

ALGORITMO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES  
ADQUIRIDOS MEDIANTE GEOMETRÍAS COMPRESIVAS EN EL DOMINIO  
CROSS-SPREAD EMPLEANDO OPTIMIZACIÓN CONVEXA CON  
APROXIMACIÓN DE BAJO RANGO

IVÁN DAVID ORTIZ PINEDA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2022

ALGORITMO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES  
ADQUIRIDOS MEDIANTE GEOMETRÍAS COMPRESIVAS EN EL DOMINIO  
CROSS-SPREAD EMPLEANDO OPTIMIZACIÓN CONVEXA CON  
APROXIMACIÓN DE BAJO RANGO

IVÁN DAVID ORTIZ PINEDA

Trabajo de Grado para optar al título de  
Ingeniero de Sistemas

Director:

Tatiana Carolina Gelvez Barrera

*Doctora en Ingeniería*

Codirectores:

Henry Arguello Fuentes

*Doctor en Ingeniería Eléctrica y Computación*

Laura Viviana Galvis Carreño

*Doctora en Ingeniería Eléctrica y Computación*

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2022

## **DEDICATORIA**

*Este trabajo está dedicado a mi familia,  
por haber sido mi apoyo a lo largo de mi carrera  
y a lo largo de toda mi vida.*

*A todas las personitas especiales  
que me acompañaron en esta etapa,  
aportando tanto a mi formación profesional  
como de ser humano.*

## **AGRADECIMIENTOS**

Este trabajo de grado fue realizado en el marco del proyecto con código 110287780575 - “Nuevas tecnologías computacionales para el diseño de geometrías de adquisición sísmica 3D con muestreo compresivo para la reducción de costos económicos e impactos ambientales en la exploración de hidrocarburos en cuencas terrestres colombianas”, financiado con recursos provenientes del Convenio 785-2019, suscrito entre la Agencia Nacional de Hidrocarburos, el MINISTERIO y el Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas.

De forma personal, agradezco a mi directora de tesis Tatiana Gelvez, una gran persona a la que llevo con alta estima y admiración, por ser parte fundamental en mi formación tanto académica como personal. En especial, agradezco toda la dedicación y esfuerzo en la dirección de este trabajo de grado.

A las personas del grupo de investigación HDSP, por el apoyo y aprendizaje académico que me brindaron desde el momento que ingresé. Especialmente a Laura Galvis y Paul Goyes, por su revisión y valiosos aportes en la realización de este trabajo de grado.

A Alejandra Hernandez y Brayan Monroy, por su amistad, consejo y apoyo en esta etapa.

## CONTENIDO

	pág.
<b>INTRODUCCIÓN</b>	<b>12</b>
Planteamiento del problema	14
<b>1 OBJETIVOS</b>	<b>17</b>
<b>2 MARCO DE REFERENCIA</b>	<b>18</b>
2.1 FUNDAMENTOS TEÓRICOS	18
2.1.1 Datos sísmicos.	18
2.1.2 Adquisición sísmica con Geometría Ortogonal.	19
2.1.3 Problema de reconstrucción de datos sísmicos.	20
2.1.4 Bajo rango.	21
2.1.5 Algoritmo de ajuste de bloques y filtrado 3D (BM3D).	23
2.2 ANTECEDENTES DE LA INTERPOLACIÓN EN DATOS SÍSMICOS	24
<b>3 METODOLOGÍA PROPUESTA</b>	<b>29</b>
3.1 FORMULACIÓN MATEMÁTICA DEL MODELO DE ADQUISICIÓN	29
3.1.1 Modelo en notación vectorial.	30
3.1.2 Modelo en notación tensorial.	31
3.2 ANÁLISIS DE LA PROPIEDAD DE BAJO RANGO EN DATOS SÍSMICOS	32
3.3 PROBLEMA DE OPTIMIZACIÓN CONVEXA PARA LA RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES	35
3.3.1 Factorización tensorial sobre el modo $m$ .	35
3.3.1.1 Cálculo de la matriz de proyección $P_m$ .	35
3.3.2 Formulación del problema de optimización.	36

3.4 ALGORITMO NUMÉRICO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES	38
3.4.1 Paso de actualización de $\mathcal{X}$ .	40
3.4.2 Paso de actualización de $\mathcal{Z}_1$ .	42
3.4.3 Paso de actualización de $\mathcal{Z}_2$ .	43
3.4.4 Pseudocódigo del algoritmo.	44
3.4.4.1 Análisis de convergencia.	45
3.4.4.2 Análisis de complejidad computacional.	46
3.4.5 Implementación computacional del algoritmo.	47
<b>4 CONFIGURACIÓN EXPERIMENTAL</b>	<b>48</b>
4.1 CONJUNTOS DE DATOS	48
4.2 MÉTRICAS	49
4.3 ALGORITMOS DE COMPARACIÓN	51
4.4 DISEÑO DE EXPERIMENTOS	54
4.4.1 Configuración experimental.	54
4.4.2 Configuración estudio de ablación.	54
<b>5 RESULTADOS Y DISCUSIÓN</b>	<b>55</b>
5.1 VALIDACIÓN CRUZADA	55
5.2 RESULTADOS DE SIMULACIONES	57
5.3 ESTUDIO DE ABLACIÓN SOBRE LA TASA DE SUBMUESTREO	58
<b>6 CONCLUSIONES</b>	<b>63</b>
<b>7 TRABAJO FUTURO</b>	<b>65</b>
<b>BIBLIOGRAFÍA</b>	<b>66</b>

## LISTA DE FIGURAS

	<b>pág.</b>
Figura 1	
Conceptos de datos sísmicos. a) Proceso de sísmica de reflexión. b) Traza sísmica. c) Disparo sísmico. d) Cubo sísmico.	19
Figura 2	20
Adquisición de disparos sísmicos para un <i>cross-spread</i>	
Figura 3	
Distribución de los valores singulares SVD de una matriz a) de rango completo, y b) de bajo rango.	22
Figura 4	
Esquema del funcionamiento del algoritmo BM3D. Esta imagen fue traducida de Wang <sup>1</sup> .	24
Figura 5	
Notaciones de las secciones en un tensor sísmico. a) <i>Cross-spread</i> con $I_1$ , $I_2$ , e $I_3$ dimensiones temporal, de receptores y de fuentes, respectivamente. b) Secciones temporales, una matriz que contiene la medida de todos los receptores en respuesta a todas las fuentes en un tiempo específico. c) Secciones de receptores, una matriz que contiene la respuesta de un único receptor a todas las fuentes. d) Disparos, una matriz que contiene la respuesta de todos los receptores ante una única fuente.	30
Figura 6	
Modelo de observación para una geometría ortogonal compresiva. a) Geometría ortogonal con líneas paralelas de fuentes perpendiculares a líneas de receptores paralelas. b) <i>Cross-spread</i> submuestreado en fuentes. c) Matriz de adquisición $\Phi$ . d) Datos capturados con fuentes faltantes.	32

---

<sup>1</sup> Dong Wang, Jia Xu y Ke Xu. "An FPGA-based hardware accelerator for real-time block-matching and 3D filtering". En: *IEEE Access* 8 (2020), págs. 121987-121998.

- Figura 7 Propiedad de bajo rango para los datos sísmicos. La caída de SVD es estimada para el modo temporal, de receptores y fuentes sobre las dos bases de datos. La amplitud de los valores singulares se encuentra normalizada y se presenta en escala logarítmica. El porcentaje de decaimiento entre el primer y el segundo valor singular se muestra para cada modo sobre el acercamiento realizado en las caídas. a) El decaimiento de la base de datos del SEAM phase II muestra el promedio y desviación de 72 *cross-spread* presentes en la adquisición. b) La caída de Stratton muestra la caída de valores singulares para un único *cross-spread*. 34
- Figura 8 Mapa de calor sobre la métrica PSNR para diferentes combinaciones de los parámetros  $\mu_1$ ,  $\mu_2$  y  $\rho$  sobre los dos conjuntos de datos. Se puede observar que se presentan cambios suaves, donde las mejores reconstrucciones se obtuvieron para  $\rho = 0.001$ . Además, el algoritmo presenta un mejor desempeño cuando alguno de los parámetros  $\mu_\ell$  se configura en el orden de  $10^{-6}$ . 60
- Figura 9 Resultados visuales para un disparo 43 reconstruido del SEAM phase II. 61
- Figura 10 Resultados visuales para un disparo 4 reconstruido de Stratton. 62

## LISTA DE TABLAS

	<b>pág.</b>
Tabla 1    Resumen de notación tensorial	29
Tabla 2    Comparación de la calidad en los disparos recuperados para los conjuntos de datos del SEAM phase II y Stratton. Cada valor es el promedio de todos los disparos eliminados. El PSNR y SNR es medido en decibelios, el tiempo computacional se presenta en segundos, y el MSE se encuentra escalado en el orden de $10^{-3}$ .	57
Tabla 3    Comparación de la calidad en los disparos recuperados para los conjuntos de datos del SEAM phase II y Stratton en función del porcentaje de fuentes faltantes. Cada valor es el promedio de todos los disparos eliminados. El PSNR es medido en decibelios.	59

## RESUMEN

**TÍTULO:** ALGORITMO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES ADQUIRIDOS MEDIANTE GEOMETRÍAS COMPRESIVAS EN EL DOMINIO CROSS-SPREAD EMPLEANDO OPTIMIZACIÓN CONVEXA CON APROXIMACIÓN DE BAJO RANGO \*

**AUTOR:** IVÁN DAVID ORTIZ PINEDA \*\*

**PALABRAS CLAVE:** Aproximación tensorial de bajo rango, recuperación de datos sísmicos, geometría de adquisición ortogonal, dominio *cross-spread*, problemas de optimización.

### DESCRIPCIÓN:

La adquisición de datos sísmicos captura un campo de ondas que permite detectar las propiedades de la tierra que sirven de apoyo a las exploraciones geológicas. Los datos capturados pueden estar incompletos debido a las limitaciones ambientales, el mal funcionamiento del hardware o, de especial interés para la industria, la eliminación de elementos del sistema de adquisición para reducir costos e impacto ambiental. Los datos faltantes pueden recuperarse mediante técnicas computacionales, como el uso de principios de muestreo compresivo. No obstante, la mayoría de estos métodos se centra en el problema de recuperación de datos a causa de receptores deficientes en lugar de fuentes sísmicas eliminadas. Por otro lado, las campañas sísmicas se realizan generalmente en superficies adyacentes con propiedades similares, generando altas correlaciones en los datos sísmicos que no han sido incorporados en los métodos de recuperación. Por lo tanto, en esta tesis se desarrolló un método de recuperación de datos sísmicos que aprovecha las altas redundancias estructurales exhibidas en las secciones del dominio *cross-spread*. Específicamente, se introdujo una aproximación tensorial de bajo rango sobre las componentes principales en los diferentes modos del tensor sísmico. Los experimentos validaron la eficacia del método de recuperación de datos sísmicos utilizando conjuntos de datos sintéticos-realistas y realistas, destacando que la propuesta supera el estado del arte en la recuperación de datos faltantes de una fuente eliminada, obteniendo una ganancia de hasta 1.5 dB y 0.07 en la calidad de la señal (PSNR) y similitud estructural (SSIM).

---

\* Trabajo de grado

\*\* Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Tatiana Carolina Gelvez Barrera. Codirectores: Henry Arguello Fuentes, Laura Viviana Galvis Carreño

## ABSTRACT

**TITLE:** RECONSTRUCTION ALGORITHM FOR MISSING SEISMIC DATA ACQUIRED THROUGH COMPRESSIVE GEOMETRIES IN THE CROSS-SPREAD DOMAIN USING CONVEX OPTIMIZATION WITH LOW-RANK APPROXIMATION \*

**AUTHOR:** IVÁN DAVID ORTIZ PINEDA \*\*

**KEYWORDS:** Low-rank tensor approximation, Seismic source recovery, Orthogonal acquisition geometry, Cross-spread domain, Optimization problems.

### DESCRIPTION:

Seismic data acquisition captures a wavefield for detecting earth properties that support geological explorations. The captured data can be incomplete because of environmental limitations, hardware malfunction, or, of particular industry interest, the elimination of devices to reduce the acquisition costs and environmental impacts. The missing data can be computationally recovered by applying compressive sensing principles. However, most of literature methods focus on recovering missing data from deficient receivers instead of recovering missing data from a removed source. On the other hand, seismic surveys are generally performed on adjacent surfaces with similar properties, generating high correlations in the seismic data that have not been incorporated in the recovery methods. Therefore, this thesis developed a seismic data recovery method that takes advantage of high structural redundancies exhibited in the cross-spread domain slices. Specifically, the proposed method introduces a low-rank tensor approximation over the principal components in the different seismic tensor modes. Experiments validated the seismic data recovery method's effectiveness using synthetic-realistic and realistic datasets, highlighting that the proposal outperforms the state-of-the-art in recovering the missing data from a removed source, obtaining a gain of up to 1.5 dB and 0.07 in signal quality (PSNR) and structural similarity (SSIM).

---

\* Bachelor Thesis

\*\* Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Tatiana Carolina Gelvez Barrera. Co-advisors: Henry Arguello Fuentes, Laura Viviana Galvis Carreño

## INTRODUCCIÓN

La exploración de cuencas sedimentarias terrestres busca el diseño y desarrollo de esquemas eficientes para la adquisición de datos sísmicos en torno a dos ejes principales: (i) el reto que enfrentan las empresas para adquirir datos de alta calidad, bajo un presupuesto establecido, y (ii) las regulaciones ambientales que restringen el uso de equipos de levantamiento sísmico en ciertas áreas de exploración <sup>1</sup>. La geometría ortogonal es el esquema de adquisición más usado en Colombia, la cual posiciona un conjunto de líneas de receptores paralelas entre sí, de forma ortogonal a un conjunto de líneas de fuentes paralelas entre sí <sup>2</sup>.

La geometría ortogonal permite la obtención y procesamiento de datos sísmicos en el dominio cross-spread, el cual consiste en un volumen de datos sísmicos definido por la intersección de una única línea de fuentes con una única línea de receptores <sup>3</sup>. En este dominio, una traza es la medición en el tiempo que registra un receptor para una única fuente; estas trazas son apiladas en una matriz que normalmente se conoce como disparo sísmico, y el apilamiento de estos disparos forma lo que se denomina un cubo sísmico, representado matemáticamente como un tensor. La calidad de los datos adquiridos depende de la cantidad de fuentes y receptores empleados.

---

<sup>1</sup> Xander Campman y col. "Sparse seismic wavefield sampling". En: *The Leading Edge* 36.8 (2017), págs. 654-660.

<sup>2</sup> Kareth León López y col. "Regular multi-shot subsampling and reconstruction on 3d orthogonal symmetric seismic grids via compressive sensing". En: *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE. 2019, págs. 1-5.

<sup>3</sup> Gijts JO Vermeer. *3D seismic survey design*. Society of Exploration Geophysicists, 2012.

Sin embargo, las fuentes se generan normalmente con explosivos costosos o vibradores sísmicos difíciles de transportar en ciertas regiones, que además causan un impacto negativo al ambiente <sup>4</sup>. Por lo tanto, en el ámbito investigativo e industrial existe un gran interés en el desarrollo de esquemas de adquisición que reduzcan el número de fuentes, en donde la calidad se mantenga mediante un algoritmo de reconstrucción.

Matemáticamente, la reconstrucción de datos sísmicos faltantes es un problema mal condicionado, por lo que es necesario el uso de información *a priori* relacionada con propiedades naturales de los datos mediante un término de regularización <sup>5</sup>. Por ejemplo, la propiedad de escasez indica que los datos sísmicos se pueden representar con pocos coeficientes en una base de transformación. Por otro lado, la propiedad de bajo rango asume que los datos sísmicos contienen altas correlaciones espaciales debido a que son medidos en superficies adyacentes, tal que se pueden representar en un subespacio de baja dimensión, o aproximar con una estructura de bajo rango <sup>6</sup>.

La propiedad de bajo rango se ha incluido en métodos de reconstrucción mediante la regularización en un dominio de transformación, como el dominio de Hankel, enfrentando varias limitaciones <sup>7</sup>. Por ejemplo, la proyección sobre la matriz de Hankel

---

<sup>4</sup> Ofelia P Villarreal y col. "Compressive sensing seismic acquisition by using regular sampling in an orthogonal grid". En: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, págs. 1-5.

<sup>5</sup> Paul Goyes-Peñafiel y col. "A Consensus Equilibrium Approach for 3-D Land Seismic Shots Recovery". En: *IEEE Geoscience and Remote Sensing Letters* 19 (2021), págs. 1-5.

<sup>6</sup> Jonathan Popa, Susan E Minkoff y Yifei Lou. "An improved seismic data completion algorithm using low-rank tensor optimization: Cost reduction and optimal data orientation". En: *Geophysics* 86.3 (2021), págs. V219-V232.

<sup>7</sup> Yangkang Chen y col. "The interpolation of sparse geophysical data". En: *Surveys in Geophysics*

genera altas redundancias, implicando el procesamiento de matrices de gran tamaño. Además, el proceso de regularización requiere la descomposición en valores singulares, incrementando el costo computacional. Finalmente, la proyección sobre la matriz se realiza sobre pequeñas subregiones, teniendo en cuenta únicamente información local y desaprovechando las correlaciones no-locales.

Por lo tanto, en este trabajo de grado se desarrolló un algoritmo de reconstrucción de datos sísmicos en el dominio *cross-spread* utilizando aproximación de bajo rango. A diferencia de los métodos en la literatura, se propuso un término de regularización sobre una descomposición tensorial que hace uso de las componentes principales de las dimensiones del tensor para aprovechar las correlaciones y redundancias estructurales entre los modos del tensor representando el cubo sísmico.

