Efficient matrix completion for seismic data reconstruction

Rajiv Kumar[1,0], Curt Da Silva[2,0], Okan Akalin[3],Aleksandr Y. Aravkin[4], Hassan Mansour[5], Ben Recht[6], Felix J. Herrmann[1],

[1] *Department of Earth and Ocean Sciences,*

[2] *Department of Mathematics,*

University of British Columbia, Canada

[3] *Department of Computer Sciences, University of Wisconsin-Madison*

[4] *IBM T.J. Watson Research, USA*

[5] *Mitsubishi Electric Research Laboratories, USA*

[6] *Department of Electrical Engineering and Computer Science, Department of Statsitics,*

University of California, Berkeley, USA

[0] *Equal contributors*

(January 23, 2015)

Running head: **Efficient factorization based interpolation**

ABSTRACT

Despite recent developments in improved acquisition, seismic data often remains undersampled along source and/or receiver coordinates, resulting in incomplete data for key applications such as migration and multiple prediction requiring densely sampled, alias-free wide azimuth data. Missing-trace interpolation can be cast into a matrix completion problem by exploiting the low-rank behaviour of seismic data in the appropriate domain. In this work, we outline three practical principles for successful recovery using low-rank optimization techniques to successfully recover seismic data using

low-rank matrix completion techniques. We ensure that our optimization algorithms are practical by avoiding repeated singular value decompositions, which are prohibitively expensive for large-scale problems.

Current approaches that exploit low-rank structure are based on repeated singular value decompositions, which become prohibitively expensive for large-scale problems. While computationally manageable, our theory and experiments show degraded results when the windows sizes become too small. To overcome this problem, we carry out our interpolations for each frequency independently while working with the complete data in the midpoint-offset domain instead of windowing. To make our proposed method computationally viable and practical, we introduce a factorization-based approach that avoids computing the singular values, and that therefore scales to large seismic data problems as long as the factors can be stored in memory. Tests on realistic two- and three-dimensional seismic data show that our method compares favorably, both in terms of computational speed and recovery quality, to existing curvelet-based and tensor-based techniques.

INTRODUCTION

Coarsely sampled seismic data creates substantial problems for seismic applications such as migration and inversion (Canning and Gardner, 1998; Sacchi and Liu, 2005). In order to mitigate acquisition related artifacts, we rely on interpolation algorithms to reproduce the missing traces accurately. The aim of these interpolation algorithms is to reduce acquisition costs and to provide densely sampled seismic data to improve the resolution of seismic images and mitigate subsampling related artifacts such as aliasing. A variety of methodologies, each based on various mathematical techniques, have been proposed to interpolate seismic data. Some of the methods require transforming the data into different domains, such as the Radon (Bardan, 1987; Kabir and Verschuur, 1995), Fourier (Duijndam et al., 1999; Sacchi et al., 1998; Curry, 2009; Trad, 2009) and curvelet domains (Herrmann and Hennenfent, 2008; Sacchi et al., 2009; Wang et al., 2010). The CS approach exploits the resulting sparsity of the signal, i.e. small number of nonzeros (Donoho, 2006) in these domains. In the CS framework, the goal for effective recovery is to first find a representation in which the signal of interest is sparse, or well-approximated by a sparse signal, and where the the mask encoding missing traces makes the signal much less sparse. Hennenfent and Herrmann (2006a); Herrmann and Hennenfent (2008) successfully applied the ideas of CS to the reconstruction of missing seismic traces in the curvelet domain.

More recently, rank-minimization-based techniques have been applied to interpolating seismic data (Trickett et al., 2010; Oropeza and Sacchi, 2011; Kreimer and Sacchi, 2012b,a; Yang et al., 2013). Rank minimization is a 2D analogue of the inherently 1D compressed sensing theory (see Recht et al. (2010) and the references within). The key idea is to exploit the low-rank structure of seismic data when organized as a matrix, i.e. a small number of nonzero singular values or quickly

decaying singular values. Oropeza and Sacchi (2011) identified that seismic temporal frequency slices organized into a block Hankel matrix, under ideal conditions, is a matrix of rank k, where k is the number of different plane waves in the window of analysis. These authors showed that additive noise and missing samples increase the rank of the block Hankel matrix, and the authors presented an iterative algorithm that resembles seismic data reconstruction with the method of projection onto convex sets, where they use a low-rank approximation of the Hankel matrix via the randomized singular value decomposition (Liberty et al., 2007; Halko et al., 2011; Mahoney, 2011) to interpolate seismic temporal frequency slices. While this technique may be effective for interpolating data with a limited number of distinct dips, first, the approach requires embedding the data into an even larger space where each dimension of size n is mapped to a matrix of size $n \times n$, so a frequency slice with 4 dimensions becomes a Hankel tensor with 8 dimensions. Second, the process involves partitioning the input data in to smaller subsets that can be processed independently. As we know the theory of matrix completion is predicated upon the notion of an $m \times n$ matrix being *relatively* low rank in order to ensure successful recovery. That is, the ratio of rank of the matrix to the ambient dimension, $\min(m, n)$, should be small for rank-minimizing techniques to be successful in recovering the matrix from appropriately subsampled data. With the practice of windowing, we are inherently *increasing* the relative rank by decreasing the ambient dimension. Although mathematically desirable due to the seismic signal being stationary in sufficiently small windows, the act of windowing from a matrix rank point of view can lead to lower quality results, as we will see later in experiments. Choosing window sizes apriori is also a difficult task, as it is not altogether obvious how to ensure that the resulting sub-volume is approximately a plane-wave. Previously proposed methods for automatic window size selection include Sinha et al. (2005); Wang et al. (2011) in the context of time-frequency analysis.

Other than the Hankel transformation, Yang et al. (2013) used a texture-patch based transformation of the data, initially proposed by Schaeffer and Osher (2013), to exploit the low-rank structure of seismic data. They showed that seismic data can be expressed as a combination of a few textures, due to continuity of seismic data. They divided the signal matrix into small $r \times r$ submatrices, which they then vectorized in to the columns of a matrix with r^2 rows using the same ordering, and approximated the resulting matrix using low rank techniques. Although experimentally promising, this organization has no theoretically motivated underpinning and its performance is difficult to predict as a function of the submatrix size. The authors proposed two algorithms to solve this matrix completion problem, namely accelerated proximal gradient method (APG) and low-rank matrix fitting (LMaFit). APG does not scale well to large scale seismic data because it involves repeated singular value decompositions, which are very expensive. LMaFit, on the other hand, parametrizes the matrix in terms of two low-rank factors and uses nonlinear successive-over-relaxation to reconstruct the seismic data, but *without* penalizing the nuclear norm of the matrix. As shown in Aravkin et al. (2014), without a nuclear norm penalty, choosing an incorrect rank parameter k can lead to *overfitting* of the data and degrading the interpolated result. Moreover, Mishra et al. (2013) demonstrates the poor performance of LMaFit, both in terms of speed and solution quality, compared to more modern matrix completion techniques that penalize the nuclear norm.

Another popular approach to seismic data interpolation is to exploit the multi-dimensional nature of seismic data and parametrize it as a *low-rank tensor*. Many of the ideas from low rank matrices carry over to the multidimensional case, although there is no unique extension of the SVD to tensors. It is beyond the scope of this paper to examine all of the various tensor formats in this paper, but we refer to a few tensor-based seismic interpolation methods here. Kreimer and Sacchi

(2012a) stipulates that the seismic data volume of interest is well captured by a $k-$rank Tucker tensor and subsequently propose a projection on to non-convex sets algorithm for interpolating missing traces. Da Silva and Herrmann (2014) develop an algorithm for interpolating Hierarchical Tucker tensors, which are similar to Tucker tensors but have much smaller dimensionality. Trickett et al. (2013) proposes to take a structured outer product of the data volume, using a tensor ordering similar to Hankel matrices, and performs tensor completion in the CP-Parafac tensor format. The method of Kreimer et al. (2013), wherein the authors consider a nuclear norm-penalizing approach in each matricization of the tensor, that is to say, the reshaping of the tensor, along each dimension, in to a matrix.

These previous CS-based approaches, using sparsity or rank-minimization, incur computational difficulties when applied to large scale seismic data volumes. Methods that involve redundant transforms, such as curvelets, or that add additional dimensions, such as taking outer products of tensors, are not computationally tractable for large data volumes with four or more dimensions. Moreover, a number of previous rank-minimization approaches are based on heuristic techniques and are not necessarily adequately grounded in theoretical considerations. Algorithmic components such as parameter selection can significantly affect the computed solution and "hand-tuning" parameters, in addition to incurring unnecessary computational overhead, may lead to suboptimal results (Owen and Perry, 2009; Kanagal and Sindhwani, 2010).

Contributions

Our contributions in this work are three-fold. First, we outline a practical framework for recovering *seismic data* volumes using *matrix* and *tensor* completion techniques built upon the theoretical

ideas from CS. In particular, understanding this framework allows us to determine apriori when the recovery of signals sampled at sub-Nyquist will succeed or fail and provides the principles upon which we can design practical experiments to ensure successful recovery. The ideas themselves have been established for some time in the literature, albeit implicitly by means of the somewhat technical conditions of CS and matrix completion. We explicitly describe these ideas on a high level in a qualitative manner in the hopes of broadening the accessibility of these techniques to a wider audience. These principles are all equally necessary in order for CS-based approaches of signal recovery to succeed.

Second, we address the computational challenges of using these matrix-based techniques for seismic-data reconstruction, since traditional rank minimization algorithms rely on computing the singular value decomposition (SVD), which is prohibitively expensive for large matrices. To overcome this issue we propose to use either a fast optimization approach that combines the (SVD-free) matrix factorization approach recently developed by Lee et al. (2010) with the Pareto curve approach proposed by van den Berg and Friedlander (2008) and the factorization-based parallel matrix completion framework dubbed Jellyfish (Recht and Ré, 2013). We demonstrate the superior computational performances of both of these approaches compared to the tensor-based interpolation of Kreimer et al. (2013) as well as traditional curvelet-based approaches on realistic 2D and 3D seismic data sets.

Third, we examine the popular approach of *windowing* a large data volume in to smaller data volumes to be processed in parallel and empirically demonstrate how such a process does not respect the inherent redundancy present in the data, degrading reconstruction quality as a result.

Notation

In this paper, we use lower case boldface letters to represent vectors (i.e. one-dimensional quantities), e.g., $\mathbf{b}, \mathbf{f}, \mathbf{x}, \mathbf{y}, \ldots$. We denote matrices and tensors using upper case boldface letters, e.g., $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$ and operators that act on vectors, matrices, or tensors will be denoted using calligraphic upper case letters, e.g., $\mathcal{A}$. 2D seismic volumes have one source and one receiver dimensions, denoted $x_{\text{src}}, x_{\text{rec}}$, respectively, and time, denoted t. 3D seismic volumes have two source dimensions, denoted $x_{\text{src}}, y_{\text{src}}$, two receiver dimensions, denoted $x_{\text{rec}}, y_{\text{rec}}$, and time t. We also denote midpoint and offset coordinates as $x_{\text{midpt}}, x_{\text{offset}}$ for the x-dimensions and similarly for the y-dimensions.

STRUCTURED SIGNAL RECOVERY

In this section, we codify a qualitative framework for signal recovery for matrix/tensor completion techniques. The three requirements that we outline below, which are based on an understanding of Compressed Sensing that we first review, are necessary for recovering a low-rank signal.

Compressed Sensing (CS) is a fundamentally different approach for signal recovery compared to traditional, Shannon-Nyquist based approaches. CS allows us to sample at *sub*-Nyquist rates for signals that exhibit additional structure besides being merely band-limited and subsequently recover fully-sampled discrete signals. Any successful, CS-based signal recovery scheme consists of three main components.

1. **Signal structure - sparsity**

   The underlying signal of interest is either $k-$sparse, that is, it has at most k nonzeros or is *compressible* — i.e. the magnitude of the coefficients aside from its k largest ones should be

small. For the seismic case, both seismic data and seismic images tend to be very sparse in curvelets (Hennenfent and Herrmann, 2006b; Mansour et al., 2012; Herrmann and Li, 2012; Li et al., 2012). On the other hand, the *redundancy* of curvelets in particular, i.e. the size of $\mathbf{x}$ is at least seven times the size of $\mathbf{f}$, makes their application to large-scale, 3D data challenging.

2. **Structure-destroying sampling operator**

   Once we know that our signal of interest is sparse or compressible, we need a sampling process that breaks this structure, that is, decreases the sparsity of the original signal. In case of seismic, periodic subsampling generate erroneous spikes, i.e. aliasing, which are intermingled with the correct spikes of the original signal and are subsequently difficult to remove (Mansour et al., 2012). Randomized subsampling, on the other hand, turn those *aliases* in to low levels diffuse Gaussian noise at every location. In this sense, the sparsity of *signal* has been destroyed and recovery can be guaranteed with high probability (Herrmann and Hennenfent, 2008; Hennenfent and Herrmann, 2008).

3. **Structure-promoting optimization program** Once the above two principles hold, we can solve an optimization problem in order to recover the fully sampled signal of interest, where the aim is to look for the sparsest signal that fits our data, i.e., we solve the following problem,

$$\underset{\text{coefficient vectors } \mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_1$$

$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

   The sampling-transformation operator $\mathbf{A}$ is defined as $\mathbf{A} = \mathbf{R}\mathbf{M}\mathbf{D}^H$, where $\mathbf{R}$ restricts to a subset of measurement entries $\mathbf{M}$, $\mathbf{D}$ transforms the data from the physical domain to the the sparsity domain and ^{H} represents the conjugate-transpose. Minimizing the ℓ_1 norm promotes sparse solutions. One of the most popular and efficient methods for solving the above problem

is $SPG\ell_1$ van den Berg and Friedlander (2008), which uses a spectral projected gradient method combined with an efficient parameter cooling method to solve the above optimization program and is particularly well-suited for large-scale problems.

While CS is a theory concerned primarily with one-dimensional functions, it is not the only approach to recovering subsampled signals. In particular in seismic cases, our signals of interest range anywhere from two to five dimensions. It is then natural to exploit this additional structure for signal recovery, which leads us to the notion of *matrix completion*. In this setting, we are interested in completing a matrix $\mathbf{X}$ when we only view a subset of its entries. For instance, in the 2D seismic data case, $\mathbf{X}$ is typically a frequency slice and missing shots correspond to missing columns from this matrix. Matrix completion arises as a natural extension of CS ideas to recovering two dimensional signals. As in the previous case, there are three core components of matrix completion.

