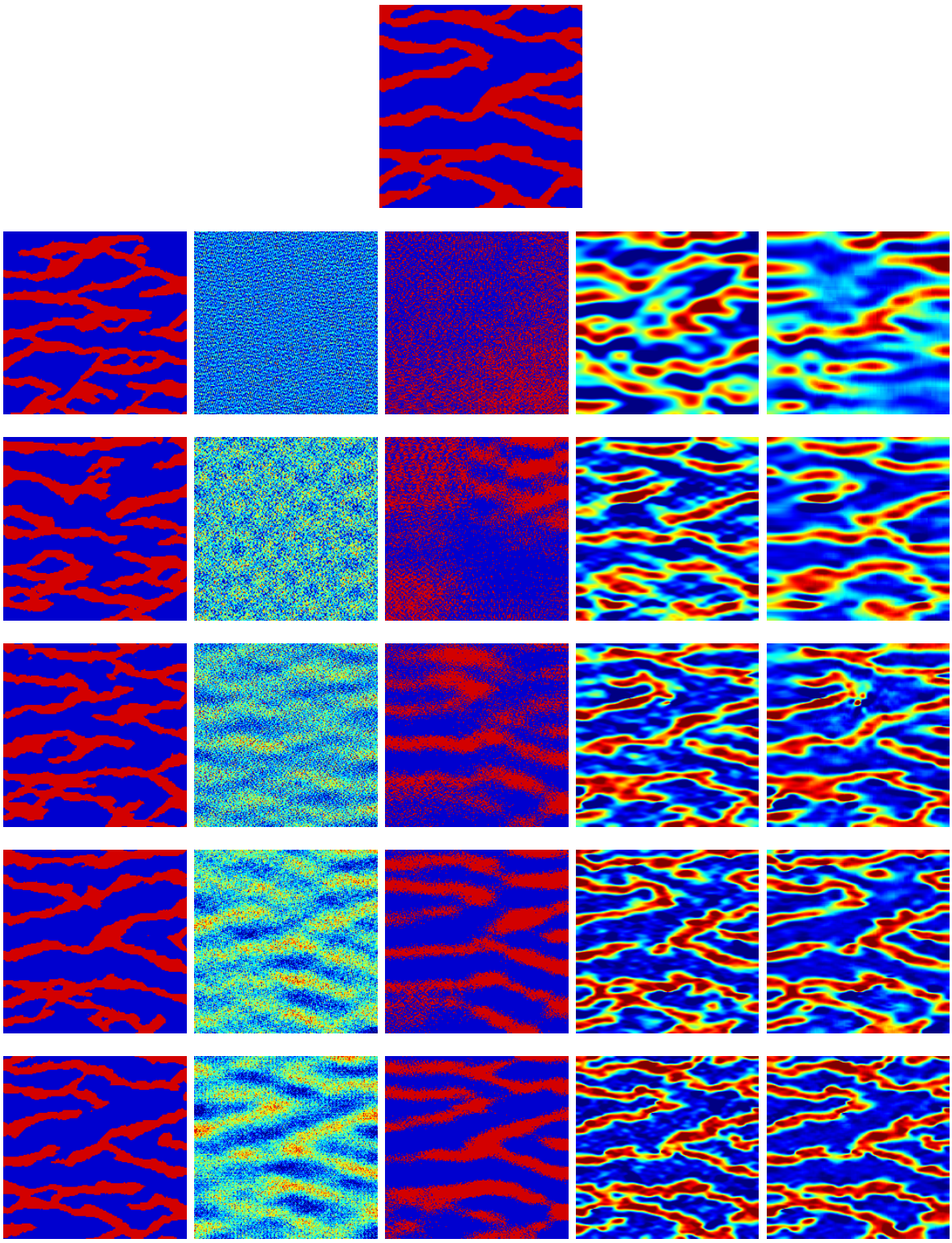


Figure 4.9: Same caption as Fig. 4.8 but using Multichannel 1.