## **Planteamiento del problema**

Los procesos de exploración geofísica permiten conocer la estructura superficial de la tierra y detectar la posición de valiosos depósitos geológicos como hidrocarburos, combustibles fósiles, minerales o reservas de agua subterránea. Por lo tanto, la adquisición de datos sísmicos es un campo de gran interés para las empresas de hidrocarburos, especialmente para Colombia, considerando que la industria petrolera es uno de los principales ejes en la actividad económica del país. Según cifras de la Agencia Nacional de Hidrocarburos, el potencial petrolífero se estima en más de 37000 millones de barriles de petróleo cada cuanto, siendo el principal producto de exportación con un 55.4 % del total de exportaciones <sup>8</sup>.

---

40.1 (2019), págs. 73-105.

<sup>8</sup> *La Agencia Nacional de Hidrocarburos de Colombia implementa tecnología Blockchain para la asignación de áreas de explotación | ACIS.* <https://acis.org.co/portal/content/noticiasdeinteres/la-agencia-nacional-de-hidrocarburos-de-colombia-implementa-tecnolog%C3%ADa-blockchain->

Las empresas de hidrocarburos buscan el desarrollo de geometrías de adquisición eficientes en costo-beneficio, que minimicen el impacto ambiental. Por ejemplo, las recientes geometrías compresivas buscan reducir la cantidad de receptores y/o fuentes usadas en el proceso de adquisición, estimando los datos faltantes mediante el procesamiento digital con un algoritmo computacional <sup>9</sup>. En ese sentido, este trabajo de grado fue realizado en el marco del proyecto con código 110287780575 - “Nuevas tecnologías computacionales para el diseño de geometrías de adquisición sísmica 3D con muestreo compresivo para la reducción de costos económicos e impactos ambientales en la exploración de hidrocarburos en cuencas terrestres colombianas” financiado con recursos provenientes del Convenio 785-2019, suscrito entre la Agencia Nacional de Hidrocarburos, el MINISTERIO y el Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas. Específicamente, se apoyó el objetivo específico “Desarrollar e implementar algoritmos computacionales basados en optimización convexa y redes neuronales para reconstrucción de fuentes sísmicas en geometrías 3D”.

Particularmente, en este trabajo de grado se desarrolló e implementó un algoritmo computacional basado en optimización convexa incluyendo la propiedad de bajo rango. Este enfoque busca aprovechar las redundancias estructurales en la adquisición de datos sísmicos en el dominio cross-spread generadas en superficies adyacentes <sup>6</sup>. Es decir, la alta correlación entre los datos permite describir una traza como una combinación lineal de otro conjunto de trazas. A diferencia de los métodos en la

---

para-la. Accessed: 2021-02-16.

<sup>9</sup> Xiaogang Huang y col. “Efficient seismic acquisition based on compressive sensing”. En: *SEG 2020 Workshop: Broadband and Wide-azimuth Deepwater Seismic Technology*. Society of Exploration Geophysicists. 2020, págs. 35-37.

literatura actual como los basados en la transformación de Hankel <sup>7</sup> o de textura a parches <sup>10</sup>, el método propone una regularización en el dominio directo de los datos, considerando además correlaciones no locales. Puntualmente, con este trabajo se buscó responder la pregunta de investigación:

¿Es posible reconstruir datos sísmicos a partir de medidas adquiridas en el dominio cross-spread por medio de un algoritmo de optimización convexa basado en la aproximación de bajo rango?

Este documento se encuentra organizado de la siguiente manera: en la Sección 1 se enumeran los objetivos formulados para este trabajo de grado. La Sección 2 describe los conceptos relacionados con la adquisición de datos sísmicos, el problema de reconstrucción de datos sísmicos faltantes, la teoría del rango matricial y vectorial, así como una revisión del estado del arte sobre métodos de reconstrucción de datos sísmicos faltantes. La Sección 3 describe la metodología propuesta para el desarrollo del algoritmo de reconstrucción de datos sísmicos faltantes utilizando optimización convexa con aproximación de bajo rango. La sección 4 se describe la configuración experimental de las simulaciones realizadas. La Sección 5 presenta los resultados para la evaluación del método, así como comparaciones con otros métodos del estado del arte. Finalmente, la Secciones 6 y 7 presenta conclusiones y trabajo a futuro, respectivamente.

---

<sup>10</sup> Jianwei Ma. "Three-dimensional irregular seismic data reconstruction via low-rank matrix completion". En: *Geophysics* 78.5 (2013), págs. V181-V192.

## 1. OBJETIVOS

En esta tesis se cumplieron los siguientes objetivos:

### Objetivo general

- Desarrollar un algoritmo basado en optimización convexa con aproximación de bajo rango para reconstruir datos sísmicos a partir de medidas adquiridas con geometrías compresivas en el dominio *cross-spread*.

### Objetivos específicos

1. Formular matemáticamente el modelo de adquisición de datos sísmicos parcialmente observados en el dominio *cross-spread* usando geometrías compresivas. (Ver sección 3.1)
2. Analizar y validar la propiedad de bajo rango en datos sísmicos en el dominio *cross-spread*. (Ver sección 3.2)
3. Plantear un problema de optimización convexa para la reconstrucción de datos sísmicos incorporando la propiedad de bajo rango a través de un término de regularización. (Ver sección 3.3)
4. Implementar un algoritmo numérico para resolver el problema de optimización planteado. (Ver sección 3.4)
5. Evaluar la efectividad del algoritmo implementado comparando el rendimiento con respecto a métodos del estado del arte. (Ver sección 5)

## 2. MARCO DE REFERENCIA

### 2.1. FUNDAMENTOS TEÓRICOS

Este trabajo de grado se fundamenta en el procesamiento computacional de datos sísmicos, la formulación matemática de problemas de optimización y la implementación de algoritmos iterativos. Específicamente, se emplea el concepto de rango de una matriz, matriz de bajo rango, y regularización de bajo rango para abordar el mal condicionamiento de problemas inversos.

**2.1.1. Datos sísmicos.** La **sísmica de reflexión** es un método de exploración geofísica, basado en principios físicos de ondas sísmicas reflejadas, que sirve para la estimación de propiedades del suelo y patrones de las rocas. Específicamente, la sísmica de reflexión utiliza fuentes de energía sísmica controlada como explosivos o excitadores sísmicos para inducir vibraciones. Tales vibraciones viajan a través de las capas del suelo y generan reflexiones ante fenómenos físicos, como cambios de impedancia acústica, que son registradas por receptores en la superficie conocidos como geófonos. Estos arribos de energía medidos por los receptores son lo que se conocen como eventos sísmicos.

Se conoce como **traza sísmica** a los datos sísmicos medidos en un solo canal, es decir, el registro realizado por un único geófono en respuesta a la energía inducida sobre el suelo por una única fuente. Las trazas sísmicas medidas en superficies adyacentes son apiladas verticalmente para formar **datos sísmicos bidimensionales** o **disparos sísmicos**. Computacionalmente, los disparos sísmicos son almacenados como una matriz, y en estos los eventos sísmicos se pueden ver como ondículas que siguen cierta geometría (líneas o hipérbolas).

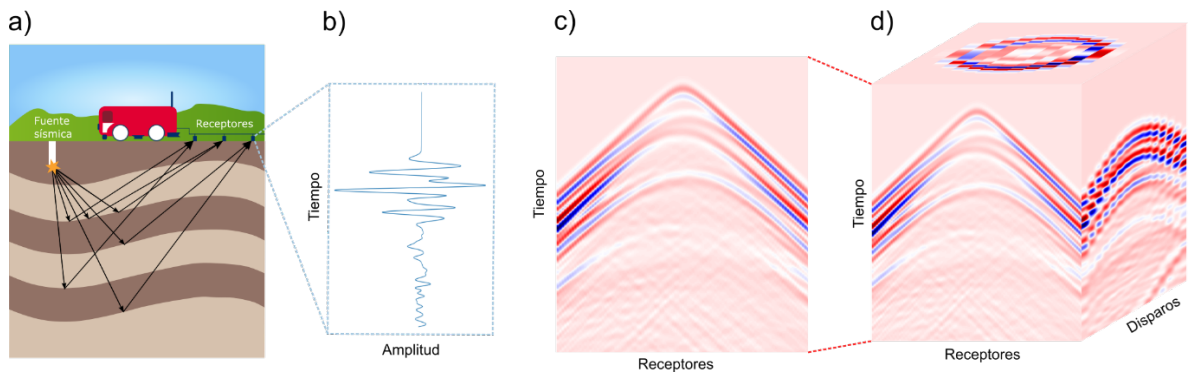


Figura 1. Conceptos de datos sísmicos. a) Proceso de sísmica de reflexión. b) Trazas sísmicas. c) Disparo sísmico. d) Cubo sísmico.

De forma análoga, los **datos sísmicos tridimensionales** son apilamientos de diferentes disparos a lo largo de un cubo, donde cada dimensión representa la muestra temporal, el receptor y la fuente empleada. Sobre estos tensores de datos se definen tres tipos de matrices dependiendo de la dimensión que permanece constante: las secciones en el tiempo, las secciones de receptores y las secciones de fuentes o disparos.

En la Figura 1 se ilustran estos conceptos, donde en a) se muestra el proceso de sísmica de reflexión con una fuente sísmica que induce ondas a la superficie terrestre, y cada geófono registra la amplitud de la onda en una traza sísmica como se observa en b). En c) se muestra el apilamiento de estas trazas en un disparo sísmico, donde los colores de la imagen codifican el valor de la amplitud. En d) se muestra el cubo de datos sísmicos obtenido de apilar diferentes disparos.

**2.1.2. Adquisición sísmica con Geometría Ortogonal.** Según la disposición en las que se ubican las líneas de receptores y fuentes se diseña una geometría de adquisición sísmica. La **geometría ortogonal** propone una adquisición sísmica en la que las líneas de receptores y de fuentes se ubican de forma perpendicular entre sí.

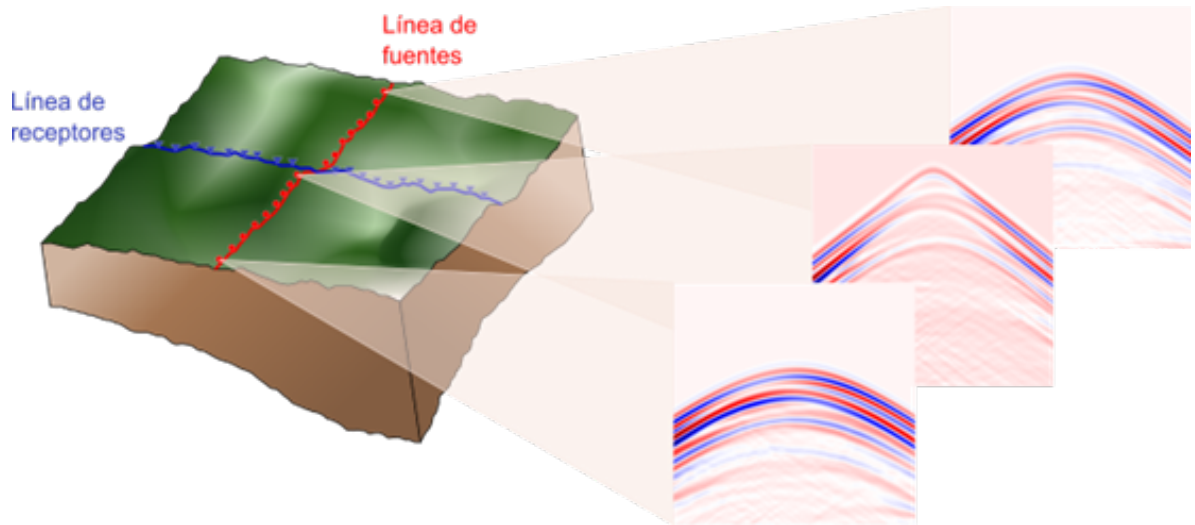


Figura 2. Adquisición de disparos sísmicos para un *cross-spread*

El *cross-spread* es la unidad de adquisición de la geometría ortogonal correspondiente a la intersección de una línea de fuentes y una de receptores. El dato sísmico adquirido usando el *cross-spread* es un cubo sísmico con dimensiones dependientes de las muestras temporales, el número de geófonos (receptores) y el número de fuentes.

En la Figura 2 se muestra la distribución de un *cross-spread* compuesto por una línea de receptores (azul) y una línea de fuentes (roja). En ella, cada fuente induce un sismo controlado que es registrado por la línea de receptores, resultando un disparo sísmico.

**2.1.3. Problema de reconstrucción de datos sísmicos.** El problema de reconstrucción de datos sísmicos parte de un modelo de observación en el que se quiere recuperar un *cross-spread*  $\mathbf{x} \in \mathbb{R}^{I_t I_r I_s}$ , donde  $I_t$ ,  $I_r$  e  $I_s$  son el número de muestras temporales, receptores y fuentes sísmicas respectivamente, a partir de un vector de medidas comprimidas  $\mathbf{y} \in \mathbb{R}^{I_t I_r I_s}$  donde las trazas faltantes son representadas con ceros.

Así, se define la relación lineal entre los datos a reconstruir y las medidas comprimidas mediante la ecuación

$$\mathbf{y} = \tilde{\Phi}\mathbf{x} \quad (1)$$

donde  $\tilde{\Phi} \in \{0, 1\}^{I_t I_r I_s \times I_t I_r I_s}$  es una matriz binaria que codifica con 1 las posiciones de trazas conocidas y con 0 las no observadas.

El problema de optimización relacionado con la reconstrucción de datos sísmicos parte de un enfoque variacional, donde usa como función de costo un término de fidelidad de los datos que asegura que las medidas adquiridas coincidan con las reconstruidas y un término de regularización que permita mejorar el condicionamiento del problema agregando asunciones sobre los datos.

Matemáticamente, el problema de optimización está definido como:

$$\underset{\mathbf{x}}{\text{minimizar}} \left\{ \frac{1}{2} \|\mathbf{y} - \tilde{\Phi}\mathbf{x}\|_2^2 + \tau \mathfrak{R}(\mathbf{x}) \right\} \quad (2)$$

donde el primer término usa el modelo observacional para plantear el término de fidelidad de los datos, el segundo es el regularizador que mejora el condicionamiento del problema y  $\tau > 0$  es un parámetro que controla el peso dado al regularizador.

Entre los regularizadores usados en el estado del arte se puede destacar la escasez sobre una base de transformación  $\mathfrak{R}(\mathbf{x}) = \|\Psi\mathbf{x}\|_1$ , o el bajo rango sobre un dominio de transformación lineal  $\mathfrak{R}(\mathbf{x}) = \|\mathcal{T}(\mathbf{x})\|_*$  como los descritos por Chen <sup>7</sup>.

**2.1.4. Bajo rango.** En álgebra lineal, se define el rango de una matriz como la dimensión del espacio columna. Es decir, el rango de una matriz coincide con la cantidad de columnas (y de forma análoga filas) que son linealmente independientes.

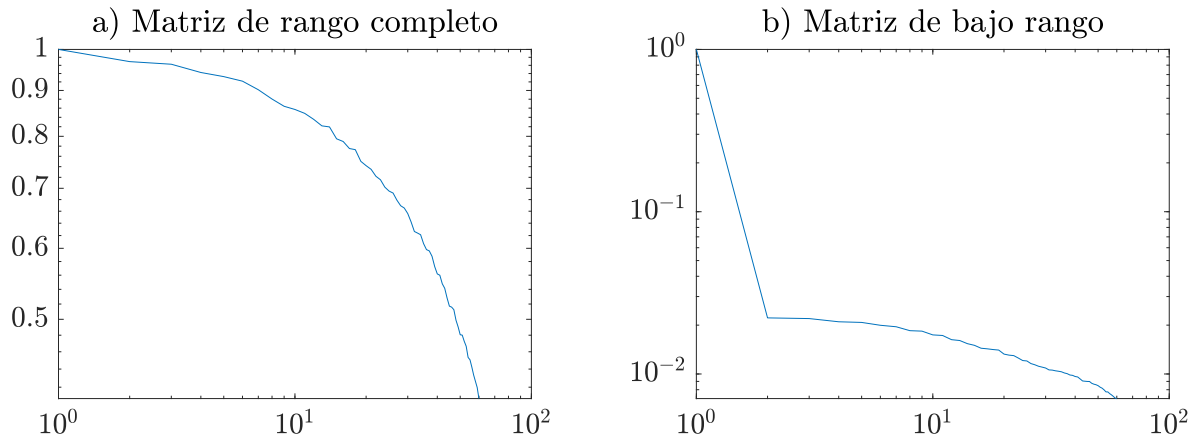


Figura 3. Distribución de los valores singulares SVD de una matriz a) de rango completo, y b) de bajo rango.

El rango de una matriz se encuentra estrechamente relacionado con su descomposición en valores singulares (SVD). Sea  $\mathbf{A}$  una matriz arbitraria, se define la SVD como la factorización de la forma:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (3)$$

donde  $\mathbf{U}$  y  $\mathbf{V}^T$  son bases ortogonales que expanden el espacio fila y el espacio columna de la matriz respectivamente, y  $\mathbf{\Sigma} \succeq 0$  es una matriz diagonal cuyos elementos se conocen como valores singulares. Cada valor singular  $\sigma_i \in \mathbf{\Sigma}$  mide la cantidad de información que aportan los correspondientes vectores columna  $u_i$  y  $v_i$  de las bases a la información de la matriz  $\mathbf{A}$ . Dada la SVD de una matriz, se puede definir el rango como la cantidad de valores singulares no nulos  $\sigma_i \neq 0$ .

Se dice que una matriz posee bajo rango cuando el rango es mucho menor que sus dimensiones. Es decir, se dice que  $\mathbf{A} \in \mathbb{R}^{m \times n}$  posee bajo rango cuando  $\text{rank}(\mathbf{A}) \ll \min(m, n)$ , donde  $\text{rank}(\mathbf{A})$  indica el rango de la matriz. Estas matrices se caracterizan por presentar redundancias estructurales y altas correlaciones a lo largo de sus columnas.