1. **Signal structure - low rank**

   For a matrix $\mathbf{X}$, a direct analogue for sparsity in a signal $\mathbf{x}$ is sparsity in the *singular values* of $\mathbf{X}$. As in CS-based recovery, we are now interested in the case where the singular values of $\mathbf{X}$ decay quickly, so that $\mathbf{X}$ is well approximated by a rank k matrix. When our matrix $\mathbf{X}$ has slowly decaying singular values, i.e. is high rank, we consider transformations that promote quickly decaying singular values, which will allow us to recover our matrix in another domain.

   Since we are sampling points from our underlying matrix $\mathbf{X}$, we want to make sure that $\mathbf{X}$ is not too "spiky" and is sufficiently "spread out". If our matrix of interest was, for instance, the matrix of all zeros with one nonzero entry, we could not hope to recover this matrix without sampling the single, nonzero entry. In the seismic case, given that our signals of interest are composed of oscillatory waveforms, they are rarely, if ever, concentrated in a single region of,

say, (source,receiver) space.

2. **Structure-destroying sampling operator**

As in the compressed sensing case, we will look for the matrix $\mathbf{X}$ of smallest rank that satisfies our sampled data, $\mathcal{A}(\mathbf{X}) = \mathbf{B}$, for a subsampling operator $\mathcal{A}$. As such, we need to employ subsampling schemes that increase the rank or decay of the singular values of the matrix. That is to say, we want to consider sampling schemes that are incoherent with respect to the left and right singular vectors. Given a subsampling operator $\mathcal{A}$, the worst possible subsampling scheme for the purposes of recovery would be removing columns (equivalently, rows) from the matrix, i.e. $\mathcal{A}(\mathbf{X}) = \mathbf{X}\mathbf{I}_k$, where $\mathbf{I}_k$ is a subset of the columns of identity matrix. Removing columns from the matrix can never allow for successful reconstruction because this operation lowers the rank, and therefore the original matrix $\mathbf{X}$ is no longer the matrix of smallest rank that matches the data (for instance, the data itself would be a candidate solution).

Unfortunately, for, say, a 2D seismic data frequency slice $\mathbf{X}$ with sources placed along the columns and receivers along the rows, data is often acquired with missing sources, which translates to missing columns of $\mathbf{X}$. Similarly, periodic subsampling can be written as $\mathcal{A}(\mathbf{X}) = \mathbf{I}_k^T\mathbf{X}\mathbf{I}_{k'}$, where $\mathbf{I}_k, \mathbf{I}_{k'}$ are subsets of the columns of the identity matrix. A similar consideration yields that this operator lowers the rank and thus rank minimizing interpolation will not succeed in this sampling regime.

The problematic aspect of the aforementioned sampling schemes is that they are separable with respect to the matrix. That is, if $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ is the singular value decomposition of $\mathbf{X}$, the previously mentioned schemes yield a subsampling operator of the form $\mathcal{A}(\mathbf{X}) = \mathbf{C}\mathbf{X}\mathbf{D}^H = (\mathbf{C}\mathbf{U})\mathbf{S}(\mathbf{D}\mathbf{V})^H$, for some matrices $\mathbf{C}, \mathbf{D}$. In the compressed sensing context, this

type of sampling is coherent with respect to the left and right singular vectors, which is an unfavourable recovery scenario.

The incoherent sampling considered in the matrix completion literature is that of uniform random sampling, wherein the individual entries of $\mathbf{X}$ are sampled from the matrix with equal probability (Candès and Recht, 2009; Recht, 2011). This particular sampling scheme, although theoretically convenient to analyze, is impractical to implement in the seismic context as it corresponds to removing (source, receiver) pairs from the data. Instead, we will consider non-separable transformations under which the missing sources operator is incoherent. The resulting transformations will simultaneously increase the decay of the singular values of our original signal, thereby lowering its rank, and slow the decay of the singular values of the subsampled signal, thereby creating a favourable recovery scenario.

3. **Structure-promoting optimization program**

Since we assume that our target signal $\mathbf{X}$ is low-rank, and that subsampling *increases* the rank, the natural interpolation problem to solve is to find the matrix of *lowest* possible rank that agrees with our observations. That is, we solve the following problem for $\mathcal{A}$, our measurement operator, and $\mathbf{B}$, our subsampled data, up to a given tolerance σ,

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \|\mathbf{X}\|_* \\ &\text{subject to } \|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_F \leq \sigma. \end{aligned} \tag{1}$$

Here $\|\mathbf{X}\|_* = \sum_{i=1}^{n} s_i$ is the *nuclear norm* of $\mathbf{X}$, the sum of its singular values, and $\|\mathbf{B}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{B}_{i,j}^2}$ is the *Frobenius* norm of $\mathbf{B}$, the matrix analogue of the vector $2-$norm. Similar to using the ℓ_1 norm in the sparse recovery case, minimizing the nuclear norm promotes low-rank structure in the final solution. Here we refer to this problem as Basis Pursuit

Denoising ($BPDN_\sigma$).

A popular approach to solve Problem 1 is to lift the constraints to the objective function itself, resulting in the QP_λ formulation,

$$\underset{\mathbf{X}}{\text{minimize}} \frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_F^2 + \lambda\|\mathbf{X}\|_*. \tag{2}$$

This lifting results in an unconstrained problem that can be solved using several convex optimization techniques. One such approach using Singular Value Thresholding is given in Cai et al. (2010). A practical approach requires a scheme to choose λ in the above formulation so that the solution to Problem 2 is equal to the solution of $BPDN_\sigma$, our actual problem of interest (Giryes et al., 2011).

The $BPDN_\sigma$ formulation has an advantage over QP_λ in that σ has the natural interpretation of the *noise level* in the data and is thus more readily estimated in practice than the penalty parameter λ. On the other hand, problem (1) is challenging to solve at first glance, as the objective is nondifferentiable and the constraints $\|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_2 \leq \sigma$ are not simple. Instead, we refer to Appendix A to solve the problem (1) using the pareto-curve based $SPG\ell_1$ framework.

In summary, these three principles are all necessary for the recovery of subsampled signals using matrix completion techniques. Omitting any one of these three principles will, in general, cause such methods to fail, which we will see in the next section. Although this framework is outlined for matrix completion, a straightforward extension of this approach also applies to the tensor completion case.

LOW-RANK PROMOTING DATA ORGANIZATION

Before we can apply matrix completion techniques to interpolate $\mathbf{F}$, our unvectorized frequency slice of fully-sampled data, we must deal with the following issues. First, in the original (src, rec)

domain, the missing sources operator, $\mathcal{A}$, removes columns from $\mathbf{F}$, which does not increase the rank of the matrix since removing random *columns* or *rows* results in singular values being set to *zero* at the low end, which is undesireable from the matrix completion point of view.

Second, $\mathbf{F}$ itself also has high rank, owing to the presence of strong diagonal entries (zero offset energy) and subsequent off-diagonal oscillations. The previous theory indicates that naively applying matrix completion techniques in this domain will yield poor results. Simply put, we are missing two of the prerequisite signal recovery components in the (src, rec) domain, which we can see by plotting the decay of singular values in Figure 1. In light of the previous discussion on CS, we will examine different transformations under which the missing sources operator increases the singular values of our data matrix and hence promotes recovery in an alternative domain.

2D Seismic Data

In this case, we use the *Midpoint-Offset* transformation, which defines new coordinates for the matrix as

$$x_{\text{midpt}} = \frac{1}{2}(x_{\text{src}} + x_{\text{rec}})$$
$$x_{\text{offset}} = \frac{1}{2}(x_{\text{src}} - x_{\text{rec}}).$$

This coordinate transformation rotates the matrix $\mathbf{F}$ by 45 degrees, which is a tight frame operator with a nullspace, as depicted in Figure 2. If we denote this operator by $\mathcal{M}$, then $\mathcal{M}^*\mathcal{M} = I$, so transforming from (src, rec) to (midpt, offset) to (src, rec) returns the original signal, but $\mathcal{M}\mathcal{M}^* \neq I$, so the transformation from (midpt, offset) to (src, rec) and back again does not return the original signal. By using this transformation, we move the strong diagonal energy to a single column in the new domain, which mitigates the slow singular value decay in the original domain. Likewise, the

restriction operator $\mathcal{A}$ now removes super-/sub-*diagonals* from $\mathbf{F}$ rather than columns, demonstrated in Figure 2, which results in an overall increase in the singular values, as seen in Figure 1, placing the interpolation problem in a favourable recovery scenario as per the previous section. Our new optimization variable is $\tilde{\mathbf{X}} = \mathcal{M}(\mathbf{X})$, which is the data volume in the midpoint-offset domain, and our optimization problem is therefore

$$\begin{aligned} &\text{minimize } \|\tilde{\mathbf{X}}\|_* \\ &\quad \text{s.t. } \|\mathcal{A}\mathcal{M}^*(\tilde{\mathbf{X}}) - \mathbf{B}\|_F \leq \sigma. \end{aligned}$$

[Figure 1 about here.]

[Figure 2 about here.]

3D seismic data

Unlike in the matrix-case, there is no unique generalization of the SVD to tensors and as a result, there is no unique notion of *rank* for tensors. Instead, can consider the rank of different *matricizations* of $\mathbf{F}$. Instead of restricting ourselves to matricizations $\mathbf{F}^{(i)}$ where $i = x_{src}, y_{src}, x_{rec}, y_{rec}$, we can also look at the case where $i = \{x_{\text{src}}, y_{\text{src}}\}, \{x_{\text{src}}, x_{\text{rec}}\}, \{x_{\text{src}}, y_{\text{rec}}\}, \{y_{\text{src}}, x_{\text{rec}}\}$. That is, we group the dimensions of $\mathbf{F}$ specified by i and vectorize them along the rows $\mathbf{F}^{(i)}$ while vectorizing the other dimensions along the columns. Owing to the reciprocity relationship between sources and receivers in $\mathbf{F}$, we only need to consider two different matricizations of $\mathbf{F}$, which are depicted in Figure 3 and Figure 4. As we see in Figure 5, the $i = (x_{rec}, y_{rec})$ organization, that is, placing both receiver coordinates along the rows, results in a matrix that has *high* rank and the missing sources operator removes *columns* from the matrix. The overall singular value decay *decreases* as a result,

which is the worst possible scenario from the matrix completion perspective. On the other hand, the $i = (y_{src}, y_{rec})$ matricization yields *fast* decay of the singular values and a subsampling operator that causes the singular values to *increase*. This scenario is much closer to the idealized matrix completion sampling, which would correspond to the nonphysical process of randomly removing $(x_{src}, y_{src}, x_{rec}, y_{rec})$ points from $\mathbf{F}$. We note that this data organization has been considered in the context of solution operators of the wave equation in Demanet (2006), which applies to our case as our data volume $\mathbf{F}$ is the restriction of a Green's function to the acquisition surface.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

LARGE SCALE DATA RECONSTRUCTION

In this section, we explore the modifications necessary to extend matrix completion to 3D seismic data as well as compare this approach to an existing tensor-based technique. Matrix-completion techniques, after some modification, easily scale to interpolate large, multidimensional seismic volumes. Akin to the midpoint-offset transformation discussed earlier, we will detail an appropriate reordering of the data to promote the low-rank behaviour of our multidimensional seismic volumes.

The tensor completion technique of Kreimer et al. (2013) does not scale well to large problem sizes, owing to the need to compute SVDs of large-scale matrices. We will compare this tensor-based approach to the matrix completion approach, both in terms of recovery quality and speed, on both a simple synthetic data and a realistically sized synthetic frequency slice from a 3D survey provided

to us by the BG Group, and demonstrate the computational advantages of techniques that avoid computing SVDs over methods that make use of them. We will also contextualize the choice by Kreimer et al. (2013) to interpolate the data volume in the midpoint-offset domain within our signal recovery framework, which has previously not been justified.

In the subsequent sections, we will consider various schemes for interpolating a frequency slice $\mathbf{F}$ with source and receiver coordinates $(x_{src}, y_{src}, x_{rec}, y_{rec})$. We will also use $i = 1, \ldots, 4$ to refer to the i^{th} dimension of this frequency slice.

Large scale matrix completion

For the matrix completion approach, the limiting component for large scale data is that of the *nuclear norm projection*. As mentioned in Appendix A, the projection on to the set $\|\mathbf{X}\|_* \leq \tau$ requires the computation of the SVD of $\mathbf{X}$, which is prohibitively expensive when $\mathbf{X}$ has thousands or even millions of rows and columns. On the assumption that $\mathbf{X}$ is approximately low-rank at a given iteration, other authors such as Stoll (2012) compute a partial SVD using a Krylov approach, which is still cost-prohibitive for large matrices.

We can avoid the need for the expensive computation of SVDs via a well known *factorization* of the nuclear norm. Specifically, we have the following characterization of the nuclear norm, due to Srebro (2004),

$$\|\mathbf{X}\|_* = \underset{\mathbf{L},\mathbf{R}}{\text{minimize}} \ \frac{1}{2}(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2)$$

$$\text{subject to } \mathbf{X} = \mathbf{L}\mathbf{R}^T.$$

This allows us to write $\mathbf{X} = \mathbf{L}\mathbf{R}^T$ for some placeholder variables $\mathbf{L}$ and $\mathbf{R}$ of a prescribed rank k. Therefore, instead of projecting on to $\|\mathbf{X}\|_* \leq \tau$, we can instead project on to the *factor* ball

$\frac{1}{2}(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) \leq \tau$. This factor ball projection only involves computing $\|\mathbf{L}\|_F^2$, $\|\mathbf{R}\|_F^2$ and scaling the factors by a constant, which is substantially cheaper than computing the SVD of $\mathbf{X}$.

Equipped with this factorization approach, we can still use the basic idea of SPGℓ_1 to flip the objective and the constraints. The factorized LS_τ subproblems can be solved *much* more efficiently, while still maintaining the quality of the solution. As mentioned previously, the resulting algorithm is dubbed SPG-LR by Aravkin et al. (2014). This reformulation allows us to apply these matrix completion techniques to large scale seismic data interpolation, whereas SVD-based approaches such as Equation 3 are much more inefficient for even moderately sized problems. By using this approach in the previous section on 2D data, we were able to vastly outperform curvelet-based techniques.