Regarding the impact of the average multiscale approach, it is clear that it outperforms the standard full-image reconstruction. As previously stated, several artefacts are mitigated for this technique due to the information of partial recovered images at different scales. Nevertheless, the enhancement is less dramatic than in the unweighted case [4]. This confirms

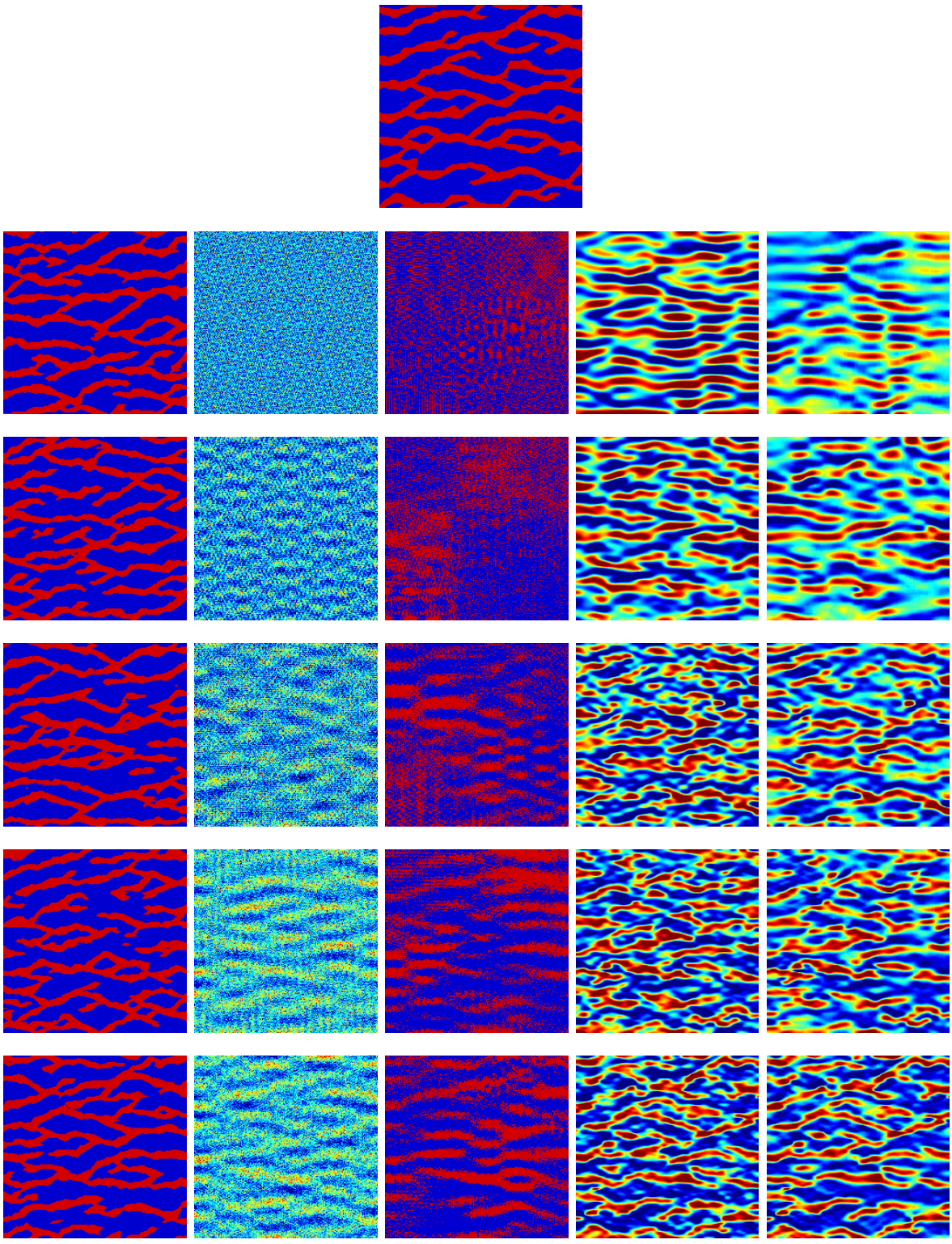


Figure 4.10: Same caption as Fig. 4.8 but using Multichannel 2.

the relevance of the MPS in the estimation of the support and the construction of the weighting matrix. In these critical sub-sampling environments, sampling rates are far below the boundaries stipulated by CS theory, and additional strategies are necessary to achieve an acceptable distortion level. In the previous unweighted CS scheme [4], the channel structure

is accentuated by the averaging process. In the proposed weighted CS-based method, incorporating support prior information allows the identification of dominant channel features without any post-processing stage. Then, averaging partial reconstructions removes some artefacts, but does not increase significantly the quality of the recovered image. In summary, main improvements are attributed to the support prior knowledge, instead of the average multiscale approach, ratifying the relevance of MPS as a key alternative information source.