Si bien, no todas las matrices poseen esta propiedad, existen algunas de estas que pueden ser muy bien aproximadas con una matriz de bajo rango. Este tipo de matrices se caracteriza porque la distribución de los valores singulares decrece rápidamente, por lo cual muchos de sus valores singulares son cercanos a cero. Esta tendencia se muestra en la Figura 3.

**2.1.5. Algoritmo de ajuste de bloques y filtrado 3D (BM3D).** El BM3D es un algoritmo de reducción de ruido en imágenes que se basa en una estrategia de concordancia por bloques, en donde se hace una separación de la imagen en parches, y realiza una búsqueda de los parches que encuentren una mayor similitud con el parche de referencia. Específicamente, el BM3D es un algoritmo que consta de 3 pasos:

1. **Paso de agrupamiento:** Se divide la imagen de entrada en parches del mismo tamaño, no necesariamente disjuntos, y en una ventana establecida se agrupan los parches que presenten mayor similitud en base a una métrica establecida.
2. **Paso de filtrado colaborativo:** A cada uno de los grupos se le aplica una transformación lineal 3D, y seguido un filtrado en una transformada (como la transformada de Wiener). Después para obtener los parches originales se aplica la inversa de la transformación lineal.
3. **Paso de agregación:** Se recupera la imagen original a partir de los parches. Si varios parches se superponen se realiza un promedio ponderado asegurando la señal distintiva en cada píxel.

Este esquema del algoritmo en forma general es mostrado en la figura 4.

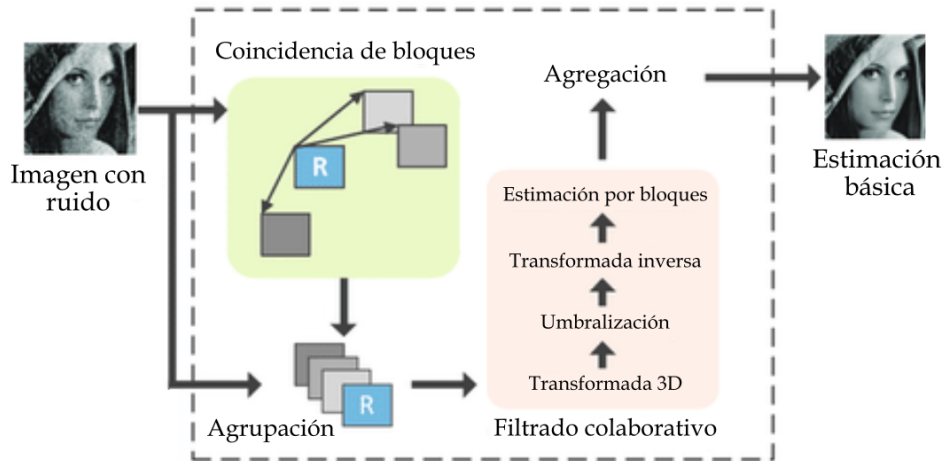


Figura 4. Esquema del funcionamiento del algoritmo BM3D. Esta imagen fue traducida de Wang <sup>11</sup>.

## 2.2. ANTECEDENTES DE LA INTERPOLACIÓN EN DATOS SÍSMICOS

Los métodos de interpolación de trazas (columnas) faltantes dentro de una imagen sísmica bidimensional (matriz) han sido ampliamente estudiados en la literatura. Matemáticamente, este problema cuenta con una cantidad de datos muestreados menor a los que presenta la imagen original, por lo que el sistema lineal de ecuaciones asociado es inconsistente, generando que el problema original de reconstrucción de los datos se encuentre mal condicionado matemáticamente.

Dada esta condición, muchos trabajos del estado del arte han optado por el uso de regularizadores dentro de la formulación del problema para garantizar la existencia de una solución. Entre los términos de regularización utilizados se puede destacar el uso de una norma ponderada para condicionar el problema de interpolación original <sup>12</sup>; el uso de un algoritmo de proyección sobre conjuntos convexos (POCS del

<sup>12</sup> Bin Liu y Mauricio D Sacchi. "Minimum weighted norm interpolation of seismic records". En: *Geophysics* 69.6 (2004), págs. 1560-1568.

inglés *Projection Onto Convex Sets*) combinado con la transformada de Fourier para reconstruir cuadrículas muestreadas regularmente <sup>13</sup>; el algoritmo de POCS ha sido extendido ya sea para mejorar su velocidad de convergencia <sup>14</sup>, o para aplicarse sobre datos muestreados irregularmente <sup>15</sup>; o se usa la reducción de rango como restricción para condicionar el problema de interpolación de trazas <sup>16</sup>.

Dentro de las técnicas de reconstrucción de datos sísmicos, otra metodología que ha sido aplicada para la reconstrucción de trazas y disparos faltantes es el muestreo compresivo (CS, del inglés *compressive sensing*). De forma general, los principios de CS son utilizados para poder diseñar adquisiciones con números de muestras menores al valor teórico dado por el teorema de Nyquist-Shannon <sup>15</sup>, permitiendo reducir considerablemente la cantidad de receptores o fuentes, sin perder la calidad del dato después de la reconstrucción. Los dos principios que se deben cumplir para garantizar la correcta reconstrucción de los datos submuestreados son: primero, disponer de una base de transformación que promueva la escasez de los coeficientes; y segundo, establecer una estrategia de submuestreo que suavice la distorsión coherente <sup>1</sup>, entre las cuales se puede usar una estrategia aleatoria o de Jitter <sup>17</sup>.

---

<sup>13</sup> Ray Abma y Nurul Kabir. "3D interpolation of irregular data with a POCS algorithm". En: *Geophysics* 71.6 (2006), E91-E97.

<sup>14</sup> Shu-Qin Wang, Xing Gao y Zhen-Xing Yao. "Accelerating POCS interpolation of 3D irregular seismic data with graphics processing units". En: *Computers & Geosciences* 36.10 (2010), págs. 1292-1300.

<sup>15</sup> Tao Jiang y col. "Compressive seismic reconstruction with extended POCS for arbitrary irregular acquisition". En: *2017 SEG International Exposition and Annual Meeting*. OnePetro. 2017.

<sup>16</sup> Stewart Trickett y col. "Rank-reduction-based trace interpolation". En: *SEG Technical Program Expanded Abstracts 2010*. Society of Exploration Geophysicists, 2010, págs. 3829-3833.

<sup>17</sup> Wei Liu y col. "Reconstruction of seismic data with missing traces based on local random sampling and curvelet transform". En: *Journal of Applied Geophysics* 115 (2015), págs. 129-139.

Respecto a los métodos de reconstrucción que se basan en los principios de CS, estos proponen resolver un problema de optimización cuya función de costo consiste en un término de fidelidad de los datos submuestreados (asociado a una norma  $\ell_2$ ) regularizado mediante un término de norma  $\ell_1$  para promover la escasez de los coeficientes sobre una base de transformación. Entre las bases de transformación utilizadas se pueden destacar Fourier <sup>18</sup>, Curvelet <sup>17,19</sup> y Wavelet <sup>2,4</sup>. Otra estrategia consiste en usar diferentes bases de transformación por cada dimensión, combinando bases de transformación como coseno discreto con una Wavelet <sup>2</sup>.

En otro trabajo se plantea el uso de una estrategia de optimización de equilibrio consenso para permitir complementar el término de escasez propuesto por la norma  $\ell_1$  junto a otros regularizadores aplicados a los diferentes modos del tensor, usando un algoritmo de ajuste de bloques y filtrado 3D (BM3D del inglés *Block-Matching and 3D filtering*) para la regularización de las amplitudes en los datos <sup>5</sup>.

Por otra parte, en los últimos años se han utilizado métodos de aprendizaje profundo para la reconstrucción de información sísmica. Entre estos trabajos se ha implementado una arquitectura de redes neuronales convolucionales para aprender información sobre el dato sísmico en un enfoque no supervisado <sup>20,21</sup>.

---

<sup>18</sup> Pengliang Yang, Jinghuai Gao y Wenchao Chen. "Irregularly sampled seismic data interpolation using iterative half thresholding regularization". En: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2013, págs. 5820-5824.

<sup>19</sup> Felix J Herrmann y Gilles Hennenfent. "Non-parametric seismic data recovery with curvelet frames". En: *Geophysical Journal International* 173.1 (2008), págs. 233-248.

<sup>20</sup> Qun Liu, Lihua Fu y Meng Zhang. "Deep-seismic-prior-based reconstruction of seismic data using convolutional neural networks". En: *Geophysics* 86.2 (2021), págs. V131-V142.

<sup>21</sup> Fantong Kong y col. "Deep prior-based seismic data interpolation via multi-res U-net". En: *SEG*

Otro enfoque encontrado en la literatura para la reconstrucción de datos sísmicos consiste en condicionar el problema de interpolación mediante regularizadores que promuevan el bajo rango de una matriz. Un enfoque tradicional de estos métodos plantea la interpolación de trazas sísmicas faltantes sobre datos sísmicos bidimensionales utilizando la transformada de Fourier como dominio para promover el bajo rango de los datos y un algoritmo de truncamiento sobre los valores singulares para disminuir el rango de la matriz <sup>16</sup>.

Existen dominios lineales que pueden ser usados para promover el bajo rango dentro de los datos sísmicos. Uno de estos es el dominio de matrices de Hankel, matrices caracterizadas por tener antidiagonales constantes, y permite el uso de algoritmo de reducción de rango para la reconstrucción de información <sup>7</sup>.

Otro dominio de transformación para algoritmos de bajo rango es dominio de parches a texturas <sup>10</sup>, el cual por medio de una transformación de parches tridimensionales proyecta el cubo sísmico a una matriz bidimensional. Se recuperan los elementos de la matriz resultante mediante un método de completado de matrices con una norma nuclear generalizada. Con este enfoque se busca redistribuir la nulidad de las trazas sísmicas dispersamente de tal forma que no se genere una situación de deficiencia de rango. Sin embargo, según la distribución de las trazas igualmente se puede llegar a esta situación.

Uno de los problemas más destacados en el uso de algoritmos de bajo rango es el alto costo que genera el cálculo de la descomposición en valores singulares de una matriz de grandes dimensiones. Para ello, trabajos del estado del arte han propuesto resolver el problema de interpolación mediante un método de factorización matricial en matrices rectangulares que fijar el rango reducido de la matriz, disminuyendo el tiempo de cómputo del algoritmo al no requerir el computo de la descomposición en valores singulares cada iteración <sup>22</sup>.

Estos principios han sido generalizados en algunos trabajos donde se han podido reconstruir datos sísmicos aplicados a un tensor de 5 dimensiones bajo un enfoque de completado de tensores (*tensor completion*) en el dominio de Fourier <sup>6</sup>.

---

<sup>22</sup> Juan Wu y col. "Fast and robust low-rank approximation for five-dimensional seismic data reconstruction". En: *IEEE Access* 8 (2020), págs. 175501-175512.

### 3. METODOLOGÍA PROPUESTA

Para facilitar la lectura de la formulación matemática e implementación del algoritmo de reconstrucción de datos sísmicos faltantes, en la Tabla 1 se introduce la notación tensorial que se utilizará a lo largo de esta sección. Las operaciones tensoriales fueron aplicadas como las define Kolda <sup>23</sup>.

Notación	Descripción
$\mathcal{X}, \mathbf{X}, \mathbf{x}, x$	Tensor, matriz, vector, escalar.
$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$	Tensor de 3 <sup>er</sup> orden con dimensiones $I_1 \times I_2 \times I_3$ .
$\mathbf{X}_{i::} \in \mathbb{R}^{I_2 \times I_3}$	$i$ -ésima sección en el tiempo del tensor $\mathcal{X}$ .
$\mathbf{X}_{:j} \in \mathbb{R}^{I_1 \times I_3}$	$j$ -ésima sección de receptor en el tensor $\mathcal{X}$ .
$\mathbf{X}_{::k} \in \mathbb{R}^{I_1 \times I_2}$	$k$ -ésimo disparo en el tensor $\mathcal{X}$ .
$\mathbf{X}_{(m)} \in \mathbb{R}^{I_m \times \tilde{I}_m}$	Matriz de desdoblamiento sobre el modo $m$ (o matriz del modo $m$ ) del tensor $\mathcal{X}$ . En esta definición $\tilde{I}_m$ es el producto de todas las dimensiones del tensor excepto $I_m$ .
$\mathcal{T} = \mathcal{X} \times_m \mathbf{A}$	Producto tensorial sobre el modo $m$ , multiplicando un tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_m \times \dots \times I_N}$ y una matriz $\mathbf{A} \in \mathbb{R}^{J \times I_m}$ . El producto cumple la correspondencia $\mathbf{T}_{(m)} = \mathbf{A} \mathbf{X}_{(m)}$ sobre el tensor resultante.
$\ \mathcal{X}\ _2^2 = \sum x_{i,j,k}^2$	Norma $\ell_2$ del tensor $\mathcal{X}$ , calculada como la suma de todos los cuadrados de los elementos del tensor.

Tabla 1. Resumen de notación tensorial

#### 3.1. FORMULACIÓN MATEMÁTICA DEL MODELO DE ADQUISICIÓN

Para el desarrollo de un algoritmo de reconstrucción de datos sísmicos faltantes, se planteó como primera etapa de este proyecto el modelamiento matemático del

<sup>23</sup> Tamara G Kolda y Brett W Bader. "Tensor decompositions and applications". En: *SIAM review* 51.3 (2009), págs. 455-500.

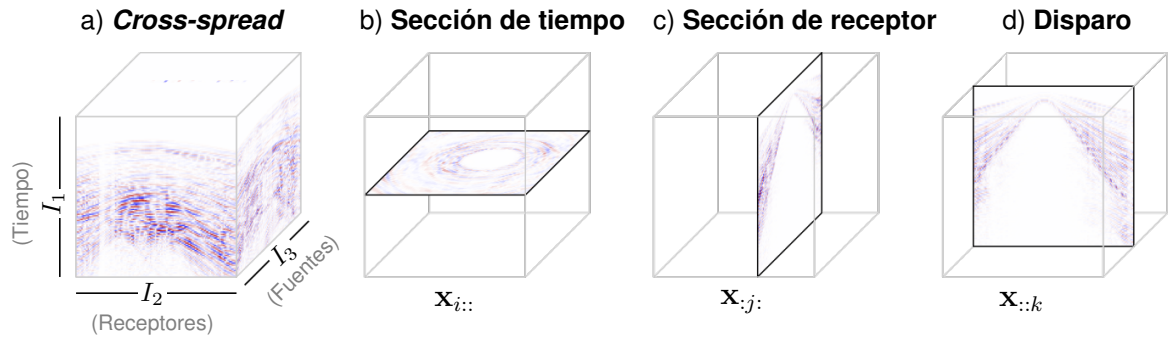


Figura 5. Notaciones de las secciones en un tensor sísmico. a) *Cross-spread* con  $I_1$ ,  $I_2$ , e  $I_3$  dimensiones temporal, de receptores y de fuentes, respectivamente. b) Secciones temporales, una matriz que contiene la medida de todos los receptores en respuesta a todas las fuentes en un tiempo específico. c) Secciones de receptores, una matriz que contiene la respuesta de un único receptor a todas las fuentes. d) Disparos, una matriz que contiene la respuesta de todos los receptores ante una única fuente.

proceso de adquisición sísmica en una geometría compresiva donde se eliminan ciertas fuentes sísmicas.

Se define el tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  como un *cross-spread* sísmico con  $I_1$  muestras en el tiempo,  $I_2$  número de receptores e  $I_3$  fuentes empleadas en la adquisición. A estas dimensiones se le nombrarán a lo largo del presente documento como dimensión temporal, de receptores y de disparos respectivamente. Adicionalmente, la Figura 5 muestra una representación visual de un *cross-spread*, junto con las secciones temporales, de receptores y disparos que definen las dimensiones del tensor sísmico.

**3.1.1. Modelo en notación vectorial.** Como fue mencionado en los antecedentes teóricos del proyecto, específicamente en la sección 2.1.3, el modelamiento de una geometría de adquisición compresiva se ha formulado en la literatura mediante un sistema lineal de la forma

$$\mathbf{y} = \tilde{\Phi} \mathbf{x}, \quad (4)$$

donde el vector  $\mathbf{y}$  representa los datos capturados en forma vectorizada; la matriz  $\tilde{\Phi}$  corresponde a la matriz de adquisición, la cual es una matriz binaria y diagonal cuyos elementos codifican con 0 los valores no medidos en la adquisición, y con 1 los que sí fueron obtenidos; y  $\mathbf{x}$  corresponde a la vectorización de los datos originales, la cual es la variable objetivo que se busca reconstruir con el método.

**3.1.2. Modelo en notación tensorial.** Este trabajo de grado pretende emplear las correlaciones presentes en las diferentes dimensiones del dato sísmico para mejorar la reconstrucción, por lo tanto, se optó por plantear el modelo de adquisición utilizando una notación tensorial en la que estas dimensiones son preservadas. Así, el modelo de adquisición tensorial estaría definido por el sistema lineal de la forma

$$\mathcal{Y} = \mathcal{X} \times_3 \Phi, \quad (5)$$

donde  $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  corresponde al tensor de datos obtenido, la matriz de adquisición  $\Phi \in \{0, 1\}^{I_3 \times I_3}$  es una matriz diagonal binaria que codifica con 1 los disparos obtenidos y con 0 los faltantes, y el tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  corresponde al tensor de datos completo que se buscará reconstruir.

Cabe resaltar que, aunque se optó por utilizar notación tensorial en el documento para facilitar el uso de descomposiciones en términos de las dimensiones del tensor, los modelos propuestos en las ecuaciones (4) y (5) son completamente análogos. Es posible establecer las relaciones lineales

$$\mathbf{y} = \text{vect}(\mathcal{Y}), \quad \mathbf{x} = \text{vect}(\mathcal{X}), \quad \tilde{\Phi} = \Phi \otimes \mathbf{I}_{I_1 I_2}, \quad (6)$$

entre las variables de ambas formulaciones.