This factorization turns the convex problem LS$_\tau$ posed in terms of $\mathbf{X}$ into a nonconvex problem in terms of the variables $\mathbf{L}, \mathbf{R}$, so there is a possibility for local minima or non-critical stationary points to arise when using this approach. As it turns out, as long as the prescribed rank k is larger than the rank of the optimal $\mathbf{X}$, any local minima encountered in the factorized problem is actually a global minimum (Burer and Monteiro, 2005; Aravkin et al., 2014). The possibility of non-critical stationary points is harder to discount, and remains an open problem. There is preliminary analysis indicating that initializing $\mathbf{L}$ and $\mathbf{R}$ so that $\mathbf{L}\mathbf{R}^T$ is sufficiently close to the true $\mathbf{X}$ will ensure that this optimization program will converge to the true solution (Sun and Luo, 2014) .

An alternative approach to factorized $BPDN_\sigma$ is to solve the QP_λ problem with the factorization scheme detailed above. The authors in Recht and Ré (2013) exploit the resulting *independance* of various subblocks of the $\mathbf{L}$ and $\mathbf{R}$ factors to create a partitioning scheme that updates components of these factors in parallel, resulting in a parallel matrix completion framework dubbed Jellyfish. By using this Jellyfish approach, each QP_λ problem for fixed λ and fixed internal rank k can be solved

very efficiently and cross-validation techniques can choose the optimal λ and rank parameters. We will compare the performance and recovery quality of the aforementioned techniques on 5D seismic data volumes in future sections.

Large scale tensor completion

Following the approach of Kreimer et al. (2013), which applies the method developed in Gandy et al. (2011) to seismic data, we can also exploit the *tensor* structure of a frequency slice $\mathbf{F}$ for interpolating missing traces. Specifically, let us consider different *matricizations* of $\mathbf{F}$, denoted $\mathbf{F}^{(i)}$ for $i = 1, \ldots, 4$, which is the matrix that reshapes $\mathbf{F}$ so that the i^{th} dimension is placed along the rows and the other dimensions are placed along the columns.

We now stipulate that each matricization $\mathbf{F}^{(i)}$ has low-rank, and thus we can proceed in an analogous way to the matrix completion case in order to interpolate the tensor $\mathbf{F}$ by solving the following problem

$$\begin{aligned} &\underset{\mathbf{F}}{\text{minimize}} \sum_{i=1}^{4} \|\mathbf{F}^{(i)}\|_* \\ &\text{subject to } \|\mathcal{A}(\mathbf{F}) - \mathbf{B}\|_2 \leq \sigma, \end{aligned}$$

i.e. look for the tensor $\mathbf{F}$ that has *simultaneously* the lowest rank in each matricization $\mathbf{F}^{(i)}$ that fits the subsampled data $\mathbf{B}$. Here $\mathcal{A}$ is our subsampling operator, typically missing sources. In the case of Kreimer et al. (2013), this interpolation is performed in the $(x_{\text{midpt}}, y_{\text{midpt}}, x_{\text{offset}}, y_{\text{offset}})$ domain on each frequency slice, which we also employ in our later experiments.

To solve this problem, the authors in Kreimer et al. (2013) use the Douglas-Rachford variable splitting technique that creates 4 additional copies of the variable $\mathbf{F}$, denoted $\mathbf{X}_i$, with each copy

corresponding to each matricization $\mathbf{F}^{(i)}$. This is an inherent feature of this approach to solve convex optimization problems with coupled objectives/constraints and thus cannot be avoided or optimized. The authors then use an Augmented Lagrangian approach to solve the *decoupled* problem

$$\underset{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathbf{X}_4,\mathbf{F}}{\text{minimize}} \sum_{i=1}^{4} \|\mathbf{X}_i\|_* + \lambda\|\mathcal{A}(\mathbf{F}) - \mathbf{B}\|_2^2 \tag{3}$$

$$\text{subject to } \mathbf{X}_i = \mathbf{F}^{(i)} \text{ for } i = 1, \ldots, 4.$$

The resulting problem is convex, and thus has a unique solution. We refer to this method as the alternating direction method of multipliers (ADMM) tensor method. This variable splitting technique can be difficult to implement for realistic problems, as the tensor $\mathbf{F}$ is often unable to be stored fully in working memory. Given the large number of elements of $\mathbf{F}$, creating at minimum four extraneous copies of $\mathbf{F}$ can quickly overload the storage and memory of even a large computing cluster. Moreover, there are theoretical and numerical results that state that this problem formulation is in fact no better than imposing the nuclear norm penalty on a *single* matricization of $\mathbf{F}$, at least in the case of Gaussian measurements (Oymak et al., 2012; Signoretto et al., 2011). We shall see a similar phenomenon in our subsequent experiments.

Penalizing the nuclear norm in this fashion, as in all methods that use an explicit nuclear norm penalty, scales very poorly as the problem size grows. Solving these optimization problems requires hundreds or thousands of soft thresholding operations of the singular values of $\mathbf{X}$. The main computational costs of soft thresholding is computing the SVD of a relatively large matrix, which has computational complexity $O(n^3)$ for an $n \times n$ matrix. When our data $\mathbf{F}$ has four or more dimensions, the cost of computing the SVD of one of its matricizations easily dominates the overall computational costs of the method. Applying this operation four times per iteration in the above problem, as is required due to the variable splitting, prevents this technique from performing

efficiently for large realistic problems.

EXPERIMENTS

We perform seismic data interpolation on five different data sets. For our 2D data sets, we show the efficacy of proposed formulation on three different seismic lines. The first data set, which is a shallow-water marine scenario, is from the Nelson field provided to us by PGS. The Nelson data set contains 401×401 sources and receivers. The temporal length of the data is 4s with a sampling interval of 0.004s. The first four seconds of a single shot gather are shown in Figure 6. The second synthetic data set is from the Gulf of Mexico (GOM) and is provided to us by the Chevron. It contains 3201 sources and 801 receivers with a spatial interval of 25m. The third data set (Figure 17) is simulated on a synthetic velocity model (see Berkhout and Verschuur (2006)) using IWave (Symes et al., 2011). An anticline salt structure over-lies the target, i.e., a fault structure. A seismic line is modelled using a fixed-spread configuration where sources and receivers are placed at an interval of 15m. This results in a data set of 361×361 sources and receivers.

Our 3D examples consist of two different data sets. The first data set is generated on a synthetic single-layer model. This data set has 50 sources and 50 receivers and we use a frequency slice at 4 Hz. This simple data set allows us to compare the running time of the various algorithms under consideration. The Compass data set is provided to us by the BG Group and is generated from an unknown but geologically complex and realistic model. We selected a few 4D monochromatic frequency slices from this data set at 4.68, 7.34, and 12.3Hz. Each monochromatic frequency slice has 401×401 receivers spaced by 25m and 68×68 sources spaced by 150m.

2D Seismic data

In this section, we compare matrix-completion based techniques to existing curvelet-based interpolation for interpolating 2D seismic data and demonstrate the inability of the matrix completion approach to recover subsampled signals if the three previously discussed conditions are not satisfied. For details on the curvelet-based reconstruction techniques, we refer to (Herrmann and Hennenfent, 2008; Mansour et al., 2013). For concreteness, we concern ourselves with the missing-sources scenario, although the missing-receivers scenario is analogous. In all the experiments, we set the data misfit parameter σ to be equal to $\eta\|\mathbf{B}\|_F$ where $\eta \in (0, 1)$ is the fraction of the input data energy to fit.

Nelson data set

Here $n_s = 0.5N_s, 0.25N_s$, so we remove 50%, 75% of the sources, respectively. For the sake of comparing curvelet-based and rank-minimization based reconstruction methods on identical data, and because the matrix completion algorithm requires a 2D input, we first interpolate a single 2D frequency slice using curvelets and with matrix completion in Figure 8 and Figure 9. When working with frequency slices, Mansour et al. (2013) proposed using weighted ℓ_1 minimization, wherein the authors exploit the correlation between the locations of significant curvelet coefficients across different partitions, such as shot records, common-offset gathers, or frequency slices of the acquired data, to improve the reconstruction quality. The authors showed that the best recovery is achieved across frequencies in the midpoint-offset domain, owing to the increased curvelet sparsity in the midpoint-offset domain compared to the source-receiver domain. Therefore, in order to draw a fair comparison with the matrix-based methods, we perform curvelet-based reconstruction in the

midpoint-offset domain.

We summarize these results of interpolating a single 2D frequency slice in Table 1. Compared to the costs associated to applying the forward and adjoint curvelet transform, SPG-LR is much more efficient and, as such, this approach significantly outperforms the ℓ_1-based curvelet interpolation. Both methods perform similarly in terms of reconstruction quality for low frequency slices, since these slices are well represented both as a sparse superposition of curvelets and as a low-rank matrix. High frequency data slices, on the other hand, are empirically high rank, which can be shown explicitly for a homogeneous medium as a result of Lemma 2.7 in Engquist and Ying (2007), and we expect matrix completion to perform more less well in this case since for a simple geological models, even-though high frequency contains oscillations away from the zero-offset. On the other hand, these oscillations can be well approximated by low-rank values. To perform the reconstruction of seismic data in the high frequency regime, Kumar et al. (2013) proposed to represent the matrix in the Hierarchical semi-separable (HSS) format, wherein data is first windowed in off-diagonal and diagonal blocks and the diagonal blocks are recursively partitioned. The interpolation is then performed on each subset separately. In the interest of brevity, we omit the inclusion of this approach here. Additionally, since the high frequency slices are very oscillatory, they are much less sparse in the curvelet dictionary.

Owing to the significantly faster performance of matrix completion compared to the curvelet-based method, we apply the former technique to an entire seismic data volume by interpolating each frequency slice in the 5-85Hz band. Figures 11 show the interpolation results in case of 75% missing traces. In order to get the best rank values to interpolation the full seismic line, we first performed the interpolation for the frequency slices at 10 Hz and 60 Hz. The best rank value we

get for these two slices is 30 and 60 respectively. Keeping this in mind, we work with all of the monochromatic frequency slices and adjust the rank linearly from 30 to 60 when moving from low to high frequencies. The running time is 2 h 18 min using SPG-LR on a 2 quad-core 2.6GHz Intel processor with 16 GB memory and implicit multithreading via LAPACK libraries. We can see that we have low-reconstruction error with little coherent energy in the residual when 75% of the sources are missing. Figure 12 shows the qualitative measurement of recovery for all frequencies in the energy-band. We can further mitigate such coherent residual energy by exploiting additional structures in the data such as symmetry, as in Kumar et al. (2014).

Remark

It is important to note that if we omit the first two principles of matrix completion by interpolating the signal in the source-receiver domain, as discussed previously, we obtain very poor results, as shown in Figure 10. Similar to CS-based interpolation, choosing an appropriate transform-domain for matrix and tensor completion is vital to ensure successful recovery.

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Table 1 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

Gulf of Mexico data set

In this case, we set $0.2N_s$ and so we remove 80% of the sources. Here, we perform the interpolation on a frequency spectrum of 5-30hz. Figure 14 shows the comparison of the reconstruction error using sparsity and rank-minimization based approaches for a frequency slice at 7Hz and 20 Hz. In each case we remove the 80% of the sources. For visualization purposes, we only show a subset of interpolated data corresponding to the square block in Figure 13, but we interpolate the monochromatic slice over all sources and receivers. Even in the highly sub-sampled case of 80%, we are still able to recover to a high SNR of 14.2 dB, 10.5dB, respectively, but we start losing coherent energy in the residual as a result of the high-subsampling ratio. These results indicate that even in complex geological environments, low-frequencies are still low-rank in nature. This can also be seen since, for a continuous function, the smoother the function is (i.e., the more derivatives it has), the faster its singular values decay (see, for instance, Chang and Ha (1999)). For comparison purposes, we plot the frequency-wavenumber spectrum of the 20Hz frequency slice in Figure 15 along with the corresponding spectra of matrix with 80% of the sources removed periodically and uniform randomly. In this case, the volume is approximately three times aliased in the bandwidth of the original signal for periodic subsampling, while the randomly subsampled case has created noisy aliases. The average sampling interval for both schemes is the same. As shown in this figure, the interpolated matrix has a significantly improved spectrum compared to the input. Figure 16 shows the interpolation result over a common receiver gather using rank-minimization based techniques.

In this case, we set the rank parameter to be 40 and use the same rank for all the frequencies. The running time on a single frequency slice in this case is 7 min using SPG-LR and 1320 min using curvelets.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

[Figure 16 about here.]

Synthetic fault model

In this setting, we remove 80% of the sources and display the results in Figure 17. For simplicity, we only perform rank-minimization based interpolation on this data set. In this case we set the rank parameter to be 30 and used the same for all frequencies. Even though the presence of faults make the geological environment complex, we are still able to successfully reconstruct the data volume using rank-minimization based techniques, which is also evident in the low-coherency of the data residual (Figure 17).

[Figure 17 about here.]

3D Seismic data

Single-layer reflector data

Before proceeding to a more realistically sized data set, we first test the performance of the SPG-LR matrix completion and the tensor completion method of Kreimer et al. (2013) on a small, synthetic data set generated from a simple, single-reflector model. We only use a frequency slice at 4 Hz. We normalize the volume to unit norm and randomly remove 50% of the sources from the data.

For the alternating direction method of multipliers (ADMM) tensor method, we complete the data volumes in the midpoint-offset domain, which is the same domain used in Kreimer et al. (2013). In the context of our framework, we note that the midpoint-offset domain for recovering 3D frequency slices has the same recovery-enhancing properties as for recovering 2D frequency slices, as mentioned previously. Specifically, missing source sampling tends to *increase* the rank of the individual source and receiver matricizations in this domain, making completion via rank-minimization possible in midpoint-offset compared to source-receiver. In the original source-receiver domain, removing (x_{src}, x_{rec}) points from the tensor does *not* increase the singular values in the x_{src} and x_{rec} matricizations and hence the reconstruction quality will suffer. On the other hand, for the matrix completion case, the midpoint-offset conversion is a tight frame that acts on the left and right singular vectors of the matricized tensor $\mathbf{F}^{(x_{\mathrm{src}},x_{\mathrm{rec}})}$ and thus does not affect the rank for this particular matricization. Also in this case, we consider the effects of windowing the input data on interpolation quality and speed. We let ADMM-w denote the ADMM method with a window size of w with an additional overlap of approximately 20%. In our experiments, we consider $w = 10$ (small windows), $w = 25$ (large windows), and $w = 50$ (no windowing).