A more fundamental result is that the method proposed in this thesis outperforms the MPS algorithm SNESIM, providing better performance in terms of error and variability induced by the sensing matrix A in Eq. (3.3). MPS algorithm provides plausible solutions according to patterns extracted from a training image, but not necessary follow the actual underlying structures. Instead, depending on the model and the rate of acquired data, the proposed method is able to estimate the actual channelized structure in regimes where MPS offers a high variability level in its solutions. Fig. 4.8, 4.9 and 4.10 illustrate all previous discussions by providing concrete examples of recovered images for different sampling ratios and methods.

To conclude, it is implicit the complementary nature of the geological and signal information. On one hand, the proposed method outperforms the standard ℓ_1-minimization, which uses only prior signal models (sparsity and compressibility hypothesis). On the other hand, this weighted sparse signal recovery approach offers better recovery performances than MPS algorithm, which utilizes only spatial information (spatial patterns statistics). Thus, addressing jointly ill-posed problems in geoscience; i.e., by using spatial and signal information, provides promising new avenues for geological fields reconstruction.

Chapter 5

Conclusion

This thesis proposes the use of ℓ_1-minimization as an attractive technique to solve ill-posed problems in goescience. This work extends the previous reached outcomes based on standard sparse signal recovery. First, a methodology to incorporate geological prior information is elaborated, exploring the weighted ℓ_1-regularization problem and outperforming dramatically the performance of classical CS. Secondly, a more challenging and realistic acquisition regime range from 0.1% to 2% of measurements is explored . Third, three different channelized models are analyzed to provide a more complete study with respect to the sensibility of the method to the complexity of the field. Main results are presented in terms of signal-to-noise ratio and structural similarity indexes, which provides a perceptual indicator of similarity between images.

One of the major contributions of this work is the incorporation of geological prior statistical information (from a training image) in the sparse signal recovery framework. Particularly, multiple-points simulation is used to generate several realizations of the field to estimate a spatial statistical model. From this information, the statistical support of the model is extracted computing the DCT transform of each element of the database. The main hypothesis of this work is that the support of the field is persistent across different realizations of the permeability map, which is validated across various experimental settings and models. Thus, MPS is presented as an excellent information source to estimate the location of the most significant coefficients in the transform domain of the actual field. Additionally, several definitions of weights were analyzed, illustrating the impact of the weighting matrix construction.

This approach has shown outstanding recovery performances even in a critical undersampling scenario below 1% of data. A study of three different channelized models is elaborated to provide a more complete analysis with respect to the sensibility of the method to the complexity of the field. From this analysis, the method is able to estimate most significant global features, such as complexity, orientation and dominant channel structure; with independence of the model if the sampling ratio is near to 1% of measurements. This bound can be reduced depending of the specific model: 0.3% in the case of single channel and 0.7% in the case of multichannel 1.

The method proposed in this thesis outperform the standard weighted and unweighted ℓ_1 minimization, the inter-block CS method [4] and the widely used multiple-point simulation in a large range of sampling ratio. The reconstruction of the image at different scales enhances the standard weighted ℓ_1-minimizer, and the incorporation of the weighting matrix provides outstanding improvement of the inter-block CS recovery scheme. These two aspects are combined to obtain a more robust recovery method, in terms of reconstruction error and variability.

Finally, it is remarkable to note that the proposed method provides better solutions that the MPS algorithm for a large range of sampling regimes. Thus, the complementary nature of the signal information and spatial statistical pattern is validated, providing a robust recovery method that outperforms both standard ℓ_1-minimizer and MPS technique.