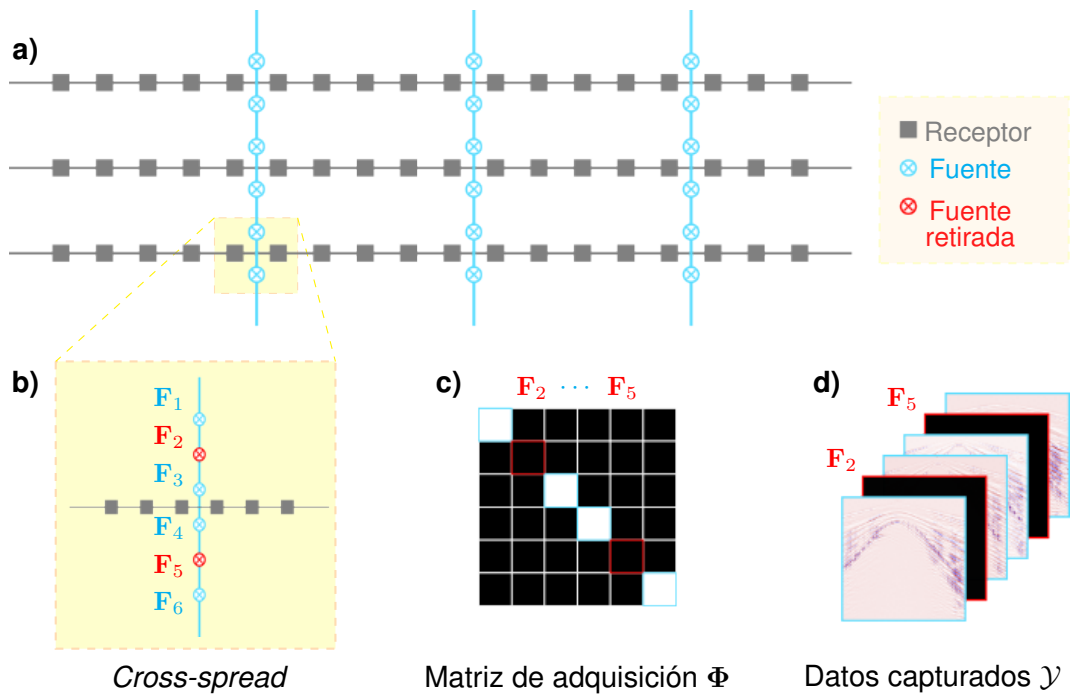


Figura 6. Modelo de observación para una geometría ortogonal compresiva. a) Geometría ortogonal con líneas paralelas de fuentes perpendiculares a líneas de receptores paralelas. b) *Cross-spread* submuestreado en fuentes. c) Matriz de adquisición  $\Phi$ . d) Datos capturados con fuentes faltantes.

En la Figura 6 se esquematiza el modelo de observación para una geometría ortogonal submuestreada en fuentes, ilustrando la relación entre la geometría, la matriz de submuestreo y los datos capturados.

### 3.2. ANÁLISIS DE LA PROPIEDAD DE BAJO RANGO EN DATOS SÍSMICOS

Considerando que los datos sísmicos son medidos sobre superficies adyacentes, en este trabajo se asume que los datos presentan redundancias estructurales entre las diferentes secciones del tensor de datos sísmicos. Esta sección busca validar la suposición de bajo rango en datos sísmicos a través de la medición de las correlaciones entre las diferentes secciones del *cross-spread*. Si las correlaciones entre las secciones del cubo sísmico son altas, la suposición de bajo rango puede ser

validada y los modos se pueden aproximar con una estructura de bajo rango.

Para esta verificación se hizo uso del concepto de rango tensorial en el modo  $m$ , definido como el rango usual del modo  $m$  del tensor. Utilizando una descomposición en valores singulares (SVD), el modo  $\mathbf{X}_{(m)}$  puede ser factorizado como

$$\mathbf{X}_{(m)} = \mathbf{U}_m \mathbf{\Sigma}_m \mathbf{V}_m, \quad (7)$$

donde  $\mathbf{\Sigma}_m \in \mathbb{R}^{I_m \times I_m}$  define una matriz diagonal que contiene los valores singulares de  $\mathbf{X}_{(m)}$  ordenados de forma descendente, y  $\mathbf{U}_m \in \mathbb{R}^{I_m \times I_m}$  y  $\mathbf{V}_m \in \mathbb{R}^{I_m \times I_m}$  son las bases ortonormales para el espacio fila y columna de la matriz.

El valor singular  $\sigma_i$  define la cantidad de información que aportan los vectores  $\mathbf{u}_i$  y  $\mathbf{v}_i$  a la estructura de la matriz. Por lo tanto, cuando se ve un decaimiento rápido en los valores singulares implica que la mayor parte de la información estructural de la matriz se encuentra en los primeros vectores de la base, por lo que se presenta una estructura con altas correlaciones.

La evaluación de la propiedad de bajo rango en datos sísmicos se realizó calculando la SVD sobre dos bases de datos de adquisiciones sísmicas 3D: i) 72 *cross-spread* sintético-realistas tomados del proyecto "*SEAM phase II-Land*"<sup>24</sup>, y ii) un *cross-spread* tomado de la base de datos "*Stratton 3D survey*"<sup>25</sup> provenientes de una adquisición terrestre real (Ver Sección 4.1).

En la Figura 7 se presentan las caídas de SVD para la dimensión temporal, de receptores y de fuentes. Como se esperaba, los valores singulares presentan un rápido decaimiento, disminuyendo los últimos valores a aproximadamente cero. Así

---

<sup>24</sup> <https://seg.org/News-Resources/Research-Data/Land-Seismic-Data>. EL proyecto 110287780575 cuenta con el permiso de la *Society of Exploration Geophysicists* (SEG) para el uso de esta base de datos en el contexto académico.

<sup>25</sup> [https://wiki.seg.org/wiki/Stratton\\_3D\\_survey#How\\_to\\_download](https://wiki.seg.org/wiki/Stratton_3D_survey#How_to_download). Esta base de datos es de dominio público.

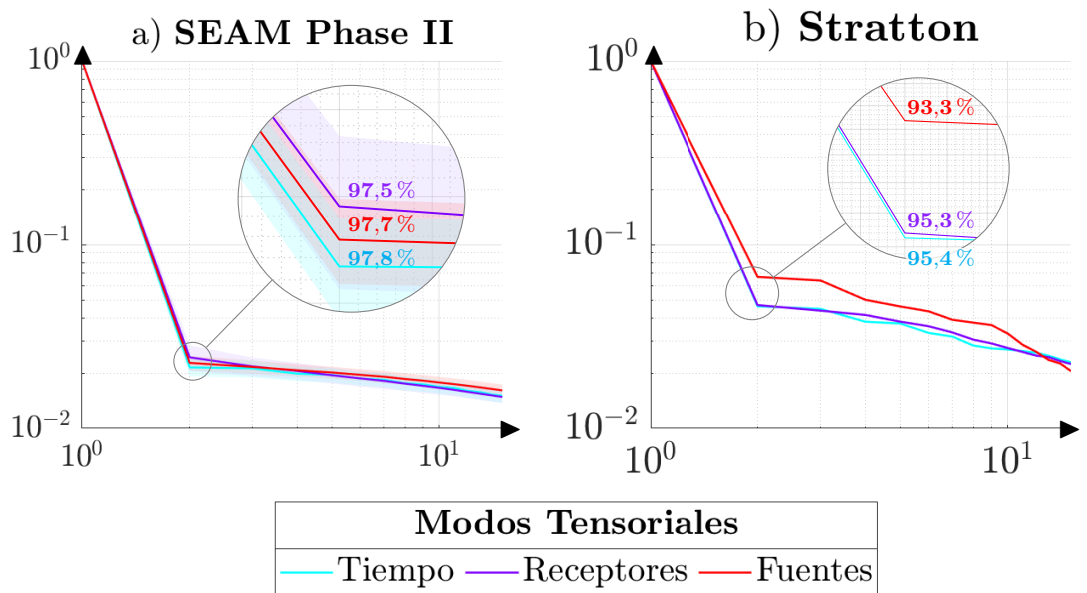


Figura 7. Propiedad de bajo rango para los datos sísmicos. La caída de SVD es estimada para el modo temporal, de receptores y fuentes sobre las dos bases de datos. La amplitud de los valores singulares se encuentra normalizada y se presenta en escala logarítmica. El porcentaje de decaimiento entre el primer y el segundo valor singular se muestra para cada modo sobre el acercamiento realizado en las caídas. a) El decaimiento de la base de datos del SEAM phase II muestra el promedio y desviación de 72 *cross-spread* presentes en la adquisición. b) La caída de Stratton muestra la caída de valores singulares para un único *cross-spread*.

mismo, se puede ver que en promedio entre el primer y el segundo valor singular se presenta una tasa de decaimiento del 95%. Este rápido decaimiento establece que las matrices  $\mathbf{X}_{(m)}$  pueden ser aproximadas de forma precisa por medio de una matriz de bajo rango con alta probabilidad, lo que indica que las secciones en el tensor sísmico se encuentran fuertemente correlacionadas, y que matemáticamente yacen en un subespacio de baja dimensionalidad.

### 3.3. PROBLEMA DE OPTIMIZACIÓN CONVEXA PARA LA RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES

Debido a que se encontraron altas correlaciones en cada una de las dimensiones del tensor de datos sísmicos, en el algoritmo propuesto se hace uso de una descomposición tensorial de cada modo para introducir la propiedad de bajo rango analizada en la sección 3.2.

**3.3.1. Factorización tensorial sobre el modo  $m$ .** Sea  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  un tensor cuyas secciones de cierto modo yacen en un subespacio de baja dimensionalidad,  $\mathcal{X}$  puede ser factorizado como

$$\mathcal{X} = \mathcal{Z}_m \times_m \mathbf{P}_m, \quad (8)$$

donde  $\mathbf{P}_m \in \mathbb{R}^{I_m \times r_m}$  es una matriz de proyección que expande la  $m$ -ésima dimensión del tensor, y  $\mathcal{Z}_m$  un tensor que expande las dimensiones restantes.

Las secciones sobre el modo  $m$  del tensor  $\mathcal{Z}_m$  definido en la factorización (8) puede preservar ciertas propiedades de acuerdo a la elección de la matriz  $\mathbf{P}_m$ . De esta forma, la descomposición puede ser usada para incorporar información previa presente en las secciones del modo  $m$  en un problema de optimización. Específicamente, en este trabajo de grado se utilizó el análisis de componentes principales (PCA) para el cálculo de la matriz  $\mathbf{P}_m$  de tal forma que las secciones sobre  $\mathcal{Z}_m$  coincidan estructuralmente con sus secciones análogas en  $\mathcal{X}$ , como se describe a continuación.

**3.3.1.1. Cálculo de la matriz de proyección  $\mathbf{P}_m$ .** Sea  $\mathbf{C}_m = \mathbf{X}_{(m)}\mathbf{X}_{(m)}^T \in \mathbb{R}^{I_m \times I_m}$  la matriz de correlaciones sobre el modo  $m$ , se define su descomposición en valores propios como

$$\mathbf{C}_m = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \quad (9)$$

donde  $\Lambda \in \mathbb{R}^{I_m \times I_m}$  y  $\mathbf{Q} \in \mathbb{R}^{I_m \times I_m}$  corresponde con la matriz de valores propios y vectores propios de la matriz  $\mathbf{C}_m$ , respectivamente. Entonces, la matriz  $\mathbf{P}_m$  es calculada obteniendo las primeras  $r_m$  columnas de la matriz  $\mathbf{Q}$

$$\mathbf{P}_m = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{r_m}], \quad (10)$$

de tal forma que las secciones en  $\mathcal{Z}_m$  resulten ser las  $r_m$  componentes principales más significativas del tensor original  $\mathcal{X}$ . La elección de esta descomposición se hizo considerando que permite ahorrar costo computacional para valores pequeños del rango tensorial  $r_m$ , y que permite mantener las relaciones estructurales en las componentes de  $\mathcal{Z}_m$ .

**3.3.2. Formulación del problema de optimización.** Considerando la propiedad de bajo rango analizada en la sección 3.2, el problema de recuperación de datos sísmicos dado un tensor de datos submuestreados  $\mathcal{Y}$  bajo una matriz de adquisición  $\Phi$  está dada por el problema de optimización

$$\underset{\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}}{\text{minimizar}} \left\{ \frac{1}{2} \|\mathcal{Y} - \mathcal{X} \times_3 \Phi\|_2^2 + \mu \psi(\mathcal{X}) \right\}. \quad (11)$$

En (11), el primer término se conoce como término de fidelidad de los datos, y basado en el modelo de observación definido en (5) busca que los disparos reconstruidos guarden la mayor similitud con los submuestreados de la adquisición; y el segundo término corresponde a un regularizador aplicado al problema de reconstrucción para mejorar el condicionamiento del problema, incorporando la propiedad de bajo rango; el parámetro  $\mu > 0$  balancea la importancia del regularizador.

En esta formulación se va a incorporar el principio de bajo rango de forma global, entre los modos del tensor como fue analizado en la sección 3.2, en conjunto con el

bajo rango no-local utilizado en trabajos como el propuesto por Chen<sup>26</sup>.

El principio de bajo rango no-local puede ser incorporado al definir el operador  $\psi$  de forma explícita siguiendo el esquema de regularizador no-local que se muestra en el trabajo de Wai He <sup>27</sup>. Para ello, se define el operador  $\psi$  en función de normas nucleares

$$\psi(\mathcal{X}) = \sum_j \|[\mathcal{X}]_{\mathcal{G}^j}\|_*, \quad (12)$$

donde  $[\mathcal{X}]_{\mathcal{G}^j}$  es una matriz que concatena los parches definidos por un grupo no-local predefinido  $\mathcal{G}^j$ .

Para incorporar el bajo rango de forma global se hace uso de la factorización mostrada en (8), redefiniendo la regularización de bajo rango en (12) como

$$\mu\psi(\mathcal{X}) = \mu_1\psi(\mathcal{Z}_1 \times_1 \mathbf{P}_1) + \mu_2\psi(\mathcal{Z}_2 \times_2 \mathbf{P}_2), \quad (13)$$

donde  $\mathcal{Z}_1 \in \mathbb{R}^{r_1 \times I_2 \times I_3}$  y  $\mathcal{Z}_2 \in \mathbb{R}^{I_1 \times r_2 \times I_3}$  son tensores que apilan las componentes principales de las secciones temporales y de receptores, respectivamente;  $\mathbf{P}_1$  y  $\mathbf{P}_2$  son matrices de proyección calculadas como se plantea en (10); y los parámetros  $\mu_1, \mu_2 > 0$  son parámetros de regularización sobre la penalización del bajo rango tal que  $\mu_1 + \mu_2 = \mu$ .

Observe que en la ecuación (13) no se utilizó una factorización sobre el tercer modo del tensor. Esto se debe a que la tercera dimensión del tensor esta asociada a las fuentes, y como en la geometría compresiva se eliminan fuentes de la adquisición, la matriz de proyección  $\mathbf{P}_3$  sobre el tensor de muestras  $\mathcal{Y}$  resultaría deficiente en rango, con filas y columnas nulas.

---

<sup>26</sup> Siyuan Chen y col. "Post-stack seismic data interpolation using a fast non-local similarity matching algorithm". En: *Studia Geophysica et Geodaetica* 65.1 (2021), págs. 59-69.

<sup>27</sup> Wei He y col. "Non-local meets global: An iterative paradigm for hyperspectral image restoration". En: *arXiv preprint arXiv:2010.12921* (2020).

Al incorporar la separación de los modos (13) en el problema de optimización original (11), el problema de optimización resultante estaría dado por

$$\begin{aligned} & \underset{\mathcal{X}, \mathcal{Z}_1, \mathcal{Z}_2}{\text{minimizar}} \quad \left\{ \frac{1}{2} \|\mathcal{Y} - \mathcal{X} \times_3 \Phi\|_2^2 + \mu_1 \psi(\mathcal{Z}_1) + \mu_2 \psi(\mathcal{Z}_2) \right\}, \\ & \text{sujeto a} \quad \mathcal{X} = \mathcal{Z}_1 \times_1 \mathbf{P}_1; \quad \mathcal{X} = \mathcal{Z}_2 \times_2 \mathbf{P}_2. \end{aligned} \quad (14)$$

Observe que en la función objetivo del problema escrito en (14), los operadores  $\psi$  que promueven el bajo rango no-local son aplicados sobre los componentes principales  $\mathcal{Z}_\ell$ , y no sobre el producto tensorial. Esta simplificación se realiza tomando ventaja de las correlaciones estructurales del tensor  $\mathcal{X}$  que se mantienen en sus componentes principales bajo la matriz de proyección propuesta, promoviendo el bajo rango tanto a nivel no-local como global.

### 3.4. ALGORITMO NUMÉRICO DE RECONSTRUCCIÓN DE DATOS SÍSMICOS FALTANTES

Para desarrollar el algoritmo de reconstrucción de datos sísmicos faltantes, se parte del problema de optimización convexa que se formuló en la sección 3.1. De este modo, cuando se encuentre un tensor  $\mathcal{X}$  que minimice la función objetivo a la vez que cumpla las restricciones impuestas en (14), ese tensor será el *cross-spread* con las fuentes faltantes reconstruidas.

El algoritmo que se utilizó como base para resolver el problema de optimización planteado fue el método de multiplicadores en dirección alternada (ADMM)<sup>28</sup> junto con el esquema de *Plug-and-play* (PnP)<sup>29</sup> para resolver algunos subproblemas.

---

<sup>28</sup> Stephen Boyd y col. "Distributed optimization and statistical learning via the alternating direction method of multipliers". En: *Foundations and Trends® in Machine learning* 3.1 (2011), págs. 1-122.

<sup>29</sup> Stanley H Chan, Xiran Wang y Omar A Elgendy. "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications". En: *IEEE Transactions on Computational Imaging* 3.1

Debido a que el término de fidelidad y los términos de regularización están en función de variables diferentes, y a su vez estos se encuentran relacionados con aplicaciones lineales (el producto tensorial en los modos 1 y 2), este algoritmo es aplicable para este problema.