In the ADMM method, the two parameters of note are λ, which control the relative penalty between data misfit and nuclear norm, and β, which controls the speed of convergence of the individual matrices $X^{(i)}$ to the tensor $\mathbf{F}$. The λ, β parameters proposed in Kreimer et al. (2013) do not appear to work for our problems, as using the stated parameters penalizes the nuclear norm terms too much compared to the data residual term, resulting in the solution tensor converging to $\mathbf{X} = 0$. Instead, we estimate the optimal λ, β parameters by cross validation, which involves removing 20% of the 50% known sources, creating a so-called "test set", and using the remaining data points as input data. We use various combinations of λ, β to solve Problem 3, using 50 iterations, and compare the SNR of the interpolant on the test set in order to determine the best parameters, i.e. we estimate the optimal λ, β without reference to the unknown entries of the tensor. Owing to the large computational costs of the "no window" case, we scan over the values of λ increasing exponentially and fix $\beta = 0.5$. For the windowed cases, we scan over exponentially increasing values of λ, β for a single window and use the estimated λ, β for interpolating the other windows. For the SPG-LR, we set our internal rank parameter to be 20 and allow the algorithm to run for 1000 iterations. As shown in Aravkin et al. (2014), as long as the chosen rank is sufficiently large, further increasing the rank parameter will not significantly change the results. We summarize our results in Table 2 and display the results in Figure 18.

Even disregarding the time spent selecting ideal parameters, SPG-LR matrix completion drastically outperforms the ADMM method on this small example. The main computational costs of the ADMM tensor approach are a result of the need to compute SVDs on large matrices, which the low-rank factorization approach does not require. Moreover, the low-rank factorization creates minimal additional copies of the full tensor, whereas the ADMM tensor approach creates at minimum 5 copies of the full tensor by design. The tensor-based, per-dimension windowing approach

also degrades the overall reconstruction quality, as the algorithm is unable to take advantage of the redundancy of the full data volume once the windows are sufficiently small. There is a very prominent tradeoff between recovery speed and reconstruction quality as the size of the windows become smaller, owing to the expensive nature of the ADMM approach itself for large data volumes and the inherent redundancy in the full data volume that makes interpolation possible which is decreased when windowing.

[Table 2 about here.]

[Figure 18 about here.]

BG Compass data

Owing to the smoothness of the data at lower frequencies, we uniformly downsample the individual frequency slices in the receiver coordinates without introducing aliasing. This reduces the overall computational complexity while simultaneously preserving the recovery quality. The 4.64Hz, 7.34Hz and 12.3Hz slices were downsampled to 101×101, 101×101 and 201×201 receiver grids, respectively. For these problems, the data was subsampled along the source coordinates by removing 25%, 50%, and 75% of the shots.

In order to apply matrix completion without windowing on the entire data set, the data was organized as a matrix using the low-rank promoting organization described previously. We used Jellyfish and SPG-LR implementations to complete the resulting incomplete matrix and compared these methods to the ADMM Tensor method and LMaFit, an alternating least-squares approach to matrix completion detailed in Wen et al. (2012). LMaFit is a fast matrix completion solver that

avoids using nuclear norm penalization but must be given an appropriate rank parameter in order to achieve reasonable results. We use the code available from the author's website. SPG-LR, ADMM, and LMaFit were run on a 2 quad-core 2.6GHz Intel processor with 16 GB memory and implicit multithreading via LAPACK libraries while Jellyfish was run on a dual Xeon X650 CPU (6 x 2 cores) with 24 GB of RAM with explicit multithreading. The hardware configurations of both of these environments are very similar, which results in SPG-LR and Jellyfish performing comparably.

For the Jellyfish experiments, the model parameter μ, which plays the same role as the λ parameter above, and the optimization parameters (initial step size and step decay) were selected by validation, which required 120 iterations of the optimization procedure for each (frequency, subsampling ratio) pair. The maximum rank value was set to the rank value used in the SPG-LR results. For the SPG-LR experiments, we interpolate a subsection of the data for various rank values and arrived at 120, 150 and 200 as the best rank parameters for each frequency. We perform the same validation techniques on the rank parameter k of LMaFit. In order to focus solely on comparing computational times, we omit reporting the parameter selection times for the ADMM method.

The results for 75% missing sources in Figure 19 demonstrate that, even in the low subsampling regime, matrix completion methods can successfully recover the missing shot data at these low frequencies. Table 3 gives an extensive summary of our results for different subsampling ratios and frequencies. The comparative results between Jellyfish and SPG-LR agree the with the theoretical results that establish the equivalence of BPDN_σ and QP_λ formulations. The runtime values include the parameter estimation procedure, which was carried out individually in each case. As we have seen previously, the ADMM approach does not perform well both in terms of computational time

and in terms of recovery owing to the extra variable creation and the computation of SVDs.

In our experiments, we noticed that successful parameter combinations work well for other problems too. Hence we can argue that in a real-world problem, once a parameter combination is selected, it can be used for different instances or it can be used as an initial point for a local parameter search.

Matrix Completion with Windowing

When windowing the data, we use the same matricizations of the data as discussed previously, but now split the volume in to nonoverlapping windows. We now use matrix completion on the resulting windows of data individually. We used Jellyfish for matrix completion on individual windows. Again, we use cross validation to select our parameters. We performed the experiments with two different window sizes. For the small window case, the matricization was partitioned into 4 segments along rows and columns, totalling 16 windows. For the large window case, the matricization was split into 16 segments along rows and columns, yielding 256 windows. This windowing is distinctly different from the windowing explored for the single-layer model, since here we are windowing the matricized form of the tensor, in the $(x_{\text{src}}, x_{\text{rec}})$ unfolding, as opposed to the per-dimension windowing in the previous section. The resulting windows created in this way contain much more sampled data than in the tensor-windowing case yet are still small enough in size to be processed efficiently.

The results in Figure 20 suggest that for this particular form of windowing, the matrix completion results are particularly degraded by only using small windows of data at a time. As mentioned previously, since we are relying on a high redundancy (with respect to the SVD) in the underlying and sampled data to perform matrix completion, we are reducing the overall redundancy of the input

data by partitioning it. On the other hand, the real-world benefits of windowing in this context become apparent when the data cannot be fit into the memory at the cost of reconstruction quality. In this case, windowing allows us to partition the problem, offsetting the I/O cost that would result from memory paging. Based on these results, whenever possible, we strive to include as much data as possible in a given problem in order to recover the original matrix/tensor adequately.

[Figure 19 about here.]

[Figure 20 about here.]

[Table 3 about here.]

DISCUSSION

As the above results demonstrate, the $\mathbf{L}, \mathbf{R}$ matrix completion approach significantly outperforms the ADMM tensor-based approach due to the need to avoid the computation of SVDs as well as the minimal duplication of variables compared to the latter method. For the simple synthetic data, the ADMM method is able to achieve a similar recovery SNR to matrix completion, albeit at a much larger computational cost. For realistically sized data sets, the difference between the two methods can mean the difference between hours and days to produce an adequate result. In terms of the difference between SPG-LR and Jellyfish matrix completion, both return results that are similar in quality, which agrees with the fact that they are both based off of $\mathbf{L}, \mathbf{R}$ factorizations and the ranks used in these experiments are identical. Compared to these two methods, LMaFit converges much faster for regimes when there is more data available, while producing a lower quality result. When there is very little data available, as is typical in realistic seismic acquisition scenarios, the

algorithm has issues converging. We note that, since it is a sparse linear algebra method, Jellyfish tends to outperform SPG-LR when the number of missing traces is high. This sparse linear algebra approach can conceivably be employed with the SPG-LR machinery. In these examples, we have not made any attempts to explicitly parallelize the SPG-LR or ADMM methods, instead relying on the efficient dense linear algebra routines used in Matlab, whereas Jellyfish is an inherently parallelized method.

Without automatic mechanisms for parameter selection as in SPG-LR, the Jellyfish, ADMM, and LMaFit algorithms rely on cross-validation techniques that involve solving many problem instances at different parameters. The inherently parallel nature of Jellyfish allows it to solve each problem instance very quickly and thus achieves very similar performance to SPG-LR. LMaFit has very fast convergence when there is sufficient data, but slows down significantly in scenarios with very little data. The ADMM method, on the other hand, scales much more poorly for large data volumes and spends much more time on parameter selection than the other methods. However, in practice, we can assume that across frequency slices, say, optimally chosen parameters for one frequency slice will likely work well for neighbouring frequency slices and thus the parameter selection time can be amortized over the whole volume.

In our experiments, aside from the simple 3D layer model and the Nelson dataset, the geological models used were not low rank. That is to say, the models had complex geology and were not simply horizontally layered media. Instead, through the use of these low rank techniques, we are exploiting the low rank structure of the data volumes, not merely any low rank structure present in the models themselves. As the temporal frequency increases, the inherent rank of the resulting frequency slices increases, which makes low rank interpolation more challenging. Despite this observation, we still

achieve reasonable results for higher frequencies using our methods.

As predicted by our theoretical considerations, the choice of windowing in this case has a negative effect on the generated results in the situation where the earth model permits a low-rank representation that is reflected in the midpoint-offset domain. In case of earth models that are not inherently low-rank, such as those with salt bodies, we can still recover the low-frequency slices as shown by the examples without performing the windowing on the data sets. As a general rule of thumb, we advise to incorporate as much of the input data is possible in to a given matrix-completion problem but there is a tradeoff between the size of the data windows, the amount of memory available to process such volumes, and the inherent complexity of the model. Additionally, one should avoid methods that needlessly create extraneous copies of the data when working with large scale volumes.

Here we have also demonstrated the importance of theoretical components for signal recovery using matrix and tensor completion methods. By ignoring these principles of matrix completion, a practitioner can unintentionally find herself in a disadvantageous scenario and produce sub-optimal results without a guiding theory to remedy the situation. However, by choosing an appropriate transform domain in which to complete the matrix or tensor, we can successfully employ this rank-minimizing machinery to interpolate a signal with missing traces in a computationally efficient manner.

From a practitioner's point of view, the purpose of this interpolation machinery is to remove the acquisition footprint from missing-trace data that is used in further downstream seismic processes such as migration and full waveform inversion. These techniques can help mitigate the lack of data coverage in certain areas that would otherwise have created artifacts or non-physical regions in a

seismic image.

CONCLUSION

Building upon existing knowledge of compressive sensing as a successful signal recovery paradigm, this work has outlined the necessary components of using matrix and tensor completion methods for interpolating large-scale seismic data volumes. As we have demonstrated numerically, without the necessary components of a low-rank domain, a rank-increasing sampling scheme, and a rank-minimizing optimization scheme, matrix completion-based techniques cannot successfully recover subsampled seismic signals. Once all of these ingredients are in place, however, we can use existing convex solvers to recover the fully sampled data volume. Since such solvers invariably involve computing singular-value decomposition of large matrices, we have presented two alternative factorized-based formulations that scale much more efficiently than their strictly convex counterparts when the data volumes are large. We have shown that our factorization-based matrix completion approach is very competitive compared to existing curvelet-based methods for 2D seismic data and alternating direction method of multipliers tensor-based methods for 3D seismic data.

From a practical point of view, this theoretical framework is exceedingly flexible. We have shown the effectiveness of midpoint-offset organization for 2D data and $(x_{\text{source}}, x_{\text{receiver}})$ matrix organization for 3D data for promoting low-rank structure in the data volumes but it is conceivable that other seismic data organizations could also be useful in this regard, e.g., midpoint-offset-azimuth. Our optimization framework also allows us to operate on both large-scale data *without* having to select a large number of parameters and we do not need to recourse to using small windows of data, which may degrade the recovery results. In the seismic context, reusing the interpolated

results from lower frequencies as a warm-start for interpolating data at higher frequencies can further reduce the overall computational costs. The proposed approach to matrix completion and the Jellyfish method are very promising for large scale data sets and can conceivably be applied to interpolate wide azimuth data sets as well.

ACKNOWLEDGEMENTS

We would like to thank PGS for permission to use the Nelson dataset, Chevron for permission to use the synthetic Gulf of Mexico dataset, and the BG Group for permission to use the synthetic Compass dataset. This work was financially supported in part by the NSERC of Canada Discovery Grant (RGPIN 261641-06) and the CRD Grant DNOISE II (CDRP J 375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, Chevron, ConocoPhillips, CGG, ION GXT, Petrobras, PGS, Statoil, Subsalt Solutions, Total SA, WesternGeco, and Woodside.

Appendix A. *SPGLR*

The insight from the $SPG\ell_1$ solver (van den Berg and Friedlander, 2008) allows us to solve the $BPDN_\sigma$ formulation by solving a sequence of *simpler* problems. By reversing the objective function and its constraints, we solve a sequence of *Lasso* subproblems,

$$\begin{aligned} v(\tau) := \underset{\mathbf{X}}{\text{minimize}} \ & \|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_F^2 \\ \text{ssubject to } & \|\mathbf{X}\|_* \leq \tau. \end{aligned} \tag{LS$_\tau$}$$

Above, $v(\tau)$ is the (minimum) value function of the data misfit that satisfies the constraint $\|\mathbf{X}\|_* \leq \tau$. We refer to this program, parametrized by the nuclear norm value τ, as LS$_\tau$. In order to compute

$v(\tau)$, we employ the *projected gradient method*, which can be written as

$$\mathbf{X}_{k+1} = P_{\|\mathbf{X}\|_* \leq \tau}(\mathbf{X}_k - \alpha_k \mathcal{A}^*(\mathcal{A}(\mathbf{X}_k) - \mathbf{B})),$$

i.e., a negative gradient step, with steplength α_k, followed by a projection on to the nuclear norm ball $\|\mathbf{X}\|_* \leq \tau$. The projection on to $\|\mathbf{X}\|_* \leq \tau$ involves computing an SVD of the matrix, which is cost-prohibitive when $\mathbf{X}$ is large. As shown above, we can relax the requirement to project exactly on to the set $\|\mathbf{X}\|_* \leq \tau$ by an appropriate *factorization* approach, which allows us to apply this method to large-scale data volumes. By writing $\mathbf{X} = \mathbf{L}\mathbf{R}^T$, we replace the projection on to the nuclear norm ball by the projection on to the $\mathbf{L}, \mathbf{R}$ ball

$$\frac{1}{2}(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) \leq \tau.$$

This projection merely involves computing the Frobenius norm of the factors $\mathbf{L}, \mathbf{R}$ and scaling them if they are larger than τ. The gradient in this case is computed in terms of $\mathbf{L}$ and $\mathbf{R}$ and refer to Aravkin et al. (2014) for the specific details.