5.1 Future Work

We note that much work have to be done to improve the post-processing stage. A more elaborated scheme can be developed to promote channalized features, such as connectivity, preserving high/low permeability values ratio, flow measurements and transport indexes; in order to exploit the categorical nature of this types of models. On this, a interesting alternative consists of imposing binary values directly in the recovery algorithm. These motivates the research of mixed integer programming and binary sparse signals recovery techniques.

Other interesting dimension to explore is the extension to 3D-models. On this line, the development of greedy algorithms is an attractive alternative to treat with the *curse of the dimensionality*, which is one of the main roadblocks on 3D-model analysis. Current solutions, such as CoSAMP, orthogonal matching pursuit, approximate message passing and block-sparse structure; can be explored to extend the approach to the 3D analysis.

Finally, most CS-based methods rely on a non-adaptive sensing scheme; i.e., the sampling process (linear in the case of CS theory) is independent to the measured object (randomly). New avenues of sequential or adaptive sampling schemes in the context of sparse signal recovery have been developed in recent years. Thus, the exploration of adaptive sparse regularization theory is presented as an attractive area of further research for geological ill-posed problems, where both samples location and recovery problem can be addressed simultaneous.

Bibliography

[1] P. K. Katanidis, *Introduction to geostatistics: applications to hydrogeology*. Cambridge Univ Press, 1997.

[2] G. Mariethoz and J. Caers, *Multiple-point geostatistics: stochastic modeling with training images*. Wiley-Blackwell, 2014.

[3] J. M. Ortiz and C. V. Deutsch, "Indicator simulation accounting for multiple-point statistics," *Math. Geol.*, vol. 36, no. 5, pp. 545–565, 2004.

[4] H. Calderón, J. F. Silva, J. M. Ortiz, and A. Egaña, "Reconstruction of channelized geological facies based on RIPless compressed sensing," *Computers & Geosciences*, vol. 77, no. 0, pp. 54 – 65, 2015.

[5] B. Jafarpour, V. K. Goyal, D. B. McLaughlin, and W. T. Freeman, "Compressed history matching: Exploiting transform-domain sparsity for regularization of nonlinear dynamic data integration problems," *Mathematical Geosciences*, vol. 42, no. 1, pp. 1–27, 2010.

[6] M. Elad, J.-L. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340 – 358, 2005.

[7] I. Daubechies and G. Teschke, "Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising," *Applied and Computational Harmonic Analysis*, vol. 19, no. 1, pp. 1 – 16, 2005.

[8] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *Signal Processing Magazine, IEEE*, vol. 20, pp. 21–36, May 2003.

[9] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203 – 4215, dec. 2005.

[10] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.

[11] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. on Pure and Appl. Math.*, vol. 59, pp. 1207–1223, 2006.

[12] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 – 1306, 2006.

[13] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[14] B. Jafarpour, V. K. Goyal, D. B. McLaughlin, and W. T. Freeman, "Transform-domain sparsity regularization for inverse problems in geosciences," *Mathematical Geosciences*, vol. 74, no. 5, pp. R69–R83, 2010.

[15] F. Herrmann, M. Friedlander, and O. Yilmaz, "Fighting the curse of dimensionality: Compressive sensing in exploration seismology," *Signal Processing Magazine, IEEE*, vol. 29, pp. 88–100, May 2012.

[16] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted $\ell 1$ minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[17] M. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models," *Signal Processing, IEEE Transactions on*, vol. 59, pp. 1985–2001, May 2011.

[18] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *ArXiv e-prints*, Sept. 2011.

[19] S. Strebelle, "Conditional simulation of complex geological structures using multiple-point statistics," *Mathematical Geology*, vol. 34, pp. 1–21, January 2002.

[20] F. Guardiano and M. Srivastava, "Multivariate geostatistics: Beyond bivariate methods," in *Geostatistics-Troia*, (Amsterdam), pp. 133–144, Klumer Academic, 1993.

[21] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. AIEE*, vol. 47, pp. 617–644, Apr. 1928.

[22] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech, J.*, vol. 27, pp. 379–423; 623–656, 1948.

[23] D. L. Donoho, "For most underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[24] E. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[25] J. D. McEwen and Y. Wiaux, "Compressed sensing for wide-field radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 413, pp. 1318–1332, May 2011.