Para aplicar el método ADMM, es necesario el cálculo del lagrangiano aumentado del problema. Para facilitar el desarrollo del método, se introduce el lagrangiano aumentado en forma escalada como sigue

$$\begin{aligned} \mathcal{L}_\rho(\mathcal{X}, \mathcal{Z}_\ell, \mathcal{W}_\ell) = & \frac{1}{2} \|\mathcal{Y} - \mathcal{X} \times_3 \Phi\|_2^2 + \mu_1 \psi(\mathcal{Z}_1) + \mu_2 \psi(\mathcal{Z}_2) \\ & + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Z}_1 \times_1 \mathbf{P}_1 + \mathcal{W}_1\|_2^2 + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Z}_2 \times_2 \mathbf{P}_2 + \mathcal{W}_2\|_2^2, \end{aligned} \quad (15)$$

donde  $\mathcal{W}_\ell \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , para  $\ell = 1, 2$ , son las variables duales del problema que buscan equilibrar el espacio de búsqueda de las variables originales para que estas terminen en el espacio factible del problema, y  $\rho > 0$  es el parámetro de paso dual que regula que tanta importancia se le da a la restricción dual.

El método consiste en actualizar iterativamente las variables objetivo del problema, resolviendo un subproblema de optimización asociado a la función lagrangiana resuelta para cada variable en particular. Los pasos de actualización de las variables en el algoritmo propuesto estarían dados por:

$$\mathcal{X}^{t+1} \in \arg \min_{\mathcal{X}} \mathcal{L}_\rho(\mathcal{X}, \mathcal{Z}_\ell^t, \mathcal{W}_\ell^t), \quad (16)$$

$$\mathcal{Z}_\ell^{t+1} \in \arg \min_{\mathcal{Z}_\ell} \mathcal{L}_\rho(\mathcal{X}^{t+1}, \mathcal{Z}_\ell, \mathcal{W}_\ell^t), \quad (17)$$

$$\mathcal{W}_\ell^{t+1} = \mathcal{W}_\ell^t + \mathcal{X}^{t+1} - \mathcal{Z}_\ell^{t+1} \times_\ell \mathbf{P}_\ell. \quad (18)$$

---

(2016), págs. 84-98.

A continuación, para cada una de las variables de optimización, se resuelve el sub-problema asociado de forma explícita llegando a una solución numérica.

**3.4.1. Paso de actualización de  $\mathcal{X}$ .** El problema de optimización asociado a la actualización de esta variable está dado por la suma de tres normas  $\ell_2$ ,

$$\begin{aligned} \underset{\mathcal{X}}{\text{minimizar}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{X} \times_3 \Phi\|_2^2 + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{W}_1^t\|_2^2 \\ & + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Z}_2^t \times_2 \mathbf{P}_2 + \mathcal{W}_2^t\|_2^2. \end{aligned} \quad (19)$$

Dado que la norma  $\ell_2$  como fue definida en la Tabla 1 es una función derivable, es posible encontrar una solución cerrada de este problema de optimización al igualar la derivada a cero y obtener una solución para  $\mathcal{X}$ .

La derivada de la función objetivo para (19) estaría dada por

$$\begin{aligned} \frac{\partial}{\partial \mathcal{X}} (\mathcal{L}_\rho(\mathcal{X}, \mathcal{Z}_\ell^t, \mathcal{W}_\ell^t)) = & (\mathcal{X} \times_3 \Phi - \mathcal{Y}) \times_3 \Phi^T + \rho(\mathcal{X} - \mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{W}_1^t) \\ & + \rho(\mathcal{X} - \mathcal{Z}_2^t \times_2 \mathbf{P}_2 + \mathcal{W}_2^t). \end{aligned} \quad (20)$$

Al igualar (20) a cero, se obtiene el siguiente sistema lineal

$$\begin{aligned} (\mathcal{X} \times_3 \Phi - \mathcal{Y}) \times_3 \Phi^T + \rho(\mathcal{X} - \mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{W}_1^t) + \rho(\mathcal{X} - \mathcal{Z}_2^t \times_2 \mathbf{P}_2 + \mathcal{W}_2^t) &= 0 \\ \mathcal{X} \times_3 \Phi \times_3 \Phi^T + \rho\mathcal{X} + \rho\mathcal{X} &= \mathcal{Y} \times_3 \Phi^T + \rho(\mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{Z}_2^t \times_2 \mathbf{P}_2 - \mathcal{W}_1^t - \mathcal{W}_2^t) \\ \mathcal{X} \times_3 \Phi^T \Phi + 2\rho\mathcal{X} &= \mathcal{Y} \times_3 \Phi^T + \rho(\mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{Z}_2^t \times_2 \mathbf{P}_2 - \mathcal{W}_1^t - \mathcal{W}_2^t) \\ \mathcal{X} \times_3 \Phi + 2\rho\mathcal{X} &= \mathcal{Y} + \rho(\mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{Z}_2^t \times_2 \mathbf{P}_2 - \mathcal{W}_1^t - \mathcal{W}_2^t) \\ \mathcal{X} \times_3 (\Phi + 2\mathbf{I}) &= \mathcal{Y} + \rho(\mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{Z}_2^t \times_2 \mathbf{P}_2 - \mathcal{W}_1^t - \mathcal{W}_2^t). \end{aligned} \quad (21)$$

La solución para el sistema lineal, y por lo tanto el paso de actualización para  $\mathcal{X}$ , estaría dado por

$$\hat{\mathcal{X}} = [\mathcal{Y} + \rho(\mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{Z}_2^t \times_2 \mathbf{P}_2 - \mathcal{W}_1^t - \mathcal{W}_2^t)] \times_3 (\Phi + 2\rho\mathbf{I})^{-1}. \quad (22)$$

Sobre el procedimiento realizado en (21) sería importante realizar las siguientes aclaraciones:

1. Como la derivada (20) está en función de un tensor con la misma dimensión de  $\mathcal{X}$ , al igualar esta derivada a cero, el  $\theta$  hace referencia a un tensor de dimensión  $I_1 \times I_2 \times I_3$  con un valor de 0 en todas las posiciones.
2. La matriz  $\mathbf{I}$  es la matriz identidad de orden  $I_3 \times I_3$ .
3. Dado que  $\Phi$  es una matriz diagonal binaria, es fácil demostrar que es una matriz simétrica e idempotente. De estas dos propiedades se puede obtener la siguiente relación:

$$\Phi^T \Phi = \Phi \Phi = \Phi^2 = \Phi$$

Este resultado fue usado en (21) para simplificar uno de los coeficientes resultantes de  $\mathbf{X}$ .

4. Por la naturaleza de los datos capturados  $\mathcal{Y}$  definidos en el modelo de observación, y la simetría de  $\Phi$ , se puede obtener la relación:

$$\mathcal{Y} \times_3 \Phi^T = \mathcal{Y} \times_3 \Phi = \mathcal{Y}$$

Este resultado fue usado en (21) para simplificar uno de términos independientes.

Del paso de actualización obtenido en (22) sería importante observar que la matriz  $(\Phi + 2\mathbf{I})$  es una matriz positiva definida, por lo tanto es siempre invertible. Además, al ser una matriz diagonal su inversa se puede computar de forma fácil al encontrar el inverso multiplicativo de cada entrada de la diagonal.

Note que el paso de actualización de  $\mathcal{X}$  dentro del algoritmo general tiene la función de asegurar la fidelidad de los datos. Es decir, que los disparos que ya se conocían de antemano en la captura  $\mathcal{Y}$  se mantengan dentro de la reconstrucción.

**3.4.2. Paso de actualización de  $\mathcal{Z}_1$ .** El problema de optimización relacionado con  $\mathcal{Z}_1$  estaría dado por el problema relacionado con la función  $\psi$

$$\underset{\mathcal{Z}_1}{\text{minimizar}} \quad \mu_1 \psi(\mathcal{Z}_1) + \frac{\rho}{2} \|\mathcal{X}^{t+1} - \mathcal{Z}_1 \times_1 \mathbf{P}_1 + \mathcal{W}_1^t\|_2^2. \quad (23)$$

Dado que la norma  $\ell_2$  sobre tensores cumple las propiedades de una norma, se tiene la propiedad  $\|\alpha \mathcal{X}\|_2^2 = |\alpha|^2 \|\mathcal{X}\|_2^2$  para cualquier escalar  $\alpha$ , así como la invarianza a la rotación. Por lo tanto, el problema (23) podría ser reformulado como

$$\underset{\mathcal{Z}_1}{\text{minimizar}} \quad \mu_1 \psi(\mathcal{Z}_1) + \frac{\rho}{2} \|\mathcal{Z}_1 - \tilde{\mathcal{Z}}_1^t\|_2^2, \quad (24)$$

utilizando la sustitución  $\tilde{\mathcal{Z}}_1^t = (\mathcal{X}^{t+1} + \mathcal{W}_1^t) \times_1 \mathbf{P}_1^T$ .

Recordando que  $\mathcal{Z}_1$  define las componentes principales para la dimensión temporal, mantiene las correlaciones estructurales entre las secciones temporales. Por lo tanto, es posible llegar a una solución de (24) separando el problema para las secciones temporales de  $\mathcal{Z}_1$ .

Como el operador  $\psi$  fue definido como un operador sobre parches predefinidos, al partir de la asunción que estos parches se realizan sobre las secciones temporales de la entrada, se podría llegar a

$$\psi(\mathcal{Z}_1) = \sum_{i=1}^{r_1} \psi_i([\mathbf{Z}_1]_{i:}), \quad (25)$$

donde  $[\mathbf{Z}_1]_{i:}$  corresponde con la  $i$ -ésima sección de tiempo del tensor  $\mathcal{Z}_1$ .

Teniendo en cuenta la definición del operador  $\psi$  en (25) y que la norma  $\ell_2$  puede ser separada en término de las secciones del tensor de entrada, es posible separar (24) en subproblemas de la forma

$$\underset{[\mathbf{Z}_1]_{i:}}{\text{minimizar}} \quad \frac{1}{2\sigma_1^2} \|[ \mathbf{Z}_1 ]_{i:} - [\tilde{\mathbf{Z}}_1^t]_{i:}\|_F^2 + \psi_i([\mathbf{Z}_1]_{i:}) \quad (26)$$

donde  $[\tilde{\mathbf{Z}}_1^{t+1}]_{i::}$  son las secciones temporales del tensor  $\tilde{\mathbf{Z}}_1^t$  y se define el parámetro  $\sigma_1 = \sqrt{\mu_1/\rho}$ .

El problema de optimización (26) tiene la forma estándar de un operador proximal para  $\psi_i$ . Por lo tanto, utilizando un esquema de PnP, este paso de actualización puede ser resuelto aplicando algún algoritmo iterativo sobre  $[\tilde{\mathbf{Z}}_1^t]_{i::}$ . Como el operador  $\psi_i$  hace una penalización sobre parches de la sección temporal, en este trabajo se escogió el algoritmo de ajuste de bloques y filtrado 3D (BM3D) debido a que es un algoritmo de eliminación de ruido basado en filtrado por parches similares.

Entonces, el paso de actualización de  $\mathcal{Z}_1$  se podría escribir como:

$$\hat{\mathcal{Z}}_1 = \llbracket \text{BM3D}_{\sigma_1}([\tilde{\mathbf{Z}}_1^t]_{1::}), \dots, \text{BM3D}_{\sigma_1}([\tilde{\mathbf{Z}}_1^t]_{i::}), \dots, \text{BM3D}_{\sigma_1}([\tilde{\mathbf{Z}}_1^t]_{r_1::}) \rrbracket. \quad (27)$$

Note que este paso de actualización tiene el objetivo de promover el bajo rango en el dominio temporal, así como promover el bajo rango no-local dentro del espacio de receptores y fuentes.

**3.4.3. Paso de actualización de  $\mathcal{Z}_2$ .** El problema de optimización relacionado con  $\mathcal{Z}_2$  estaría dado por el problema relacionado con el operador no-local  $\psi$

$$\underset{\mathcal{Z}_2}{\text{minimizar}} \quad \mu_2 \psi(\mathcal{Z}_2) + \frac{\rho}{2} \|\mathcal{X}^{t+1} - \mathcal{Z}_2 \times_2 \mathbf{P}_2 + \mathcal{W}_2^t\|_2^2. \quad (28)$$

Debido a la similitud entre (23) y (28), es posible solucionar el problema relacionado a  $\mathcal{Z}_2$  con una estrategia similar a la empleada para  $\mathcal{Z}_1$ . Sin embargo, como las secciones que se mantienen correlaciones de  $\mathcal{Z}_2$  son las secciones de receptores, la separación del problema debe ser en función de estas secciones.

Para realizar la separación, se asumirá que el operador  $\psi$  es aplicado separando parches sobre grupos definidos en las secciones de receptores. De esta forma, la función podría ser separada como:

$$\psi(\mathcal{Z}_2) = \sum_{j=1}^{r_2} \psi^j([\mathbf{Z}_2]_{:j:}), \quad (29)$$

donde  $[\mathbf{Z}_j]_{j:}$  corresponde con la  $j$ -ésima sección de tiempo del tensor  $\mathcal{Z}_2$ .

Por consiguiente, el problema de optimización descrito en (28) puede ser separado en término de las secciones de receptores mediante subproblemas de la forma

$$\underset{[\mathbf{Z}_2]_{:j:}}{\text{minimizar}} \quad \frac{1}{2\sigma_2^2} \|[\mathbf{Z}_2]_{:j:} - [\tilde{\mathbf{Z}}_2^t]_{:j:}\|_F^2 + \psi^j([\mathbf{Z}_2]_{:j:}), \quad (30)$$

con parámetro  $\sigma_2 = \sqrt{\mu_2/\rho}$ , donde  $[\tilde{\mathbf{Z}}_2^{t+1}]_{:j:}$  son las secciones de receptores del tensor  $\tilde{\mathcal{Z}}_2^t = (\mathcal{X}^{t+1} + \mathcal{W}_2^t) \times_2 \mathbf{P}_2^T$ .

De forma análoga, bajo un esquema de PnP, se puede utilizar un algoritmo BM3D para encontrar una solución cerrada del problema definido en (30). De forma general, el paso de actualización para  $\mathcal{Z}_2$  estaría dado por

$$\hat{\mathcal{Z}}_2 = \llbracket \underset{\sigma_2}{\text{BM3D}}([\tilde{\mathbf{Z}}_2^t]_{:1:}), \dots, \underset{\sigma_2}{\text{BM3D}}([\tilde{\mathbf{Z}}_2^t]_{:j:}), \dots, \underset{\sigma_2}{\text{BM3D}}([\tilde{\mathbf{Z}}_2^t]_{:r_2:}) \rrbracket^{T_{1,2}}. \quad (31)$$

Note que este paso de actualización tiene el objetivo de promover el bajo rango en la dimensión de los receptores, así como promover el bajo rango no-local dentro del espacio temporal y de fuentes.

**3.4.4. Pseudocódigo del algoritmo.** En el algoritmo 1 se resume el pseudocódigo del algoritmo de reconstrucción obtenido después de resolver todos los pasos de actualización resultantes del ADMM. Las matrices de proyección  $\mathbf{P}_1$  y  $\mathbf{P}_2$  son calculadas mediante las matrices de correlación  $\mathbf{C}_1$  y  $\mathbf{C}_2$  siguiendo (10).

La variable  $\mathcal{X}$  se inicializa con la captura  $\mathcal{Y}$ . Las variables  $\mathcal{Z}_\ell$  se inicializan utilizando la transpuesta de la matriz de proyección mediante el producto  $\mathcal{Y} \times_\ell \mathbf{P}_\ell^T$ , mientras que las variables duales  $\mathcal{W}_\ell$  se inicializan nulas.

Los pasos de actualización para  $\mathcal{X}$ ,  $\mathcal{Z}_1$  y  $\mathcal{Z}_2$  se resuelven con las soluciones encontradas en (22), (27) y (31), respectivamente. Las variables duales  $\mathcal{W}_\ell$  se actualizan bajo el esquema del ADMM mostrado en (18).

El algoritmo cuenta con 5 parámetros:

- $\mu_1$  y  $\mu_2$  son los parámetros que penalizan el bajo rango no-local relacionado a las secciones temporales y de receptores.
- $\rho$ , el parámetro que controla el paso de actualización dual, y está relacionado al bajo rango aplicado en la dimensión temporal y de receptores.
- $r_1$  y  $r_2$  son los parámetros de rango que escogen cuantas componentes principales se utilizan en  $\mathcal{Z}_1$  y  $\mathcal{Z}_2$ .

A continuación se presentan los análisis de convergencia y de complejidad computacional realizados sobre el algoritmo 1 de recuperación de datos sísmicos faltantes.