The overarching idea proposed in van den Berg and Friedlander (2008), and generalized by van den Berg and Friedlander (2011); Aravkin et al. (2013) to a wide class of regularizers and penalties, is as follows. The *value function* $v(\tau)$ is convex and monotonically decreasing when the regularizer and misfit penalty are convex, and differentiable for a wide class of problems, including (LS_τ). In order to find a solution to $BPDN_\sigma$, we need only find a value of τ such that $v(\tau) = \sigma$. Application of Newton's root-finding method gives us the iteration

$$\tau_{k+1} = \tau_k - \frac{v(\tau_k) - \sigma}{v'(\tau_k)}. \tag{4}$$

Evaluating $v(\tau_k)$ requires solving the optimization problem that defines this function (LS_τ), and results from van den Berg and Friedlander (2008) give us $v'(\tau_k)$, the derivative of $v(\tau)$ evaluated

at τ_k, in closed form, once a solution has been found. LS_τ is much easier to solve than $BPDN_\sigma$, since its objective is differentiable and the constraint is simple. Since σ can be estimated ahead of time in many applications, this reduces the computational burden one incurs in estimating the λ parameter, and makes the approach better suited for large scale problems.

The main steps of SPG-LR are as follows:

Require: Data misfit level σ, subsampling operator $\mathcal{A}$, input data $\mathbf{B}$.

$\tau_1 \leftarrow$ initial guess, $k \leftarrow 1$

repeat

Compute $v(\tau_k)$ by solving LS_τ using a projected gradient method,

using the previously computed solution to $v(\tau_{k-1})$ as an initial guess.

Update τ_{k+1} by Equation 4.

$k \leftarrow k + 1$

until $v(\tau_k) \leq \sigma$

REFERENCES

Aravkin, A., J. Burke, and M. Friedlander, 2013, Variational properties of value functions: SIAM Journal on Optimization, **23**, 1689–1717.

Aravkin, A. Y., R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, 2014, Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation.: SIAM Journal on Scientific Computing, **To appear**.

Bardan, V., 1987, Trace interpolation in seismic data processing: Geophysical Prospecting, **35**, 343–358.

Berkhout, A., and D. Verschuur, 2006, Imaging of multiple reflections: Geophysics, **71**, SI209–SI220.

Burer, S., and R. D. Monteiro, 2005, Local minima and convergence in low-rank semidefinite programming: Mathematical Programming, **103**, 427–444.

Cai, J.-F., E. J. Candès, and Z. Shen, 2010, A singular value thresholding algorithm for matrix completion: SIAM Journal on Optimization, **20**, 1956–1982.

Candès, E. J., and B. Recht, 2009, Exact matrix completion via convex optimization: Foundations of Computational mathematics, **9**, 717–772.

Canning, A., and G. Gardner, 1998, Reducing 3-d acquisition footprint for 3-d dmo and 3-d prestack migration: Geophysics, **63**, 1177–1183.

Chang, C.-H., and C.-W. Ha, 1999, Sharp inequalities of singular values of smooth kernels: Integral Equations and Operator Theory, **35**, 20–27.

Curry, W. J., 2009, Interpolation with fourier radial adaptive thresholding: SEG Technical Program Expanded Abstracts 2009, 3259–3263.

Da Silva, C., and F. J. Herrmann, 2014, Optimization on the hierarchical tucker manifold-applications to tensor completion.

Demanet, L., 2006, Curvelets, wave atoms, and wave equations: PhD thesis, California Institute of Technology.

Donoho, D., 2006, Compressed sensing.: IEEE Transactions on Information Theory, **52**, 1289–1306.

Duijndam, A., M. Schonewille, and C. Hindriks, 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: Geophysics, **64**, 524–538.

Engquist, B., and L. Ying, 2007, Fast directional multilevel algorithms for oscillatory kernels: SIAM Journal on Scientific Computing, **29**, 1710–1737.

Gandy, S., B. Recht, and I. Yamada, 2011, Tensor completion and low-n-rank tensor recovery via convex optimization: Inverse Problems, **27**, 025010.

Giryes, R., M. Elad, and Y. C. Eldar, 2011, The projected gsure for automatic parameter tuning in iterative shrinkage methods: Applied and Computational Harmonic Analysis, **30**, 407–422.

Halko, N., P.-G. Martinsson, and J. A. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: SIAM Review, **53**, 217–288.

Hennenfent, G., and F. J. Herrmann, 2006a, Application of stable signal recovery to seismic data interpolation: SEG Technical Program Expanded Abstracts, SEG, 2797–2801.

———, 2006b, Seismic denoising with nonuniformly sampled curvelets: Computing in Science & Engineering, **8**, 16–25.

———, 2008, Simply denoise: wavefield reconstruction via jittered undersampling: Geophysics, **73**, V19–V28.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.

Herrmann, F. J., and X. Li, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical Prospecting, **60**, 696–712.

Kabir, M. N., and D. Verschuur, 1995, Restoration of missing offsets by parabolic radon transform1: Geophysical Prospecting, **43**, 347–368.

Kanagal, B., and V. Sindhwani, 2010, Rank selection in low-rank matrix approximations: A study of cross-validation for nmfs.

Kreimer, N., and M. Sacchi, 2012a, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, V113–V122.

Kreimer, N., and M. D. Sacchi, 2012b, Tensor completion via nuclear norm minimization for 5d seismic data reconstruction: SEG Technical Program Expanded Abstracts, 1–5.

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction: Geophysics, **78**, V273–V284.

Kumar, R., A. Y. Aravkin, E. Esser, H. Mansour, and F. J. Herrmann, 2014, SVD-free low-rank matrix factorization : wavefield reconstruction via jittered subsampling and reciprocity: Presented at the EAGE.

Kumar, R., H. Mansour, A. Y. Aravkin, and F. J. Herrmann, 2013, Reconstruction of seismic wavefields via low-rank matrix factorization in the hierarchical-separable matrix representation: SEG Technical Program Expanded Abstracts, 3628–3633.

Lee, J., B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp, 2010, Practical large-scale optimization for max-norm regularization: Presented at the Advances in Neural Information Processing Systems, 2010.

Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: Geophysics, **77**, A13–A17.

Liberty, E., F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, 2007, Randomized algorithms for the low-rank approximation of matrices: **104**, 20167–20172.

Mahoney, M. W., 2011, Randomized algorithms for matrices and data: Foundations and Trends in Machine Learning, **3**, 123–224.

Mansour, H., F. J. Herrmann, and O. Yilmaz, 2013, Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization: Geophysics, **78**, V193–V206.

Mansour, H., H. Wason, T. T. Lin, and F. J. Herrmann, 2012, Randomized marine acquisition with compressive sampling matrices: Geophysical Prospecting, **60**, 648–662.

Mishra, B., G. Meyer, F. Bach, and R. Sepulchre, 2013, Low-rank optimization with trace norm penalty: SIAM Journal on Optimization, **23**, 2124–2149.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.

Owen, A. B., and P. O. Perry, 2009, Bi-cross-validation of the svd and the nonnegative matrix factorization: The Annals of Applied Statistics, **3**, 564–594.

Oymak, S., A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, 2012, Simultaneously structured models with application to sparse and low-rank matrices: CoRR, **abs/1212.3753**, 1–1.

Recht, B., 2011, A simpler approach to matrix completion: The Journal of Machine Learning Research, **7777777**, 3413–3430.

Recht, B., M. Fazel, and P. Parrilo, 2010, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization: SIAM Review, **52**, 471–501.

Recht, B., and C. Ré, 2013, Parallel stochastic gradient algorithms for large-scale matrix completion: Mathematical Programming Computation, **5**, 201–226.

Sacchi, M., S. Kaplan, and M. Naghizadeh, 2009, F-x gabor seismic data reconstruction: Presented at the 71st Annual International Conference and Exhibition, EAGE.

Sacchi, M., T. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution

discrete fourier transform: Signal Processing, IEEE Transactions on, **46**, 31 –38.

Sacchi, M. D., and B. Liu, 2005, Minimum weighted norm wavefield reconstruction for ava imaging: Geophysical Prospecting, **53**, 787–801.

Schaeffer, H., and S. Osher, 2013, A low patch-rank interpretation of texture: SIAM Journal on Imaging Sciences, **6**, 226–262.

Signoretto, M., R. Van de Plas, B. De Moor, and J. A. Suykens, 2011, Tensor versus matrix completion: a comparison with application to spectral data: Signal Processing Letters, IEEE, **18**, 403–406.

Sinha, S., P. S. Routh, P. D. Anno, and J. P. Castagna, 2005, Spectral decomposition of seismic data with continuous-wavelet transform: Geophysics, **70**, P19–P25.

Srebro, N., 2004, Learning with matrix factorizations, phd thesis.

Stoll, M., 2012, A krylov-schur approach to the truncated svd: Linear Algebra and its Applications, **436**, 2795 – 2806.

Sun, R., and Z.-Q. Luo, 2014, Guaranteed matrix completion via non-convex factorization: arXiv preprint arXiv:1411.8003.

Symes, W. W., D. Sun, and M. Enriquez, 2011, From modelling to inversion: designing a well-adapted simulator: Geophysical Prospecting, **59**, 814–833.

Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, **74**, V123–V132.

Trickett, S., L. Burroughs, and A. Milton, 2013, Interpolation using hankel tensor completion: Ratio, **1**, 16.

Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank reduction based trace interpolation: , SEG Technical Program Expanded Abstracts, 3829–3833.

van den Berg, E., and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.

———, 2011, Sparse optimization with least-squares constraints: SIAM Journal on Optimization, **21**, 1201–1229.

Wang, J., M. Ng, and M. Perz, 2010, Seismic data interpolation by greedy local radon transform: Geophysics, **75**, WB225–WB234.

Wang, L., J. Gao, W. Zhao, and X. Jiang, 2011, Nonstationary seismic deconvolution by adaptive molecular decomposition: Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, IEEE, 2189–2192.

Wen, Z., W. Yin, and Y. Zhang, 2012, Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm: Mathematical Programming Computation, **4**, 333–361.

Yang, Y., J. Ma, and S. Osher, 2013, Seismic data reconstruction via matrix completion: Inverse problem and imaging, **7**, 1379–1392.

LIST OF FIGURES

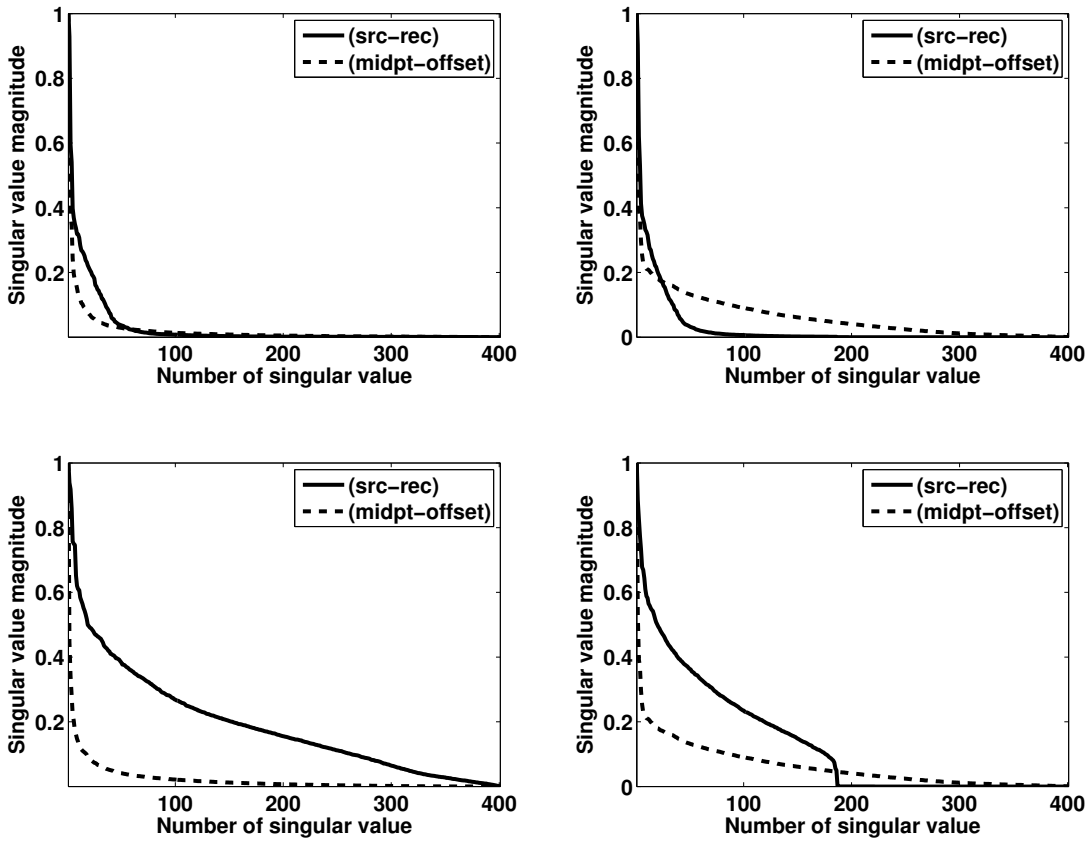


Figure 1: Singular value decay in the source-receiver and midpoint-offset domain. *Left* : fully sampled frequency slices. *Right* : 50% missing shots. *Top*: low frequency slice. *Bottom*: high frequency slice. Missing source subsampling *increases* the singular values in the (midpoint-offset) domain instead of *decreasing* them in the (src-rec) domain.