[26] M. Elad, *Sparse and Redundant Representations*. Springer New York, first edition ed.,

2010.

[27] G. K. Wallace, "The jpeg still picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, April 1991.

[28] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *Signal Processing Magazine, IEEE*, vol. 18, pp. 36 –58, sep 2001.

[29] R. A. Devore, "Nonlinear approximation," *Acta Numerica*, pp. 51–150, 1998.

[30] B. G. Bodmann, P. G. Casazza, and G. Kutyniok, "A quantitative notion of redundancy for finite frames," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 348 – 362, 2011.

[31] M. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, vol. 52, no. 1, pp. 64–79, 2009.

[32] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[33] J. McEwen and Y. Wiaux, "Compressed sensing for radio interferometric imaging: Review and future direction," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 1313–1316, Sept 2011.

[34] J. Gemmeke, H. Van Hamme, B. Cranen, and L. Boves, "Compressive sensing for missing data imputation in noise robust speech recognition," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 272–287, 2010.

[35] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 3075–3085, 2009.

[36] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.

[37] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346–2356, 2008.

[38] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible Distributions for High-dimensional Statistics," *ArXiv e-prints*, Feb. 2011.

[39] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *Signal Processing, IEEE Transactions on*, vol. 59, pp. 5193–5201, Nov 2011.

[40] A. Amini and M. Unser, "Sparsity and infinite divisibility," *Information Theory, IEEE Transactions on*, vol. 60, pp. 2346–2358, April 2014.

[41] J. Silva and M. Derpich, "On the characterization of ℓ_p-compressible ergodic sequences,"

Signal Processing, IEEE Transactions on, vol. 63, pp. 2915–2928, June 2015.

[42] H. D. Bondell and B. J. Reich, "Simultaneous Regression Shrinkage, Variable Selection, and Supervised Clustering of Predictors with OSCAR," *Biometrics*, vol. 64, pp. 115–123, 2008.

[43] M. Bogdan, E. van den Berg, W. Su, and E. Candes, "Statistical estimation and testing via the sorted L1 norm," *ArXiv e-prints*, Oct. 2013.

[44] X. Zeng and M. Figueiredo, "The atomic norm formulation of oscar regularization with application to the frank-wolfe algorithm," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pp. 780–784, Sept 2014.

[45] M. A. T. Figueiredo and R. D. Nowak, "Sparse Estimation with Strongly Correlated Variables using Ordered Weighted L1 Regularization," *ArXiv e-prints*, Sept. 2014.

[46] D. Needell, "Noisy signal recovery via iterative reweighted l1-minimization," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pp. 113–117, Nov 2009.

[47] X. Chen and W. Zhuo, "Convergence of reweighted ℓ_1 minimization algorithm and unique solution of truncated ℓ_p minimization," *Technical Report, HK Polytech*, 2010.

[48] E. Candès, "The restricted isometry property and its applications for compressed sensing," *C. R. Acad. Sci. Paris*, vol. I 346, pp. 589–592, 2008.

[49] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, pp. 211–231, July 2009.

[50] Y. Zhao and D. Li, "Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems," *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 1065–1088, 2012.

[51] B. Jafarpour, "Wavelet reconstruction of geologic facies from nonlinear dynamic flow measurements," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, pp. 1520–1535, May 2011.

[52] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[53] F. J. Herrmann and X. Li, "Efficient least-squares imaging with sparsity promotion and compressive sensing," *Geophysical Prospecting*, vol. 60, pp. 696–712, 7 2012.

[54] R. Klees and R. Haagmans, *Wavelets in the Geosciences*. Springer; 2000 edition, 2000.

[55] E. J. Candés and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *The Annals of Statistics*, vol. 30, no. 3, pp. pp. 784–842, 2002.

[56] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 –612, 2004.

[57] X. Zhou and W. Sun, "On the sampling theorem for wavelet subspaces," *The Journal of Fourier Analysis and Applications*, vol. 5, no. 4, pp. 347–354, 1999.

[58] T. Huang, D.-T. Lu, X. Li, and L. Wang, "Gpu-based snesim implementation for multiple-point statistical simulation," *Comput. Geosci.*, vol. 54, pp. 75–87, Apr. 2013.