**3.4.4.1. Análisis de convergencia.** El algoritmo 1 es una instancia del algoritmo ADMM, cuya convergencia está garantizada cuando cada uno de los subproblemas relacionados a la función lagrangiana tiene una solución teórica<sup>28</sup>. Observe que la actualización de  $\mathcal{X}$  involucra una solución algebraica cerrada mediante la derivada de la función objetivo. Sin embargo, los pasos de actualización de  $\mathcal{Z}_\ell$  se realizan al aplicar el BM3D de forma multi-canal sobre las componentes principales de la dimensión temporal y de receptores. Teóricamente, no existe ninguna garantía de que esta solución ayude a disminuir la función de costo en cada iteración. Sin embargo, de forma experimental se observó que la función objetivo seguía disminuyendo en

---

**Algoritmo 1:** Algoritmo de reconstrucción de datos sísmicos faltantes

---

**Entrada:**  $\mathcal{Y}, \Phi, r_1, r_2, \mu_1, \mu_2, \rho, \text{iter}$ **Salida :**  $\mathcal{X}^{\text{iter}}$ 

```
1  $\mathbf{P}_1 \leftarrow \text{PC}(\mathbf{Y}_{(1)}, r_1), \quad \mathbf{P}_2 \leftarrow \text{PC}(\mathbf{Y}_{(2)}, r_2)$ 
2  $t \leftarrow 0, \mathcal{X}^0 \leftarrow \mathcal{Y}, \mathcal{Z}_\ell^0 \leftarrow \mathcal{Y} \times_\ell \mathbf{P}_\ell^T, \mathcal{W}_\ell^0 \leftarrow \mathbf{0}$ 
3 mientras  $t < \text{iter}$  hacer
    // Paso de actualización de  $\mathcal{X}$ 
4    $\mathcal{S}^t \leftarrow \mathcal{Z}_1^t \times_1 \mathbf{P}_1 + \mathcal{Z}_2^t \times_2 \mathbf{P}_2 - \mathcal{W}_1^t - \mathcal{W}_2^t$ 
5    $\mathcal{X}^{t+1} \leftarrow (\mathcal{Y} + \rho \mathcal{S}^t) \times_3 (\Phi + 2\rho \mathbf{I})^{-1}$ 
    // Paso de actualización de  $\mathcal{Z}_1$ 
6    $\tilde{\mathcal{Z}}_1^t \leftarrow (\mathcal{X}^{t+1} + \mathcal{W}_1^{t+1}) \times_1 \mathbf{P}_1^T$ 
7    $\mathcal{Z}_1^{t+1} \leftarrow [\text{BM3D}_{\sigma_1}([\tilde{\mathcal{Z}}_1^t]_{1:}), \dots, \text{BM3D}_{\sigma_1}([\tilde{\mathcal{Z}}_1^t]_{i:}), \dots, \text{BM3D}_{\sigma_1}([\tilde{\mathcal{Z}}_1^t]_{r_1:})]$ 
    // Paso de actualización de  $\mathcal{Z}_2$ 
8    $\tilde{\mathcal{Z}}_2^t \leftarrow (\mathcal{X}^{t+1} + \mathcal{W}_2^{t+1}) \times_2 \mathbf{P}_2^T$ 
9    $\mathcal{Z}_2^{t+1} \leftarrow [\text{BM3D}_{\sigma_2}([\tilde{\mathcal{Z}}_2^t]_{1:}), \dots, \text{BM3D}_{\sigma_2}([\tilde{\mathcal{Z}}_2^t]_{j:}), \dots, \text{BM3D}_{\sigma_2}([\tilde{\mathcal{Z}}_2^t]_{r_2:})]^{T_{1,2}}$ 
    // Pasos de actualización de  $\mathcal{W}_\ell$ 
10   $\mathcal{W}_1^{t+1} \leftarrow \mathcal{W}_1^t + \mathcal{X}^{t+1} - \mathcal{Z}_1^{t+1} \times_1 \mathbf{P}_1$ 
11   $\mathcal{W}_2^{t+1} \leftarrow \mathcal{W}_2^t + \mathcal{X}^{t+1} - \mathcal{Z}_2^{t+1} \times_2 \mathbf{P}_2$ 
    // Actualizando el contador
12   $t \leftarrow t + 1$ 
13 fin
```

---

cada iteración aplicando el algoritmo BM3D multi-canal. Por lo tanto, se puede concluir que la convergencia del algoritmo no está comprometida en términos prácticos.

**3.4.4.2. Análisis de complejidad computacional.** En cada iteración del algoritmo 1, la complejidad computacional se encuentra determinada por dos operaciones principalmente:

1. La complejidad computacional del producto modo definido en (8), la cual se encuentra definida en el orden  $\mathcal{O}(I_1 I_2 I_3 r_m)$ . Este resultado se obtiene de la complejidad del producto matricial debido a que el producto en el modo  $m$

puede ser redefinido como un producto matricial.

2. La complejidad computacional de aplicar un BM3D multi-canal <sup>30</sup>, específicamente sobre un  $\mathcal{Z}_m$ , se encuentra en el orden  $\mathcal{O}(\tilde{I}_m(r_m - 1))$ .

Por lo tanto, de forma general, la complejidad del algoritmo se encontraría en el orden  $\mathcal{O}(nI_1I_2I_3 \max\{r_1, r_2, I_3\})$  donde  $n$  define el número de iteraciones empleadas en el algoritmo.

**3.4.5. Implementación computacional del algoritmo.** El algoritmo propuesto emplea técnicas de optimización convexa con aproximación de bajo rango, manejando los datos en notación tensorial (como un cubo de datos). Por lo tanto, el algoritmo 1 se implementó en el lenguaje de programación MATLAB, debido a que este lenguaje es especializado para el cómputo numérico, y se encuentra optimizado para realizar operaciones con arreglos de altas dimensiones.

Adicionalmente, considerando que para la solución de los subproblemas de las variables  $\mathcal{Z}_\ell$  se empleó una técnica PnP utilizando un algoritmo BM3D, y considerando que dicho algoritmo cuenta con una implementación nativa y una clara documentación en el lenguaje de MATLAB, resultó idóneo la elección de este lenguaje para la implementación del método propuesto.

Dentro de la implementación, se hizo uso del *Parallel Computing Toolbox* de MATLAB, una librería que permite realizar cálculos paralelos en equipos multinúcleo, GPUs y clusters. Esta librería fue utilizada para optimizar la velocidad del código al realizar los cálculos del BM3D en cada una de las componentes de forma paralela, lo cual ayudó a disminuir el tiempo de ejecución del algoritmo.

---

<sup>30</sup> Kostadin Dabov y col. "Image denoising by sparse 3-D transform-domain collaborative filtering". En: *IEEE Transactions on image processing* 16.8 (2007), págs. 2080-2095.

## 4. CONFIGURACIÓN EXPERIMENTAL

En esta sección se presenta la configuración experimental sobre las simulaciones realizadas con el fin de validar la efectividad del algoritmo desarrollado, realizando una descripción de los conjuntos de datos usados para la validación, las métricas escogidas, y los algoritmos del estado del arte para la comparación.

### 4.1. CONJUNTOS DE DATOS

Para el desarrollo de los experimentos de validación, se utilizaron dos conjuntos de datos de adquisiciones sísmicas 3D terrestres realistas (una modelada y la otra de una adquisición real). A continuación se realiza una descripción de cada uno de ellos.

1. **SEAM phase II:** este conjunto de datos es un conjunto sintético-realista modelado por el programa de modelado avanzado de la SEG (SEAM por sus siglas del inglés *SEG Advanced Modeling Program*), realizado durante el proyecto “*SEAM Phase II—Land Seismic Challenges*”. El modelado se centró en las características de las regiones montañosas, incluida la topografía pronunciada y el relleno aluvial blando en la superficie<sup>31</sup>. De la adquisición se tomó un *cross-spread* con 61 disparos, los cuales por facilidad fueron redimensionados a  $128 \times 128$  muestras temporales y de receptores.
2. **Stratton:** este conjunto de datos sísmicos 3D terrestres corresponde a datos reales del “Estudio 3D de Stratton” (“*Stratton 3D survey*”) realizado en el sur de Texas. En 2014, la Oficina de Geología Económica de la Universidad de

---

<sup>31</sup> Carl Regone y col. “Geologic model building in SEAM Phase II—Land seismic challenges”. En: *The Leading Edge* 36.9 (2017), págs. 738-749.

Texas en Austin dispuso públicamente este conjunto de datos para motivos educativos e investigativos en todo el mundo <sup>32</sup>. El *cross-spread* seleccionado cuenta con 18 disparos que fueron redimensionados a  $128 \times 128$  muestras temporales y número de receptores.

## 4.2. MÉTRICAS

La calidad de los datos reconstruidos por el algoritmo fue medida utilizando las siguientes cuatro métricas.

- **Error cuadrático medio (MSE):** esta métrica mide el promedio de los errores entre el valor original y el valor estimado por el algoritmo, elevado al cuadrado. Al ser una medida de error, una mejor reconstrucción se alcanza cuando el valor del MSE es menor. Matemáticamente se define como:

$$MSE = \frac{1}{I_1 I_2} \sum_{i=1}^{I_1} \sum_{j=1}^{I_2} (\hat{x}_{(i,j)} - x_{(i,j)})^2 \quad (32)$$

donde  $\hat{\mathcal{X}}$  es el dato estimado,  $\mathcal{X}$  el dato original.

- **Proporción máxima de señal a ruido (PSNR):** es una métrica que se utiliza para definir la relación que existe entre la máxima potencia que pueda tener una señal, y el ruido que se afecta en cierta representación. Debido a las posibles escalas que puede tener una señal, se representa en escala logarítmica, medida en decibelios [dB]. Matemáticamente, puede ser estimada como:

$$PSNR = 10 \log \left( \frac{\max(X)^2}{MSE} \right) \quad (33)$$

---

<sup>32</sup> *Stratton 3D survey*. [https://wiki.seg.org/wiki/Stratton\\_3D\\_survey](https://wiki.seg.org/wiki/Stratton_3D_survey).

Es importante resaltar que al ser una métrica que mide calidad, una mejor reconstrucción se alcanza cuando el PSNR es mayor.

- **Relación señal/ruido (SNR):** es una métrica que se utiliza para medir la distorsión de una señal que se transmite, midiendo la proporción que existe entre la señal original y el ruido. Igual que el PSNR, es medido en decibelios por el posible rango de las señales. Matemáticamente, esta métrica se puede definir como:

$$SNR = 10 \log \left( \frac{\sum_{i=1}^{I_1} \sum_{j=1}^{I_2} (\hat{x}_{(i,j)})^2}{MSE} \right) \quad (34)$$

Análogamente al PSNR, se desea obtener una mayor relación señal/ruido, debido a que esto implica que la medida de potencia de la señal es mayor a la diferencia de la reconstrucción.

- **Medida del índice de similitud estructural (SSIM):** esta métrica es utilizada tradicionalmente para medir la similitud entre dos imágenes, donde una reconstrucción es comparada con una imagen inicial sin distorsiones. La principal diferencia entre el SSIM con otras métricas como el MSE o el PSNR es que esta no busca medir errores, sino que busca estimar la similitud estructural entre dos imágenes, incluyendo fenómenos ópticos como la luminancia y el contraste. Matemáticamente, esta métrica es estimada como:

$$SSIM = \left( \frac{2\mu_x\mu_{\hat{x}} + c_1}{\mu_x^2 + \mu_{\hat{x}}^2 + c_1} \right) \left( \frac{2\sigma_x\sigma_{\hat{x}} + c_2}{\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2} \right) \left( \frac{\sigma_{x,\hat{x}} + c_3}{\sigma_x\sigma_{\hat{x}} + c_3} \right), \quad (35)$$

donde

- $\mu_x$  define la media de la imagen  $X$ .
- $\sigma_x$  define la desviación estándar de la imagen  $X$ .
- $\sigma_{x,y}$  define la covarianza entre las imágenes  $X$  y  $Y$ .

- $c_1 = (k_1 L)^2$ ,  $c_2 = (k_2 L)^2$  y  $c_3 = 0.5c_2$  son variables que corrigen posibles divisiones por cero.  $L$  define la intensidad máxima del píxel, y por defecto  $k_1 = 0.01$  y  $k_3 = 0.03$ .

Del producto definido en (35), el primer factor hace referencia a la luminancia, el segundo al contraste y el tercero a la estructura. Esta métrica es una métrica acotada entre 0 y 1, donde los valores cercanos a 1 indican una mayor similitud entre las imágenes que se están comparando.

Todas las métricas seleccionadas son métricas aplicadas a matrices (datos en dos dimensiones). Por lo tanto, para evaluar el desempeño del algoritmo estas métricas fueron aplicadas disparo a disparo, en donde el valor mostrado corresponde al promedio entre disparos.

### 4.3. ALGORITMOS DE COMPARACIÓN

Considerando que en el estado del arte pocos trabajos han abordado el problema de recuperar disparos completos provenientes de la eliminación de una fuente en la geometría de adquisición, en este trabajo se propuso realizar un marco de referencia comparativo con diferentes algoritmos de reconstrucción de datos sísmicos faltantes, tanto especializados en reconstrucciones sísmicas como métodos de recuperación de datos generales. Específicamente, se implementaron 3 algoritmos del estado del arte en reconstrucción de disparos sísmicos:

- El *Consensus Equilibrium*<sup>5</sup> (CE), el cual propone utilizar diferentes principios en conjunto para reconstruir disparos faltantes, utilizando el principio de escasez y eliminación de ruido.

- El método de reconstrucción de fuentes con inteligencia artificial propuesto por Wang<sup>33</sup> (ResNet), el cual se basa en aprendizaje interno utilizando similitudes estructurales entre las secciones de receptores y disparos sísmicos. De este trabajo se adaptó la forma en que se ingresaban los datos a la red neuronal para poder aplicar a un esquema de sub-muestreo aleatorio.
- El multichannel singular spectrum analysis<sup>34</sup> (MSSA), que realiza reconstrucción de datos sísmicos 3D utilizando un algoritmo de bajo rango sobre el dominio de Hankel. De este trabajo se modificó la máscara de adquisición para realizar reconstrucción de disparos completos.

Adicionalmente, se implementaron 6 métodos de completado de datos faltantes. Como algoritmos son algoritmos de completado de datos matriciales, para la reconstrucción sobre el cubo de datos se aplicó cada algoritmo a través de las secciones de receptores. Estos algoritmos se basan en 3 principios:

- Escasez. Este principio se basa en el hecho que una señal, bajo cierta base de transformación, concentra su información en unos pocos coeficientes. El problema de optimización involucra resolver un problema  $\ell_2-\ell_1$  donde el primer término induce la fidelidad de los datos, y el segundo, induce el término de escasez. Para resolver este problema se utilizaron los algoritmos ADMM<sup>28</sup>,

---

<sup>33</sup> Benfeng Wang y col. "Intelligent missing shots' reconstruction using the spatial reciprocity of Green's function based on deep learning". En: *IEEE Transactions on Geoscience and Remote Sensing* 58.3 (2019), págs. 1587-1597.

<sup>34</sup> Yangkang Chen y col. "An open-source Matlab code package for improved rank-reduction 3D seismic data denoising and reconstruction". En: *Computers & Geosciences* 95 (2016), págs. 59-66.

FISTA<sup>35</sup>, TWIST<sup>36</sup> y GAP<sup>37</sup>.

- Suavidad. Este principio asume que los cambios en una matriz se dan de forma suave, por lo tanto las diferencias entre una posición arbitraria y las posiciones vecinas es muy pequeña. Este principio se induce mediante un regularizador conodico como el *Total Variation*<sup>38</sup> (TV), que utiliza una norma  $\ell_1$  sobre la señal multiplicada por una matriz de diferencias para disminuir los cambios en vecindades de la matriz.
- Eliminación de ruido. Este principio asume que se puede incluir cierto algoritmo del estado del arte como solución a un regularizador en problema de reconstrucción. Esto se asume para poder aprovechar las características del algoritmo que se introduce. Uno de los tipos de algoritmos más usados en esta estrategia son lo de eliminación de ruido. Específicamente, se utilizó un algoritmo PnP con un filtro recursivo<sup>29</sup> (RF).

Adicionalmente, se utilizó como método comparativo una interpolación simple (init) donde cada uno de los disparos retirados se estimaba como un promedio de su anterior y el siguiente. Se agregó este procedimiento a la comparación debido a que algunos algoritmos implementados utilizan esta estrategia como inicialización.

---

<sup>35</sup> Amir Beck y Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". En: *SIAM journal on imaging sciences* 2.1 (2009), págs. 183-202.

<sup>36</sup> José M Bioucas-Dias y Mário AT Figueiredo. "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration". En: *IEEE Transactions on Image processing* 16.12 (2007), págs. 2992-3004.

<sup>37</sup> Xuejun Liao, Hui Li y Lawrence Carin. "Generalized alternating projection for weighted-2,1 minimization with applications to model-based compressive sensing". En: *SIAM Journal on Imaging Sciences* 7.2 (2014), págs. 797-823.

<sup>38</sup> Stanley Osher y col. "An iterative regularization method for total variation-based image restoration". En: *Multiscale Modeling & Simulation* 4.2 (2005), págs. 460-489.

## **4.4. DISEÑO DE EXPERIMENTOS**

**4.4.1. Configuración experimental.** Para validar la efectividad del método propuesto, se realizaron experimentos donde se simulaba una adquisición compresiva sobre los *cross-spreads* descritos en la sección 4.1. Para ello, se procedió a retirar el 25 % de los disparos en el tensor sísmico, siguiendo un esquema de submuestreo aleatorio para los índices de los disparos que se retiraban. Posteriormente, se ingresaba el dato submuestreado en cada uno de los algoritmos para completar los disparos que fueron retirados.

Las simulaciones fueron realizadas en un equipo de cómputo con las siguientes especificaciones: procesador Intel i7 de 10ma generación 3.8 GHz (16 núcleos de procesamiento) con memoria RAM de 32 GB DDR4.

**4.4.2. Configuración estudio de ablación.** La cantidad de datos faltantes es un factor crítico que afecta el desempeño de los algoritmos de recuperación de datos sísmicos. Por lo tanto, con el objetivo de validar la robustez del algoritmo propuesto ante este factor, se procedió a repetir el experimento descrito en la sección 4.4.1, pero variando la tasa de submuestreo de la adquisición. Específicamente, se utilizaron tasas de 0.15, 0.25, 0.35 y 0.5, pero esta vez se submuestreó siguiendo un esquema uniforme con el fin de poder comparar el desempeño del algoritmo en ambos esquemas.

## 5. RESULTADOS Y DISCUSIÓN

En esta sección se presentan los resultados obtenidos de las simulaciones sobre reconstrucción de datos sísmicos faltantes que permiten evaluar el desempeño del algoritmo desarrollado, en comparación con otros algoritmos de reconstrucción del estado del arte. Adicionalmente, se presenta un análisis sobre la robustez del método en función de la cantidad de datos faltantes.