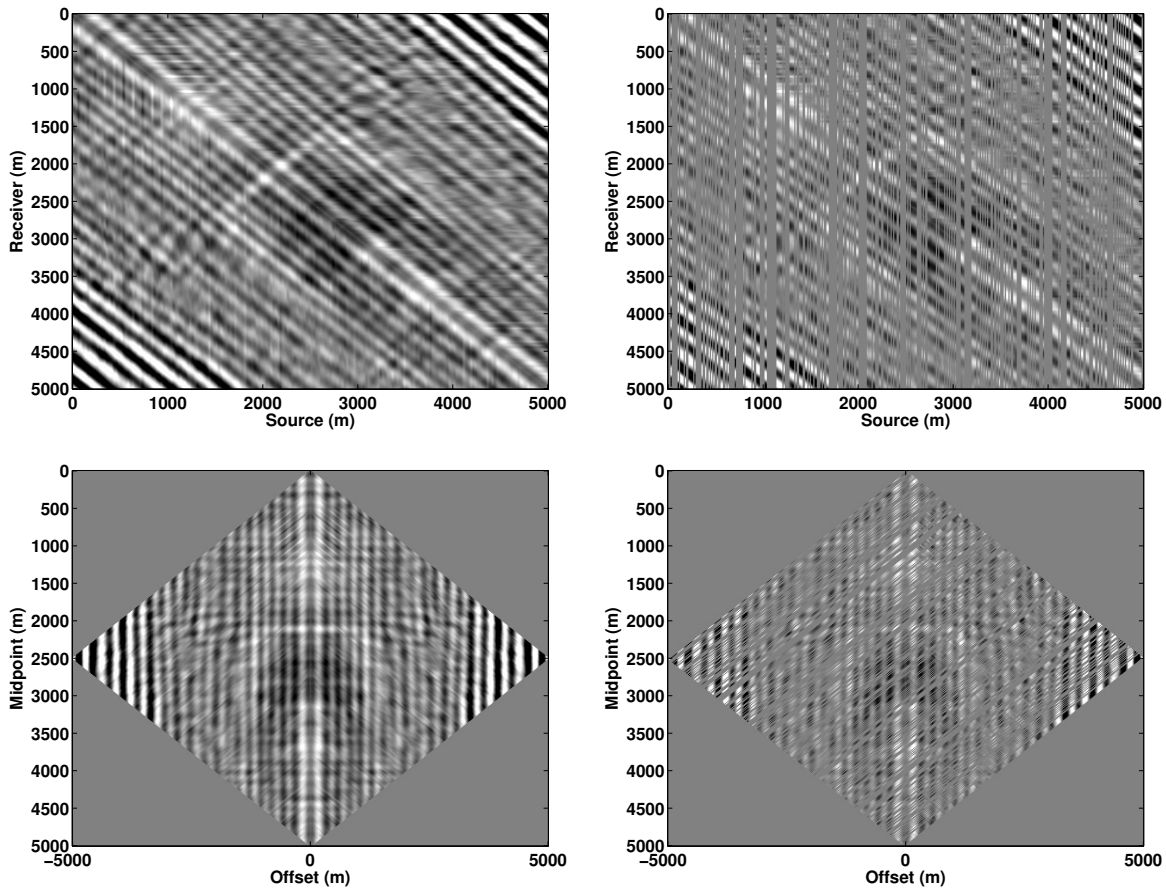


Figure 2: A frequency slice from the the seismic dataset from Nelson field. *Left*: Fully sampled data. *Right*: 50% subsampled data. *Top*: Source-receiver domain. *Bottom*: Midpoint-offset domain.

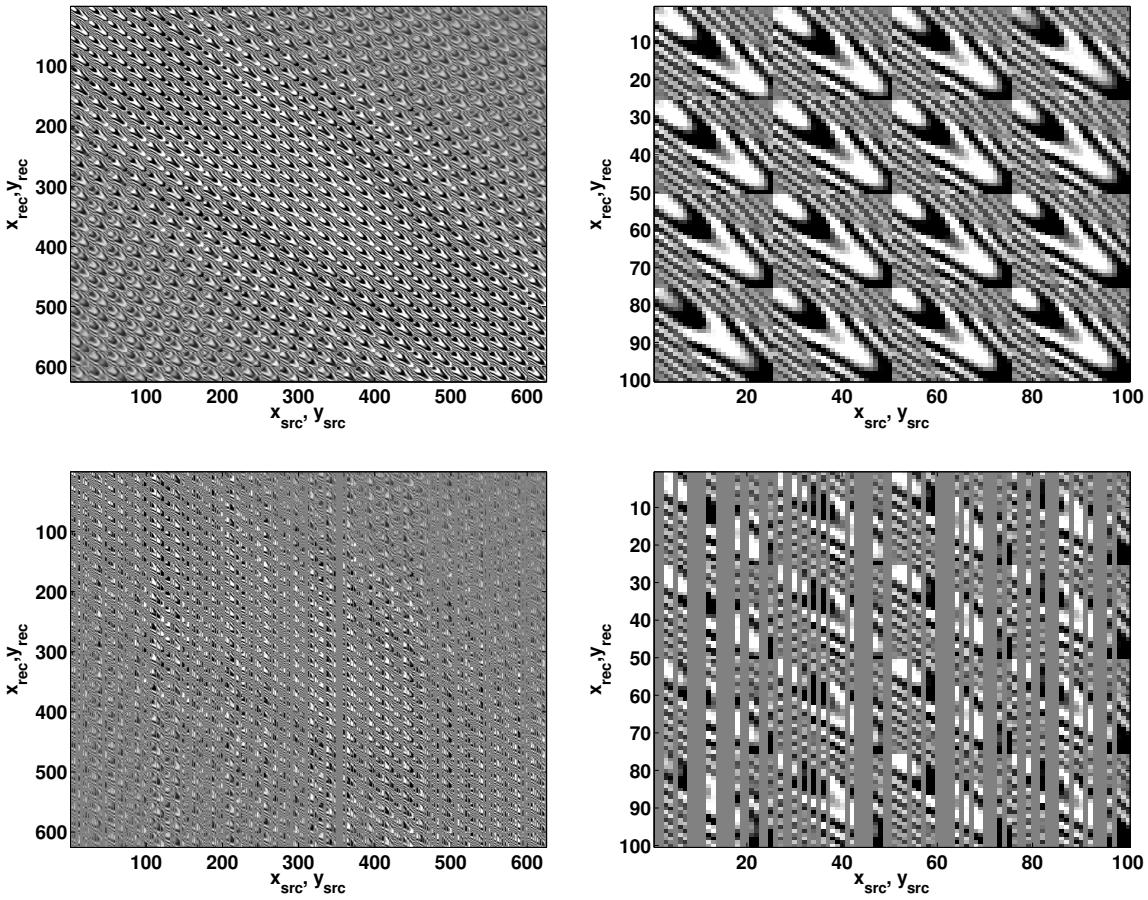


Figure 3: (x_{rec}, y_{rec}) matricization. *Top*: Full data volume. *Bottom*: 50% missing sources. *Left*: Fully sampled data. *Right*: Zoom plot

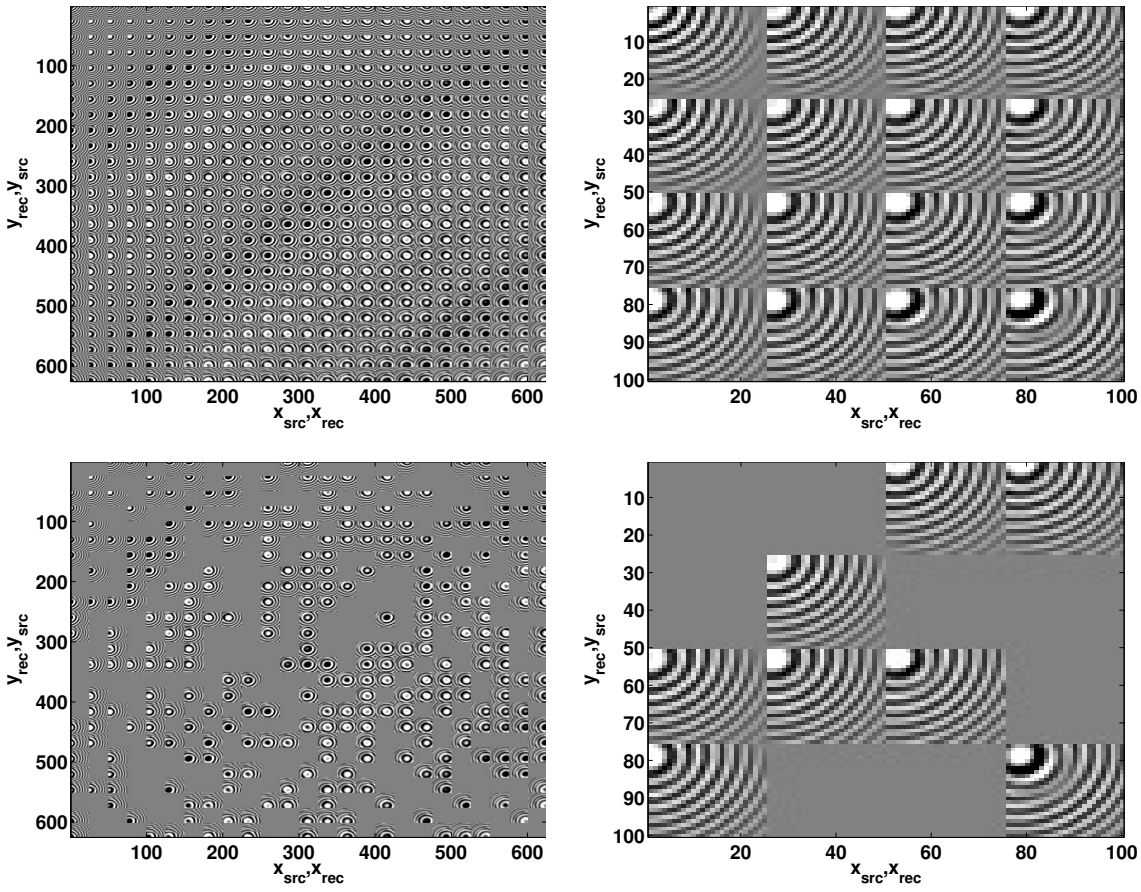


Figure 4: (y_{src}, y_{rec}) matricization. *Top*: Fully sampled data. *Bottom*: 50% missing sources. *Left*: Full data volume. *Right*: Zoom plot. In this domain, the sampling artifacts are much closer to the idealized 'pointwise' random sampling of matrix completion.

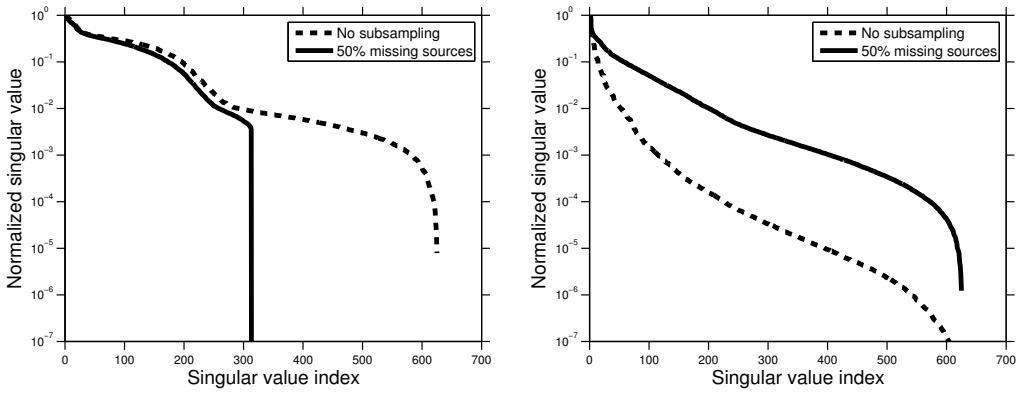


Figure 5: Singular value decay (normalized) of the *Left*: $(x_{\mathrm{rec}}, y_{\mathrm{rec}})$ matricization and *Right*: $(y_{\mathrm{src}}, y_{\mathrm{rec}})$ matricization for full data and 50% missing sources.

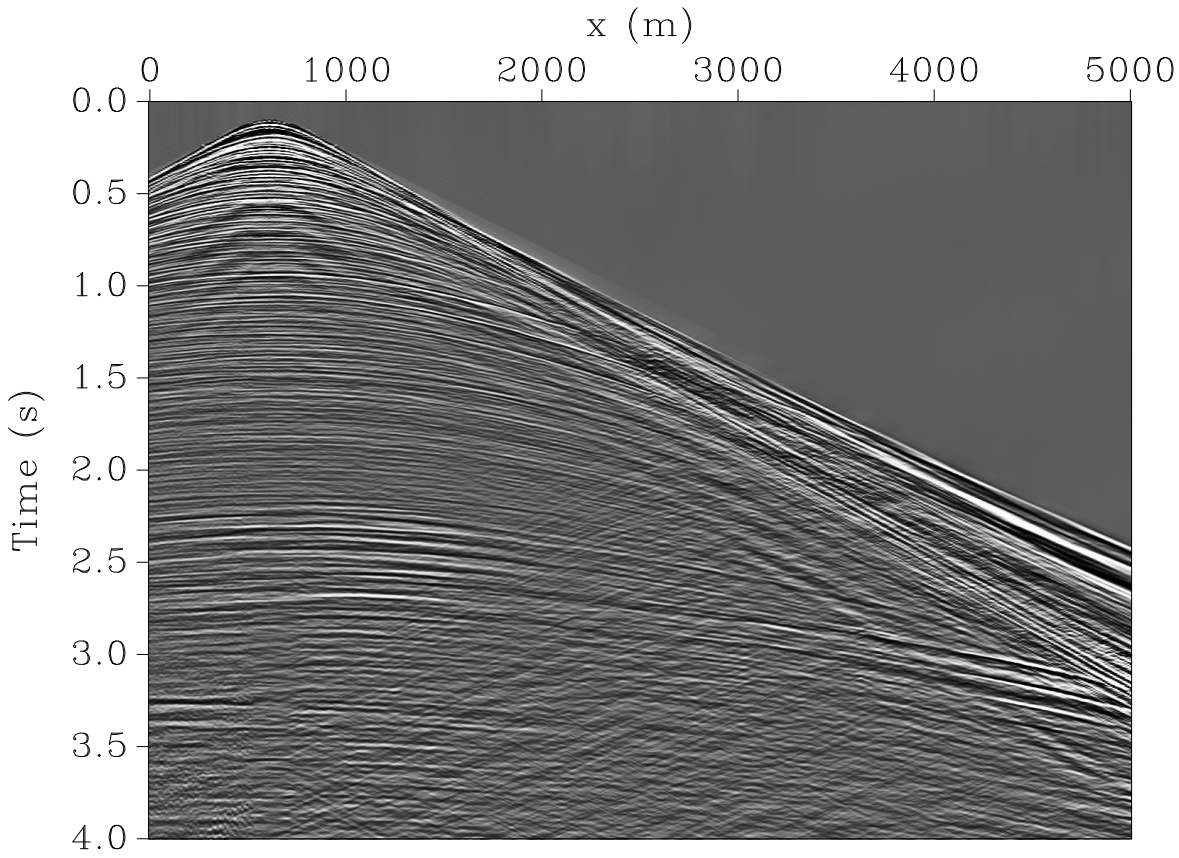


Figure 6: Original common receiver gather from the Nelson dataset.