### 5.1. VALIDACIÓN CRUZADA

Se hizo uso de validación cruzada para estimar cada uno de los parámetros en todos los métodos implementados. Para el algoritmo desarrollado en este trabajo, los parámetros  $\mu_\ell$  que regularizan el bajo rango en el dominio no-local en cada modo del tensor fueron variados en escala logarítmica sobre el rango  $[10^{-18}, 0,1]$ , mientras que el parámetro del paso dual  $\rho$  se varió logarítmicamente sobre el intervalo  $[10^{-4}, 1]$ .

La Figura 8 muestra un mapa de calor que resume la validación cruzada realizada para el algoritmo propuesto. Los valores óptimos para los parámetros se encontraron cuando se establece un  $\rho$  en el orden de  $10^{-4}$  y los parámetros  $\mu_\ell$  se configuran en el orden de  $10^{-6}$ . Esta configuración es válida para ambos conjuntos de datos.

El parámetro del rango  $r_\ell$  fue evaluado en el rango  $[25, 100]$  con pasos de 5 unidades entre cada prueba. En las simulaciones se pudo notar una tendencia de aumentar la calidad de los datos a medida que se aumentaba el rango hasta un valor de  $r_\ell = 75$ . Después de este valor, en ambos conjuntos de datos no se obtuvo una ganancia significativa. Por lo tanto, se tomó este valor como el óptimo para el método.

Respecto a los otros métodos, a continuación se describen uno a uno el rango en donde se variaron los parámetros en la validación, o la estrategia que se utilizó para escoger la configuración de parámetros.

- Para el CE, los parámetros de paso dual, regularización de escasez y regularización de eliminación de ruido fueron seleccionados en la misma escala que propusieron los autores en el trabajo original <sup>5</sup>.
- Para ResNet, los parámetros de la arquitectura de red neuronal como la cantidad de capas convolucionales ocultas, número de filtros y tamaño de los filtros se ajustaron siguiendo lo descrito por los autores en el trabajo original <sup>33</sup>. Para el entrenamiento, se utilizó como optimizador un gradiente estocástico con una tasa de aprendizaje de  $10^{-2}$  con 5000 épocas.
- Para MSSA, las frecuencias de corte fueron analizadas sobre los disparos no eliminados, y se escogieron valores de  $f_{\text{low}} = 0[\text{Hz}]$  y  $f_{\text{high}} = 80[\text{Hz}]$ . El parámetro del rango  $N$  fue variado entre  $[5, 80]$  con pasos de 5 unidades entre cada prueba.
- Para todos los otros métodos, el parámetro de regularización  $\lambda$  se varió en una escala logarítmica sobre  $[10^{-5}, 10^{-0,5}]$ , mientras que:
  - En el ADMM, TV y RF, el parámetro dual  $\rho$  se varió linealmente entre  $[0.5, 1.5]$ .
  - En el FISTA, el parámetro  $\mu$  se varió logarítmicamente en la misma escala  $[10^{-5}, 10^{-0,5}]$ .
  - En el TWIST, los parámetros  $\alpha$  y  $\beta$  se variaron linealmente en los intervalos  $[1, 2]$  y  $[0.8, 2]$ , respectivamente.

Es importante resaltar que no se reportan más datos para el algoritmo GAP debido a que este no posee parámetros adicionales.

## 5.2. RESULTADOS DE SIMULACIONES

En la Tabla 2 se presentan los resultados obtenidos para la reconstrucción de disparos, utilizando todos los métodos implementados como se diseñó el experimento en la sección 4.4.1. En términos generales, se puede observar que el algoritmo propuesto obtiene una mejora ente 0.3 y 1.5 dB de PSNR, y entre 0.03 y 0.07 de SSIM respecto a los métodos comparativos. Esta ganancia implica un mejor desempeño para la recuperación de disparos en ambos conjuntos de datos.

Respecto a la complejidad computacional se puede notar que, si bien no es el método con un menor tiempo de ejecución, el algoritmo presenta el menor tiempo de cómputo respecto a los métodos especializados de interpolación sísmica del estado del arte, obteniendo a su vez la mejor calidad.

Tabla 2. Comparación de la calidad en los disparos recuperados para los conjuntos de datos del SEAM phase II y Stratton. Cada valor es el promedio de todos los disparos eliminados. El PSNR y SNR es medido en decibelios, el tiempo computacional se presenta en segundos, y el MSE se encuentra escalado en el orden de  $10^{-3}$ .

		Propuesto	Interpolación sísmica			Algoritmos de escasez				TV	RF	Init
			CE <sup>39</sup>	ResNet <sup>40</sup>	MSSA <sup>41</sup>	ADMM	FISTA	TWIST	GAP			
SEAM phase II	MSE <sup>↓</sup>	<b>1.54</b>	2.29	21.57	<u>2.20</u>	2.87	3.58	2.69	2.74	3.42	3.20	11.49
	PSNR <sup>↑</sup>	<b>28.12</b>	26.40	16.66	<u>26.57</u>	25.42	24.46	25.70	25.61	24.66	24.94	19.40
	SNR <sup>↑</sup>	<b>22.18</b>	20.45	10.71	<u>20.62</u>	19.47	18.51	19.75	19.66	18.71	18.99	13.45
	SSIM <sup>↑</sup>	<b>0.81</b>	<u>0.74</u>	0.37	0.65	0.53	0.45	0.59	0.55	0.57	0.62	0.44
	Tiempo	808	3305	6262	5362	2752	9559	130	838	591	47	0.003
Stratton	MSE <sup>↓</sup>	<b>3.08</b>	<u>3.29</u>	4.10	3.52	4.24	4.60	3.66	4.07	3.66	3.56	4.95
	PSNR <sup>↑</sup>	<b>25.10</b>	<u>24.82</u>	23.87	24.54	23.72	23.38	24.37	23.91	24.37	24.49	23.05
	SNR <sup>↑</sup>	<b>19.23</b>	<u>18.94</u>	17.99	18.67	17.85	17.50	18.49	18.03	18.50	18.61	17.18
	SSIM <sup>↑</sup>	<b>0.67</b>	<u>0.64</u>	0.61	0.56	0.43	0.38	0.51	0.45	0.59	0.61	0.61
	Tiempo	198	915	704	261	1339	3959	59	405	281	22	0.002

En las Figuras 9 y 10 se presenta una comparación visual para todos los métodos para el conjunto de datos del SEAM phase II y Stratton, respectivamente. Las figuras muestran un mapa de error en donde se resaltan las áreas donde los errores presentan mayor magnitud. Adicionalmente, se presenta el análisis del dominio de frecuencia-longitud de onda (FK).

Observe que el método propuesto presenta los eventos de reflexión más definidos para los disparos en ambas bases de datos, mejorando no sólo en términos cuantitativos sino cualitativos la recuperación de disparos sísmicos causados por la eliminación de fuentes. Lo anterior se puede evidenciar en los mapas de error, en donde el error es mitigado, especialmente en los primeros arribos (eventos más definidos al inicio del disparo), los cuales, suelen ser los más afectados. Finalmente, observe en el dominio FK que el método propuesto mantiene las frecuencias de corte en ambas bases de datos. Particularmente, mantiene las frecuencias en la base de datos de Stratton, en la cual se observa, que otros métodos del estado del arte fallan.

### **5.3. ESTUDIO DE ABLACIÓN SOBRE LA TASA DE SUBMUESTREO**

En la Tabla 3 se presentan el PSNR y SSIM para para los dos conjuntos de datos en función a la tasa de submuestreo, como se configuró en la sección 4.4.2. Como se esperaba, el desempeño de los métodos mejora a medida que la cantidad de fuentes retiradas es menor. Sin embargo, es importante resaltar que el método propuesto obtiene el mejor desempeño en ambas métricas para tasas de 0.15, 0.25 y 0.35. Adicionalmente, se puede observar que la diferencia de la ganancia entre el método propuesto y los otros métodos del marco de referencia comparativo es mayor a medida que se dispone de más disparos en la reconstrucción. Este hecho implica que el algoritmo de reconstrucción propuesto aprovecha de mejor manera una mayor cantidad de datos observados, siendo ideal para adquisiciones con una tasa de submuestreo baja, de 0.15 a 0.25 aproximadamente.

Sería importante resaltar que los resultados en la Tabla 2 podrían ser comparados con los resultados obtenidos para 0.25 en la Tabla 3. Respecto a la comparativa, se ve cierta ganancia en los resultados de la Tabla 3, lo que podría implicar que el método tiene un mejor desempeño para el submuestreo uniforme. Sin embargo, de forma general se puede ver un desempeño bastante similar en ambas simulaciones, lo que implica que el algoritmo no es tan sensible respecto al esquema de submuestreo elegido.

Tabla 3. Comparación de la calidad en los disparos recuperados para los conjuntos de datos del SEAM phase II y Stratton en función del porcentaje de fuentes faltantes. Cada valor es el promedio de todos los disparos eliminados. El PSNR es medido en decibelios.

		Propuesto	Interpolación sísmica			Algoritmos de escasez				TV	RF	Init	
			CE <sup>42</sup>	ResNet <sup>43</sup>	MSSA <sup>44</sup>	ADMM	FISTA	TWIST	GAP				
SEAM phase II	0.15	PSNR <sup>†</sup>	<b>29.47</b>	19.67	25.36	<u>27.84</u>	26.12	25.79	26.86	26.39	26.46	26.80	24.61
		SSIM <sup>†</sup>	<b>0.77</b>	<u>0.71</u>	0.54	0.69	0.47	0.44	0.57	0.50	0.59	0.61	0.52
	0.25	PSNR <sup>†</sup>	<b>28.90</b>	<u>27.83</u>	25.35	27.19	25.68	25.17	26.46	25.94	26.05	26.36	24.20
		SSIM <sup>†</sup>	<b>0.76</b>	<u>0.74</u>	0.60	0.67	0.45	0.41	0.57	0.49	0.60	0.62	0.53
	0.35	PSNR <sup>†</sup>	<b>28.28</b>	<u>26.90</u>	25.13	26.75	25.43	24.64	26.20	25.68	25.91	26.10	24.03
		SSIM <sup>†</sup>	<b>0.74</b>	<u>0.70</u>	0.62	0.66	0.44	0.38	0.56	0.47	0.62	0.63	0.56
	0.5	PSNR <sup>†</sup>	<b>26.09</b>	25.86	24.88	25.52	24.72	6.08	25.65	24.96	<u>25.87</u>	25.21	23.57
		SSIM <sup>†</sup>	0.62	<b>0.64</b>	0.62	0.55	0.40	0.17	0.54	0.43	<u>0.63</u>	0.61	0.57
Stratton	0.15	PSNR <sup>†</sup>	<b>27.71</b>	18.77	24.81	25.55	24.53	24.30	25.29	24.73	25.57	<u>25.81</u>	24.26
		SSIM <sup>†</sup>	<b>0.75</b>	<u>0.69</u>	0.67	0.60	0.44	0.42	0.54	0.47	0.63	0.67	0.68
	0.25	PSNR <sup>†</sup>	<b>25.26</b>	<u>24.99</u>	23.89	24.54	23.65	22.91	24.31	23.85	24.37	24.57	23.10
		SSIM <sup>†</sup>	<b>0.67</b>	<u>0.66</u>	0.60	0.56	0.42	0.35	0.51	0.45	0.60	0.62	0.63
	0.35	PSNR <sup>†</sup>	<b>25.77</b>	<u>25.74</u>	24.43	24.55	23.51	23.18	24.29	23.72	24.66	24.94	23.43
		SSIM <sup>†</sup>	<b>0.68</b>	<u>0.67</u>	0.64	0.54	0.40	0.35	0.49	0.42	0.61	0.63	0.66
	0.5	PSNR <sup>†</sup>	<b>24.65</b>	<u>24.60</u>	23.27	24.75	23.17	5.84	24.11	23.40	24.55	<u>24.60</u>	22.80
		SSIM <sup>†</sup>	0.56	<u>0.61</u>	0.56	0.58	0.36	0.01	0.47	0.39	0.60	<u>0.61</u>	<b>0.62</b>

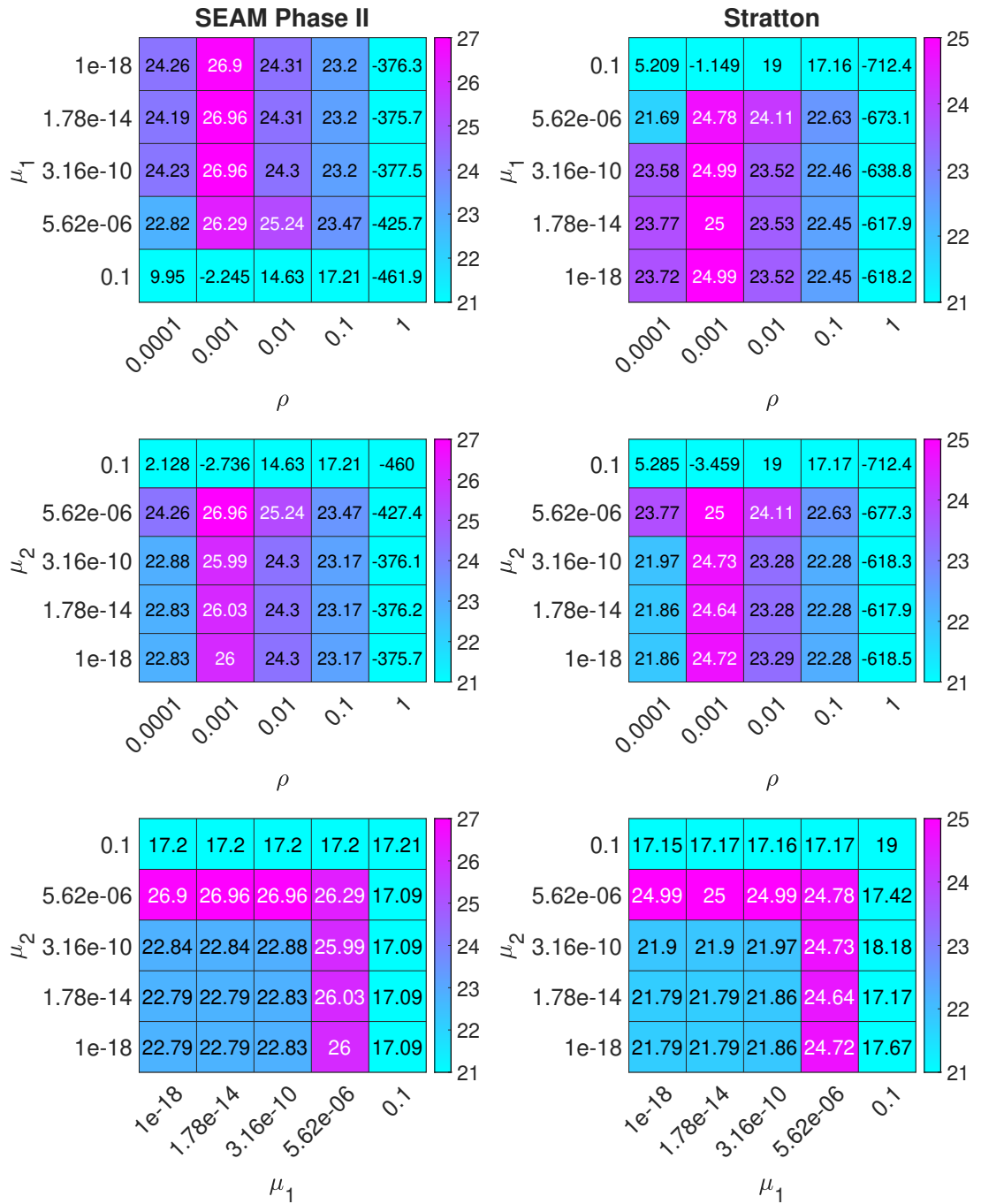


Figura 8. Mapa de calor sobre la metrica PSNR para diferentes combinaciones de los parámetros  $\mu_1$ ,  $\mu_2$  y  $\rho$  sobre los dos conjuntos de datos. Se puede observar que se presentan cambios suaves, donde las mejores reconstrucciones se obtuvieron para  $\rho = 0.001$ . Además, el algoritmo presenta un mejor desempeño cuando alguno de los parámetros  $\mu_\ell$  se configura en el orden de  $10^{-6}$ .