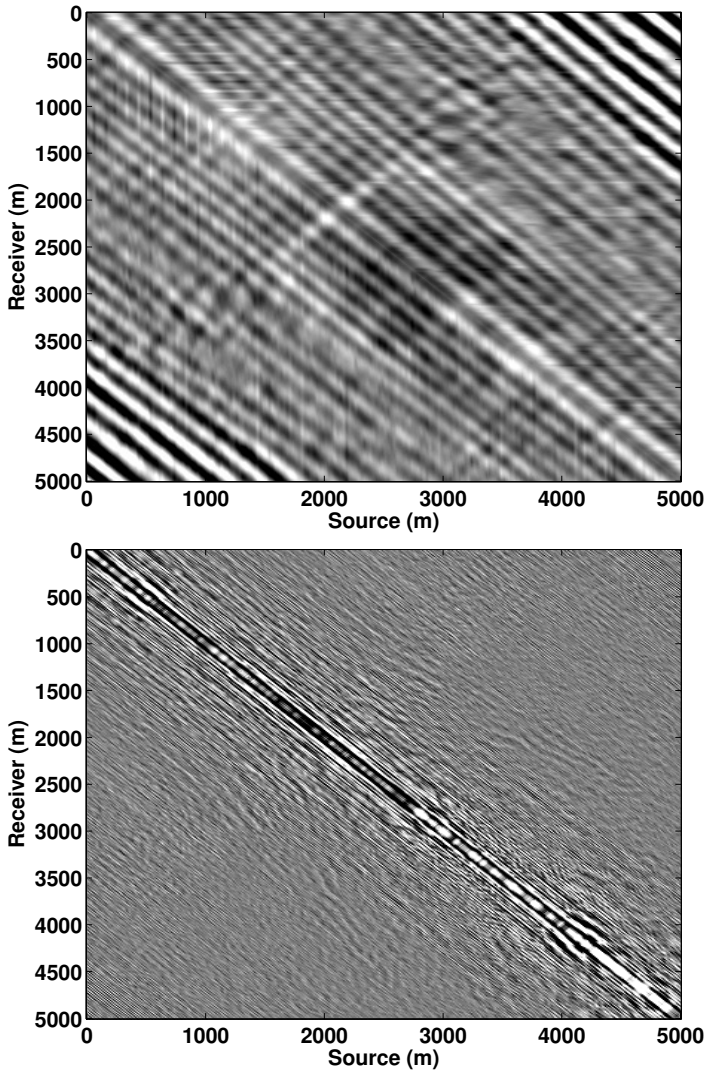


Figure 7: Ground truth. *Top*: Low-frequency slice at 10 Hz. *Bottom*: High frequency slice at 60Hz.

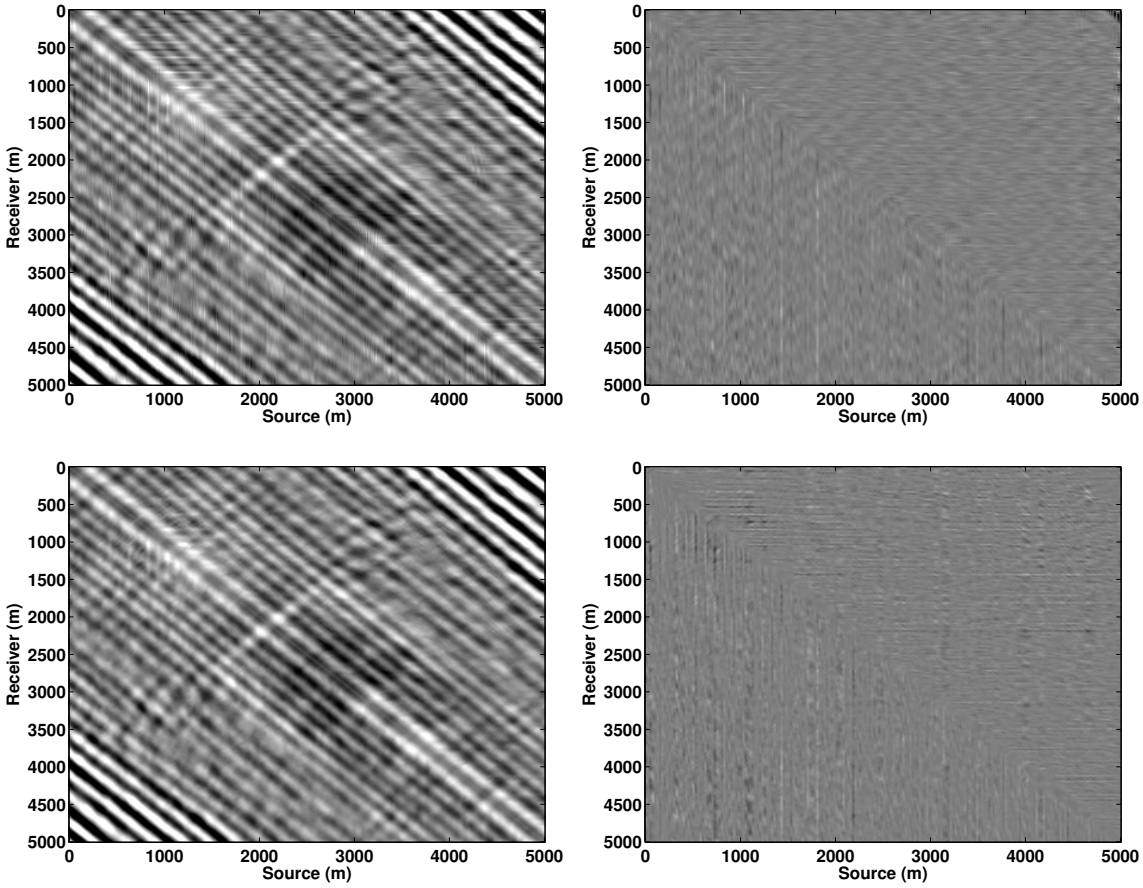


Figure 8: Interpolated low-frequency slice at 10Hz. *Left*: Interpolation is performed in the midpoint-offset domain. *Right*: Difference between true and interpolated slices. *Top*: curvelet-based interpolation (SNR = 18.2 dB). *Bottom*: matrix completion based interpolation (SNR = 18.6 dB).

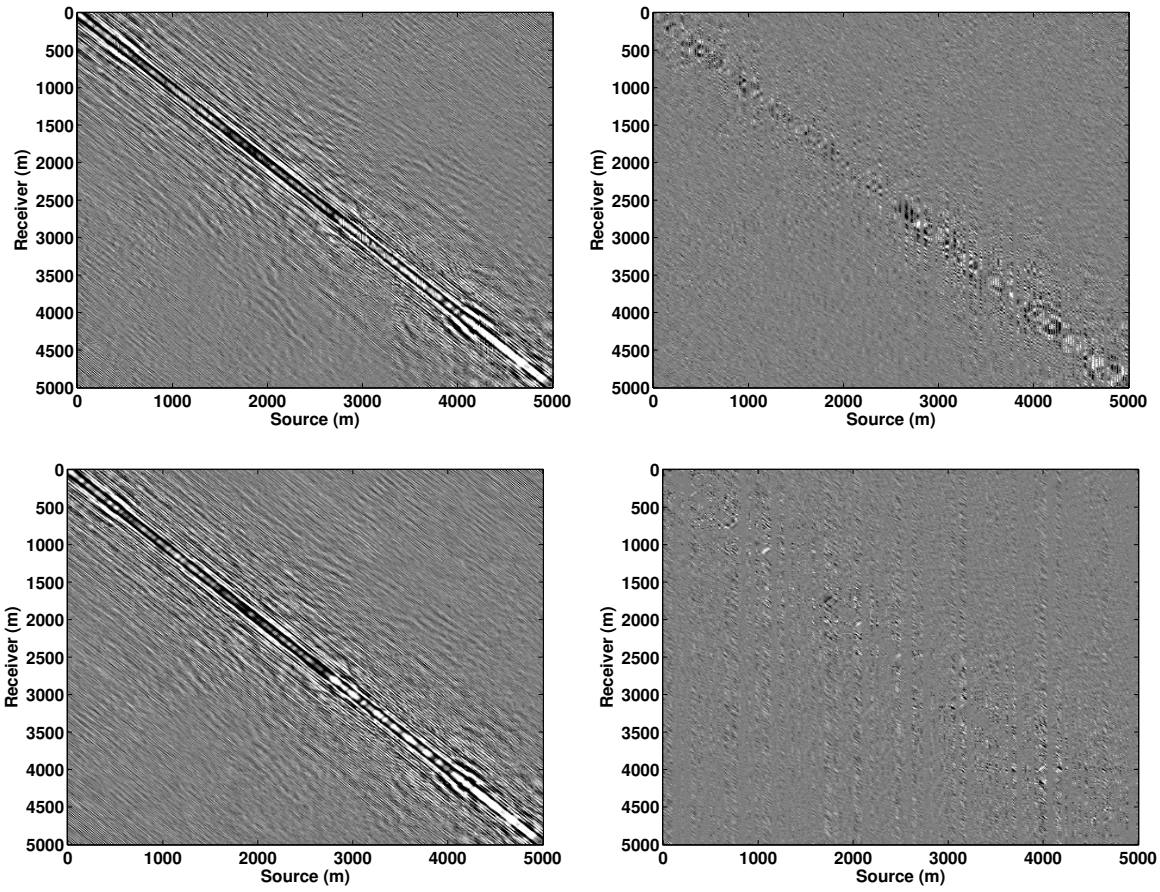


Figure 9: Interpolated high-frequency slice at 60Hz. *Left*: Interpolation is performed in the midpoint-offset domain. *Right*: Difference between true and interpolated slices. *Top*: curvelet-based interpolation (SNR = 10.5 dB). *Bottom*: matrix completion based interpolation (SNR = 12.5 dB). Owing to the curvelet transform, the computational costs of this sparsity-promoting paradigm are much larger compared to the matrix completion approach.

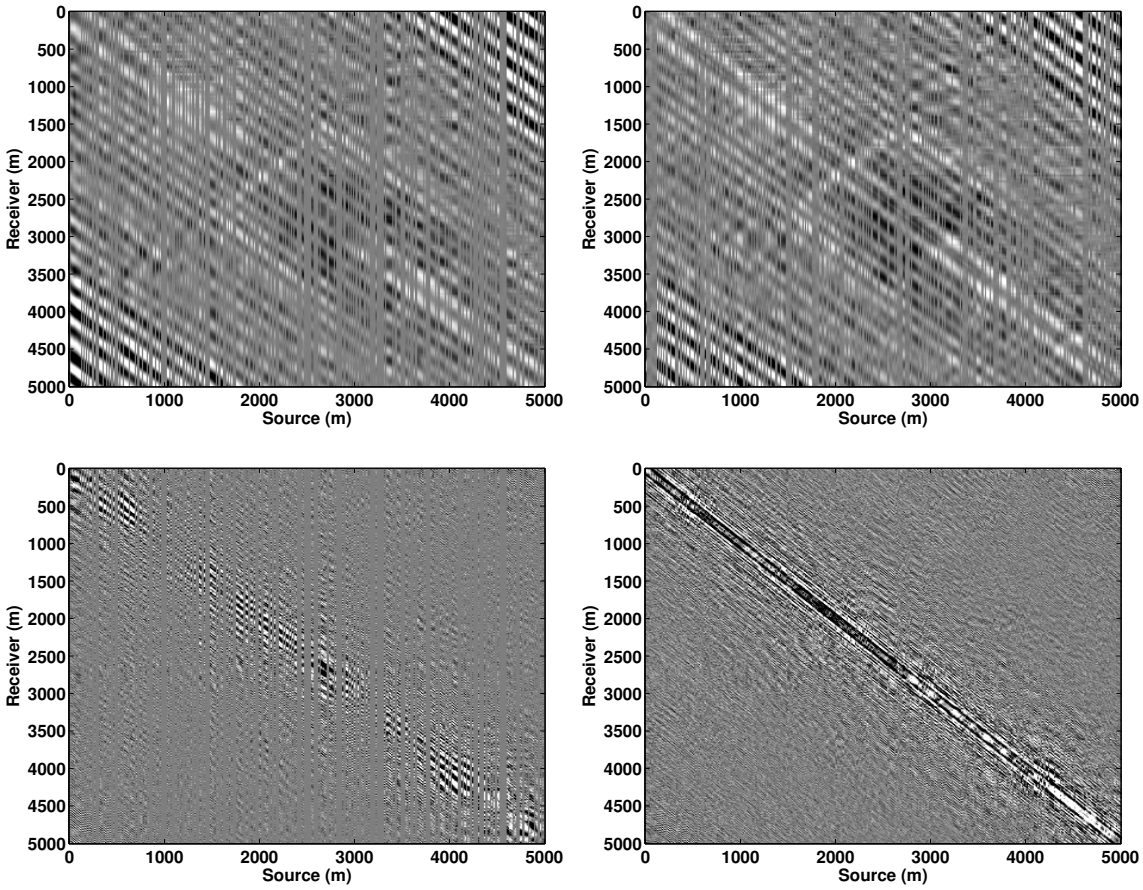


Figure 10: Recovery results using matrix-completion techniques. *Left*: Interpolation in the source-receiver domain, overall low-frequency SNR 3.1 dB and high-frequency SNR 1.74 dB. *Right*: Difference between true and interpolated slices. Since the sampling artifacts in the source-receiver domain do *not* increase the singular values, matrix completion in this domain is unsuccesful. This example highlights the necessity of having the appropriate principles of low-rank recovery in place before a seismic signal can be interpolated effectively.