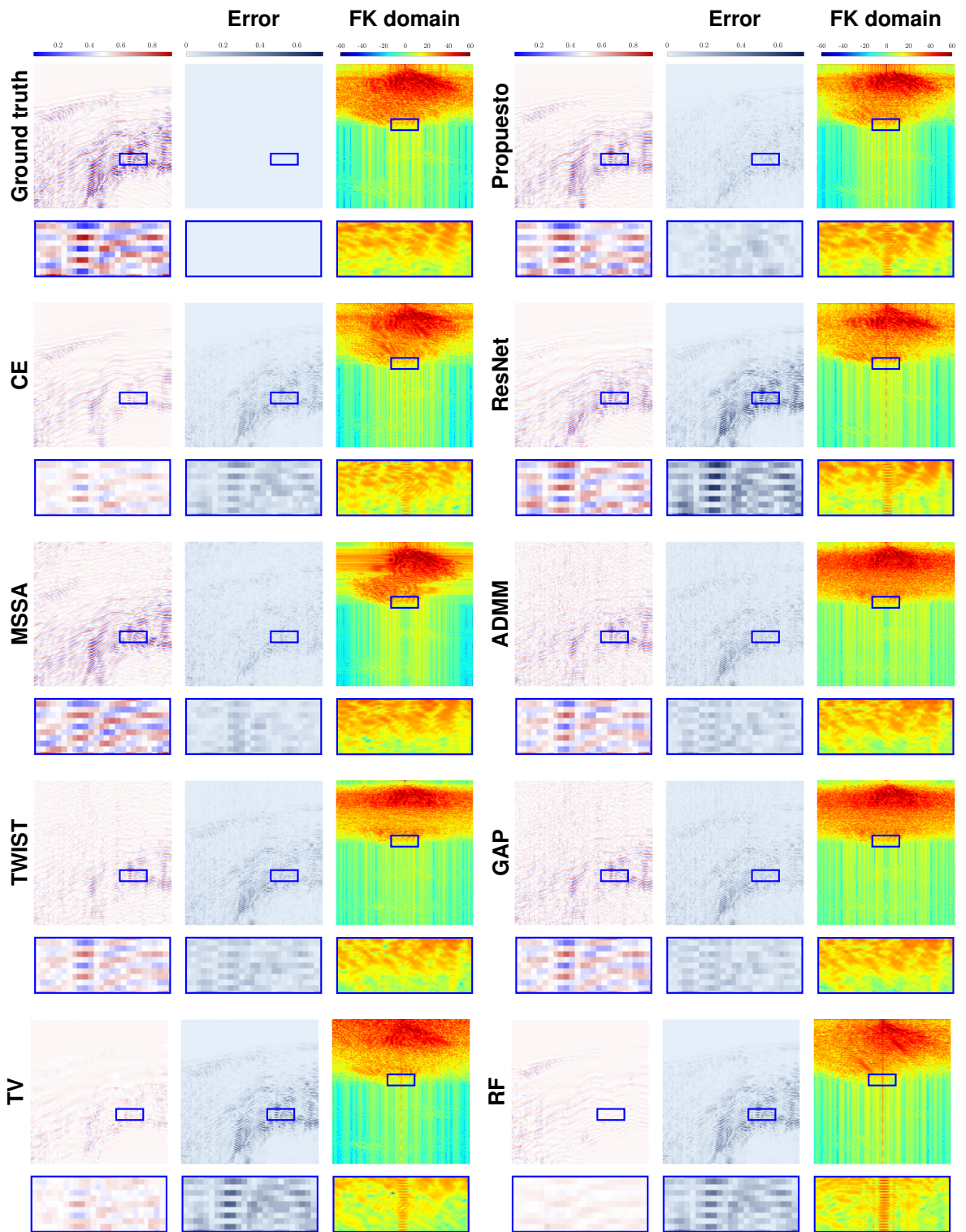


Figura 9. Resultados visuales para un disparo 43 reconstruido del SEAM phase II.

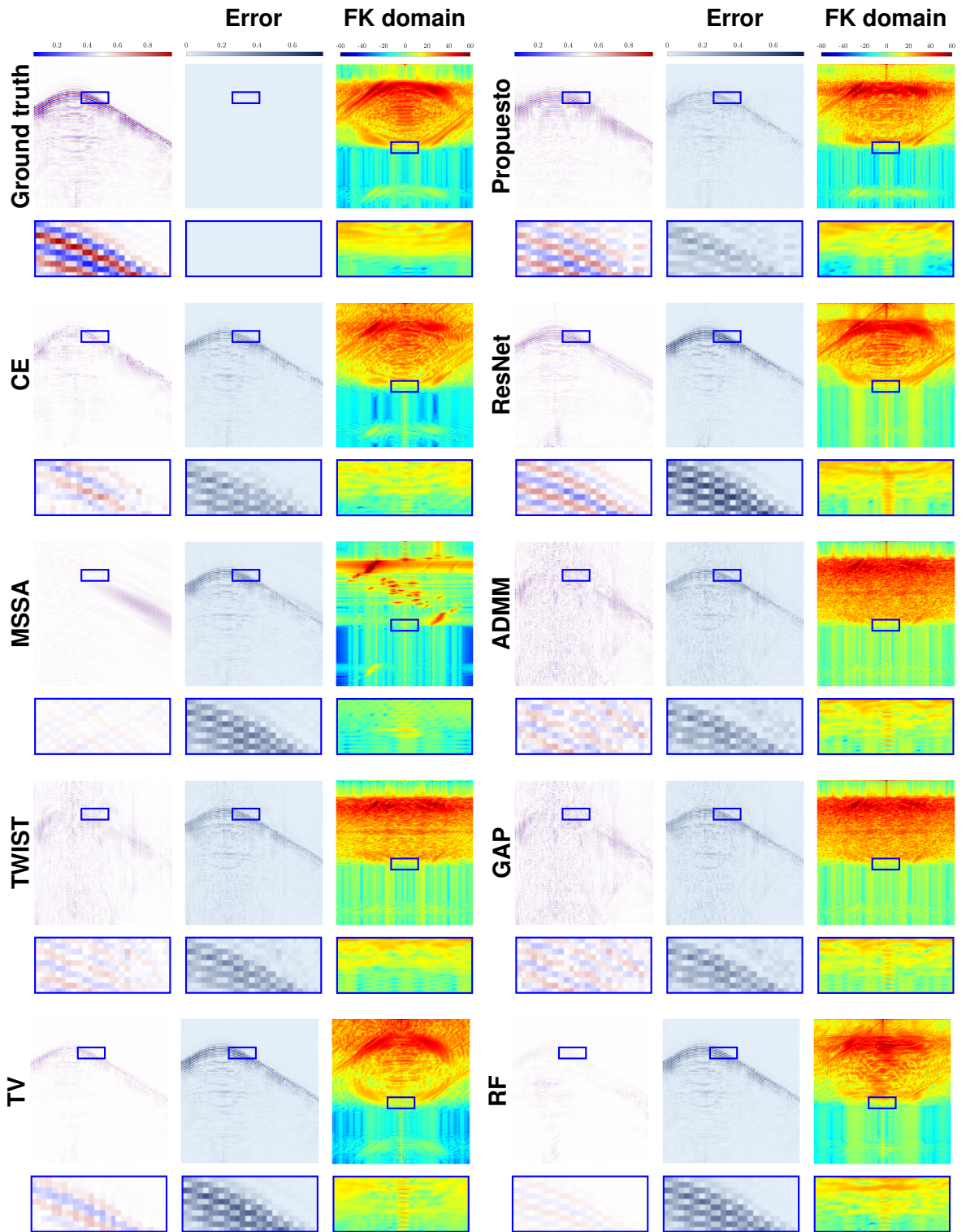


Figura 10. Resultados visuales para un disparo 4 reconstruido de Stratton.

## 6. CONCLUSIONES

En este trabajo se desarrolló un algoritmo de recuperación de datos sísmicos adquiridos mediante geometrías compresivas en el dominio *cross-spread*, utilizando un enfoque de optimización convexa con aproximación de bajo rango.

En primera instancia, se formuló el modelo de la adquisición compresiva utilizando notación tensorial, debido a que esta organización de los datos permite diferenciar las dimensiones del dato sísmico, y de este modo se pueden destacar las correlaciones entre las secciones de los diferentes modos del tensor. Con esta formulación se dio por cumplido el objetivo 1.

Seguidamente, se exploró la propiedad de bajo rango entre las diferentes dimensiones de los datos sísmicos haciendo uso de la definición de rango en modo  $m$ . Con este análisis se pudo concluir que los datos sísmicos presentan una rápida caída entre los valores singulares de sus modos, por lo cual se presentan altas correlaciones estructurales entre las secciones del dato adquirido. Por lo tanto, el dato sísmico puede ser aproximado de manera eficiente por un tensor de bajo rango. Con este análisis se dio por cumplido el objetivo 2.

Teniendo en cuenta lo anterior, la propiedad de bajo rango se incorporó en el problema de reconstrucción de datos sísmicos mediante un regularizador que utiliza una factorización tensorial. Se escogió una factorización tensorial que hace uso de las componentes principales del tensor para preservar dentro de la descomposición la coherencia estructural propia del dato original. Se concluyó que la regularización propuesta ayuda a condicionar mejor el problema de reconstrucción original. Con este planteamiento del problema se dio por cumplido el objetivo 3.

Para este problema de optimización convexa se encontró una solución numérica usando un algoritmo de ADMM, el cual permite establecer una estrategia de optimización alternada, separando el problema en problemas más sencillos. Se encontró que el algoritmo ADMM era el más apropiado para el problema formulado porque la descomposición propuesta sigue la forma lineal en las restricciones que establece este algoritmo. Con esta implementación se dio por cumplido el objetivo 4.

Finalmente, se evaluó el rendimiento del algoritmo de recuperación de datos sísmicos utilizando la base de datos sintético-realista del "*SEAM Phase II Land*", y los datos de la adquisición real "*Stratton 3D survey*". Los experimentos realizados muestran un mejor desempeño en la reconstrucción de datos sísmicos faltantes en comparación con los métodos del estado del arte que fueron implementados. Específicamente, se alcanza una ganancia de hasta 1.5 dB y 0.07 en la calidad de la señal (PSNR) y similitud estructural (SSIM) en las reconstrucciones. Del mismo modo, las reconstrucciones obtenidas visualmente muestran una mejor definición de los eventos sísmicos que se reconstruyen. Sería importante resaltar que el tiempo de cómputo del algoritmo propuesto es considerablemente menor al empleado por los métodos de interpolación sísmica implementados. Adicionalmente, se evaluó el rendimiento variando el porcentaje de submuestreo. En estos resultados se pudo observar que en un caso crítico, con un submuestreo del 50% de las fuentes retiradas, se logra un desempeño comparable con los métodos de reconstrucción de la literatura. Sin embargo, a medida que se tiene un mayor porcentaje de fuentes en la adquisición, el algoritmo obtiene una mayor ganancia en la reconstrucción. De igual forma, al comparar los esquemas de submuestreo uniforme y aleatorio, se pudo observar que el algoritmo no presenta alguna variación con respecto al esquema de submuestreo que se utilice para retirar las fuentes. Con esta validación se dio por cumplido el objetivo 5.

## 7. TRABAJO FUTURO

Como trabajo futuro, se podría evaluar el efecto que tiene la geometría de la adquisición en la calidad de la reconstrucción de los datos faltantes, y específicamente conocer las posiciones en que los disparos se ven favorecidos para ser reconstruidos por el algoritmo. Por otra parte, el algoritmo implementado podría ser adaptado para realizar reconstrucciones en datos sísmicos multi-ortogonales, en donde se apilan los *cross-spreads* de toda la adquisición, resultando en dos dimensiones adicionales el cubo de datos, el número de la línea de receptores, y el número de la línea de fuentes.

## BIBLIOGRAFÍA

- Abma, Ray y Nurul Kabir. "3D interpolation of irregular data with a POCS algorithm". En: *Geophysics* 71.6 (2006), E91-E97 (vid. pág. 25).
- Beck, Amir y Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". En: *SIAM journal on imaging sciences* 2.1 (2009), págs. 183-202 (vid. pág. 53).
- Bioucas-Dias, José M y Mário AT Figueiredo. "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration". En: *IEEE Transactions on Image processing* 16.12 (2007), págs. 2992-3004 (vid. pág. 53).
- Boyd, Stephen y col. "Distributed optimization and statistical learning via the alternating direction method of multipliers". En: *Foundations and Trends® in Machine learning* 3.1 (2011), págs. 1-122 (vid. págs. 38, 45, 52).
- Campman, Xander y col. "Sparse seismic wavefield sampling". En: *The Leading Edge* 36.8 (2017), págs. 654-660 (vid. págs. 12, 25).
- Chan, Stanley H, Xiran Wang y Omar A Elgendy. "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications". En: *IEEE Transactions on Computational Imaging* 3.1 (2016), págs. 84-98 (vid. págs. 38, 53).
- Chen, Siyuan y col. "Post-stack seismic data interpolation using a fast non-local similarity matching algorithm". En: *Studia Geophysica et Geodaetica* 65.1 (2021), págs. 59-69 (vid. pág. 37).

- Chen, Yangkang y col. "An open-source Matlab code package for improved rank-reduction 3D seismic data denoising and reconstruction". En: *Computers & Geosciences* 95 (2016), págs. 59-66 (vid. págs. 52, 57, 59).
- Chen, Yangkang y col. "The interpolation of sparse geophysical data". En: *Surveys in Geophysics* 40.1 (2019), págs. 73-105 (vid. págs. 13, 16, 21, 27).
- Dabov, Kostadin y col. "Image denoising by sparse 3-D transform-domain collaborative filtering". En: *IEEE Transactions on image processing* 16.8 (2007), págs. 2080-2095 (vid. pág. 47).
- Goyes-Peñafiel, Paul y col. "A Consensus Equilibrium Approach for 3-D Land Seismic Shots Recovery". En: *IEEE Geoscience and Remote Sensing Letters* 19 (2021), págs. 1-5 (vid. págs. 13, 26, 51, 56, 57, 59).
- He, Wei y col. "Non-local meets global: An iterative paradigm for hyperspectral image restoration". En: *arXiv preprint arXiv:2010.12921* (2020) (vid. pág. 37).
- Herrmann, Felix J y Gilles Hennenfent. "Non-parametric seismic data recovery with curvelet frames". En: *Geophysical Journal International* 173.1 (2008), págs. 233-248 (vid. pág. 26).
- Huang, Xiaogang y col. "Efficient seismic acquisition based on compressive sensing". En: *SEG 2020 Workshop: Broadband and Wide-azimuth Deepwater Seismic Technology*. Society of Exploration Geophysicists. 2020, págs. 35-37 (vid. pág. 15).
- Jiang, Tao y col. "Compressive seismic reconstruction with extended POCS for arbitrary irregular acquisition". En: *2017 SEG International Exposition and Annual Meeting*. OnePetro. 2017 (vid. pág. 25).

Kolda, Tamara G y Brett W Bader. "Tensor decompositions and applications". En: *SIAM review* 51.3 (2009), págs. 455-500 (vid. pág. 29).

Kong, Fantong y col. "Deep prior-based seismic data interpolation via multi-res U-net". En: *SEG International Exposition and Annual Meeting*. OnePetro. 2020 (vid. pág. 26).

*La Agencia Nacional de Hidrocarburos de Colombia implementa tecnología Blockchain para la asignación de áreas de explotación | ACIS*. <https://acis.org.co/portal/content/no-agencia-nacional-de-hidrocarburos-de-colombia-implementa-tecnolog%C3%ADa-blockchain-para-la>. Accessed: 2021-02-16 (vid. pág. 14).

Liao, Xuejun, Hui Li y Lawrence Carin. "Generalized alternating projection for weighted-2,1 minimization with applications to model-based compressive sensing". En: *SIAM Journal on Imaging Sciences* 7.2 (2014), págs. 797-823 (vid. pág. 53).

Liu, Bin y Mauricio D Sacchi. "Minimum weighted norm interpolation of seismic records". En: *Geophysics* 69.6 (2004), págs. 1560-1568 (vid. pág. 24).

Liu, Qun, Lihua Fu y Meng Zhang. "Deep-seismic-prior-based reconstruction of seismic data using convolutional neural networks". En: *Geophysics* 86.2 (2021), págs. V131-V144 (vid. pág. 26).

Liu, Wei y col. "Reconstruction of seismic data with missing traces based on local random sampling and curvelet transform". En: *Journal of Applied Geophysics* 115 (2015), págs. 129-139 (vid. págs. 25, 26).

López, Kareth León y col. "Regular multi-shot subsampling and reconstruction on 3d orthogonal symmetric seismic grids via compressive sensing". En: *2019 XXII*

- Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE. 2019, págs. 1-5 (vid. págs. 12, 26).
- Ma, Jianwei. “Three-dimensional irregular seismic data reconstruction via low-rank matrix completion”. En: *Geophysics* 78.5 (2013), págs. V181-V192 (vid. págs. 16, 27).
- Osher, Stanley y col. “An iterative regularization method for total variation-based image restoration”. En: *Multiscale Modeling & Simulation* 4.2 (2005), págs. 460-489 (vid. pág. 53).
- Popa, Jonathan, Susan E Minkoff y Yifei Lou. “An improved seismic data completion algorithm using low-rank tensor optimization: Cost reduction and optimal data orientation”. En: *Geophysics* 86.3 (2021), págs. V219-V232 (vid. págs. 13, 15, 28).
- Regone, Carl y col. “Geologic model building in SEAM Phase II—Land seismic challenges”. En: *The Leading Edge* 36.9 (2017), págs. 738-749 (vid. pág. 48).
- Stratton 3D survey*. [https://wiki.seg.org/wiki/Stratton\\_3D\\_survey](https://wiki.seg.org/wiki/Stratton_3D_survey) (vid. pág. 49).
- Trickett, Stewart y col. “Rank-reduction-based trace interpolation”. En: *SEG Technical Program Expanded Abstracts 2010*. Society of Exploration Geophysicists, 2010, págs. 3829-3833 (vid. págs. 25, 27).
- Vermeer, Gijs JO. *3D seismic survey design*. Society of Exploration Geophysicists, 2012 (vid. pág. 12).
- Villarreal, Ofelia P y col. “Compressive sensing seismic acquisition by using regular sampling in an orthogonal grid”. En: *2017 IEEE 7th International Workshop on*

- Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, págs. 1-5 (vid. págs. 13, 26).
- Wang, Benfeng y col. "Intelligent missing shots' reconstruction using the spatial reciprocity of Green's function based on deep learning". En: *IEEE Transactions on Geoscience and Remote Sensing* 58.3 (2019), págs. 1587-1597 (vid. págs. 52, 56, 57, 59).
- Wang, Dong, Jia Xu y Ke Xu. "An FPGA-based hardware accelerator for real-time block-matching and 3D filtering". En: *IEEE Access* 8 (2020), págs. 121987-121998 (vid. pág. 24).
- Wang, Shu-Qin, Xing Gao y Zhen-Xing Yao. "Accelerating POCS interpolation of 3D irregular seismic data with graphics processing units". En: *Computers & Geosciences* 36.10 (2010), págs. 1292-1300 (vid. pág. 25).
- Wu, Juan y col. "Fast and robust low-rank approximation for five-dimensional seismic data reconstruction". En: *IEEE Access* 8 (2020), págs. 175501-175512 (vid. pág. 28).
- Yang, Pengliang, Jinghuai Gao y Wenchao Chen. "Irregularly sampled seismic data interpolation using iterative half thresholding regularization". En: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2013, págs. 5820-5824 (vid. pág. 26).