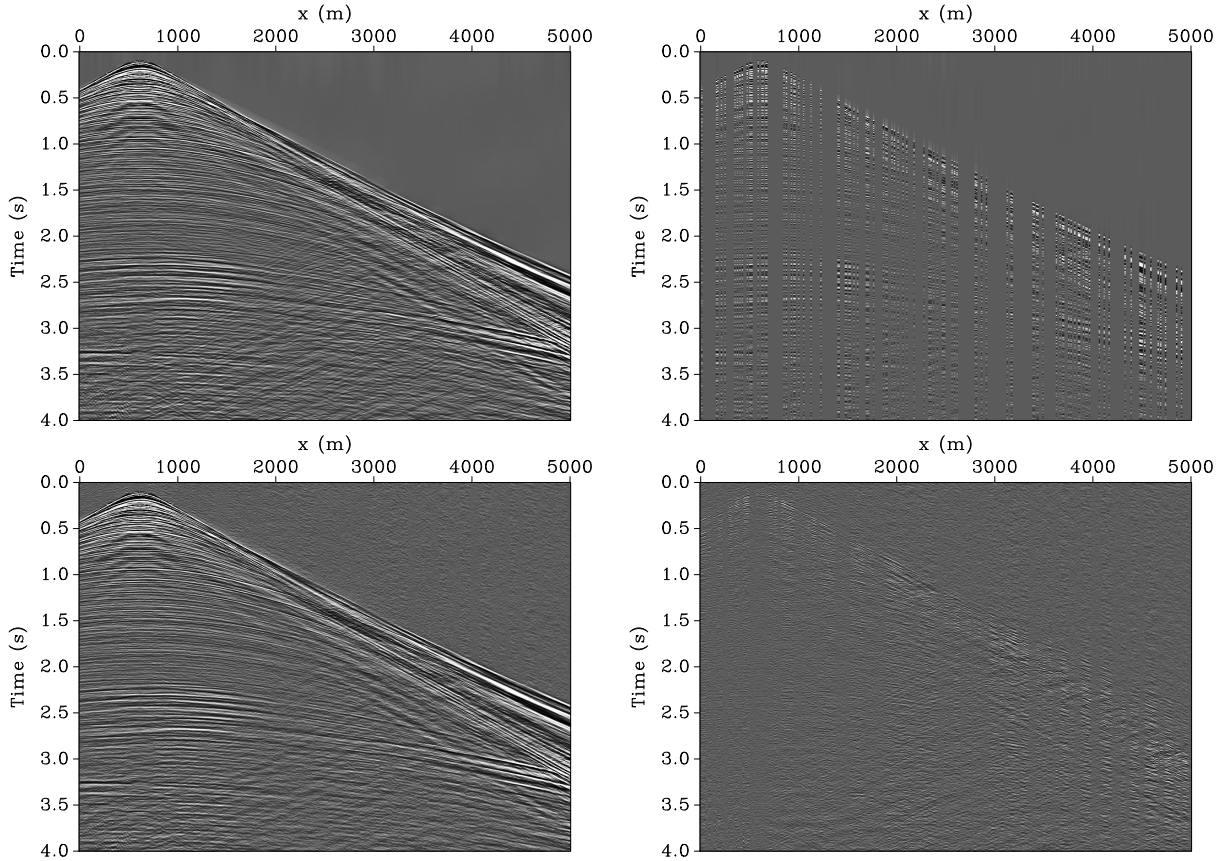


Figure 11: Missing-trace interpolation. *Top* : Fully sampled data and 75% subsampled common receiver gather. *Bottom* Recovery and residual results with a SNR of 9.4 dB.

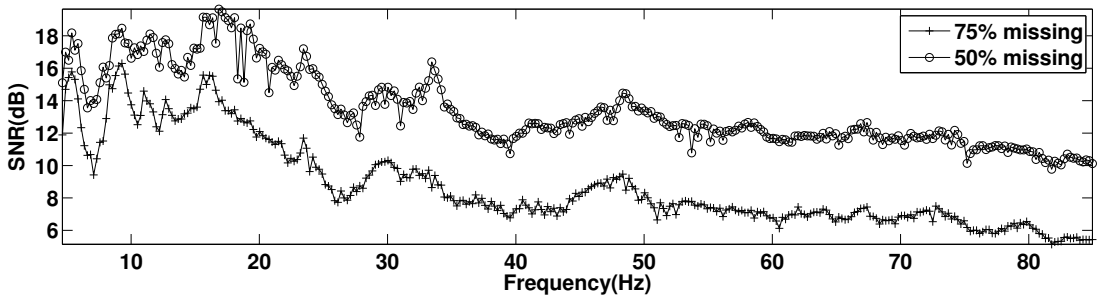


Figure 12: Qualitative performance of 2D seismic data interpolation for 5-85 Hz frequency band for 50% and 75% subsampled data.

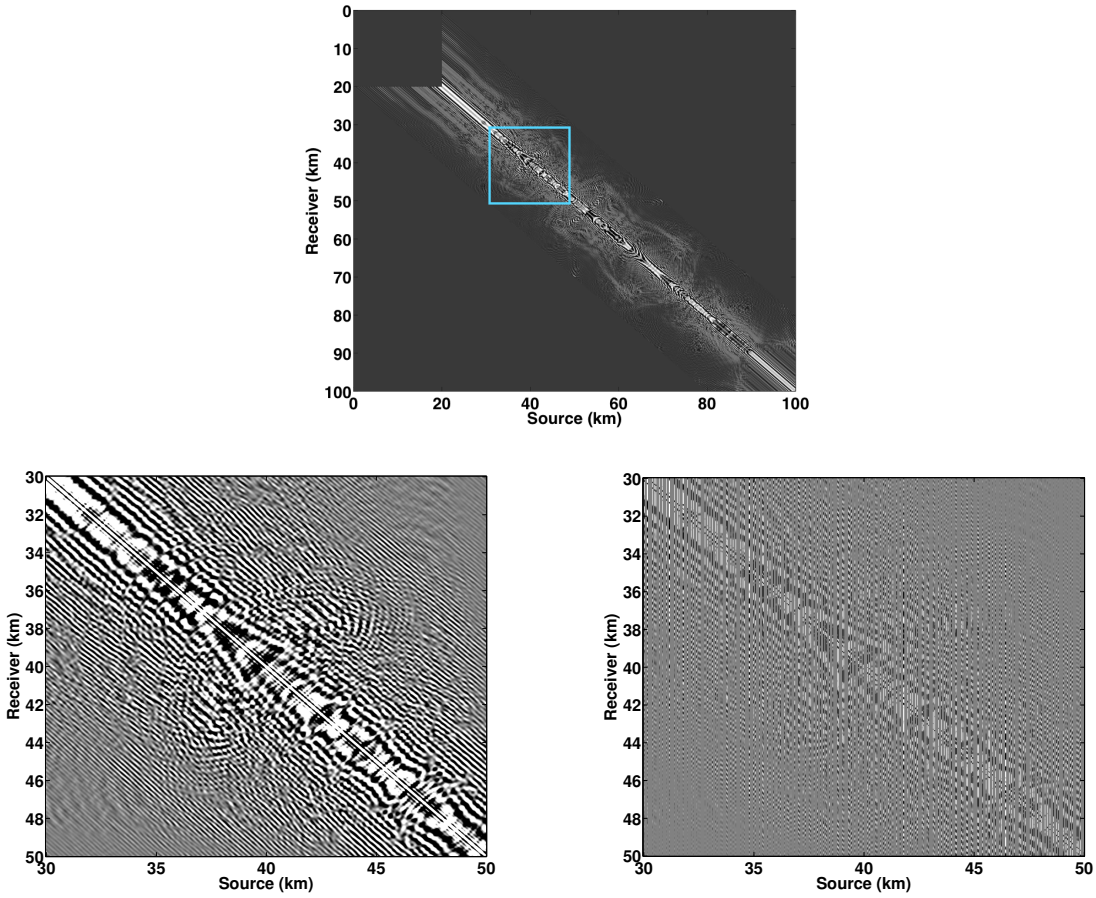


Figure 13: Gulf of Mexico data set. *Top*: Fully sampled monochromatic slice at 7 Hz. *Bottom left*: Fully sampled data (zoomed in the square block). *Bottom right*: 80% subsampled sources. For visualization purpose, the subsequent figures only show the interpolated result in the square block.

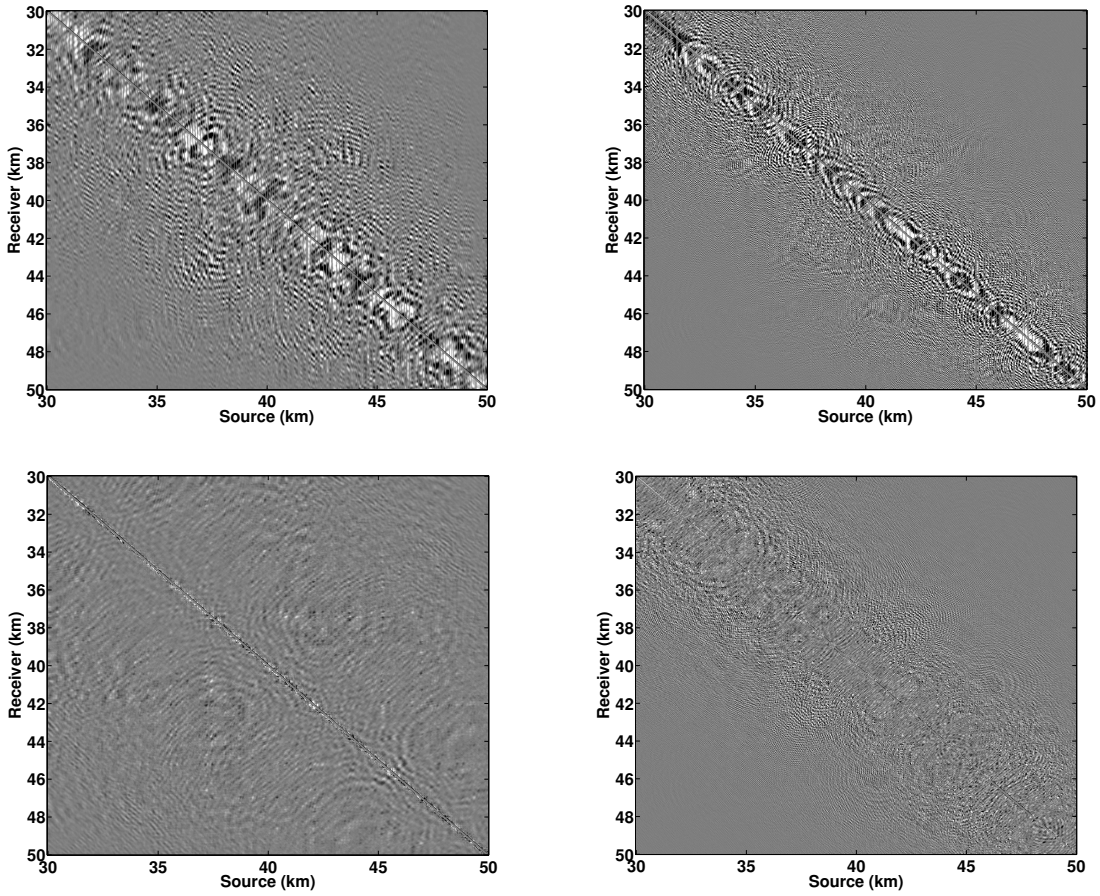


Figure 14: Reconstruction errors for frequency slice at 7Hz (left) and 20Hz (right) in case of 80% subsampled sources. *Top*: Curvelet based recovery with a SNR of 9.4 dB and 6.0 dB respectively. Bottom: Rank-minimization based recovery with a SNR of 14.2 dB and 11.0 dB respectively.

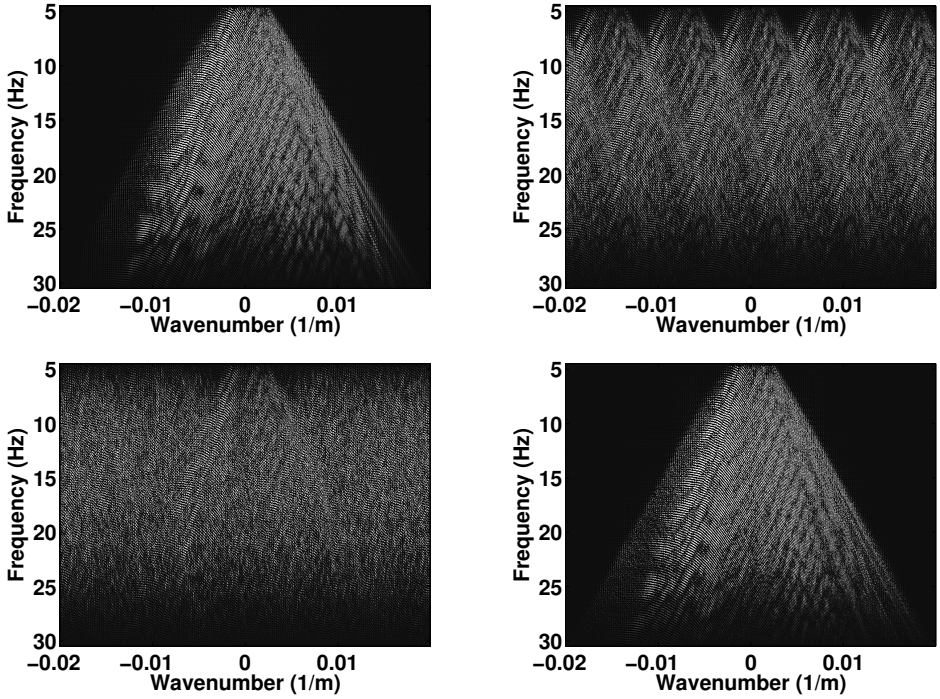


Figure 15: Frequency-wavenumber spectrum of the common receiver gather. *Top left*: Fully-sampled data. *Top right*: Periodic subsampled data with 80% missing sources. *Bottom left*: Uniform-random subsampled data with 80% missing sources. *Bottom Right*: Reconstruction of uniformly-random subsampled data using rank-minimization based techniques. While periodic subsampling creates aliasing, uniform-random subsampling turns the aliases in to incoherent noise across the spectrum.

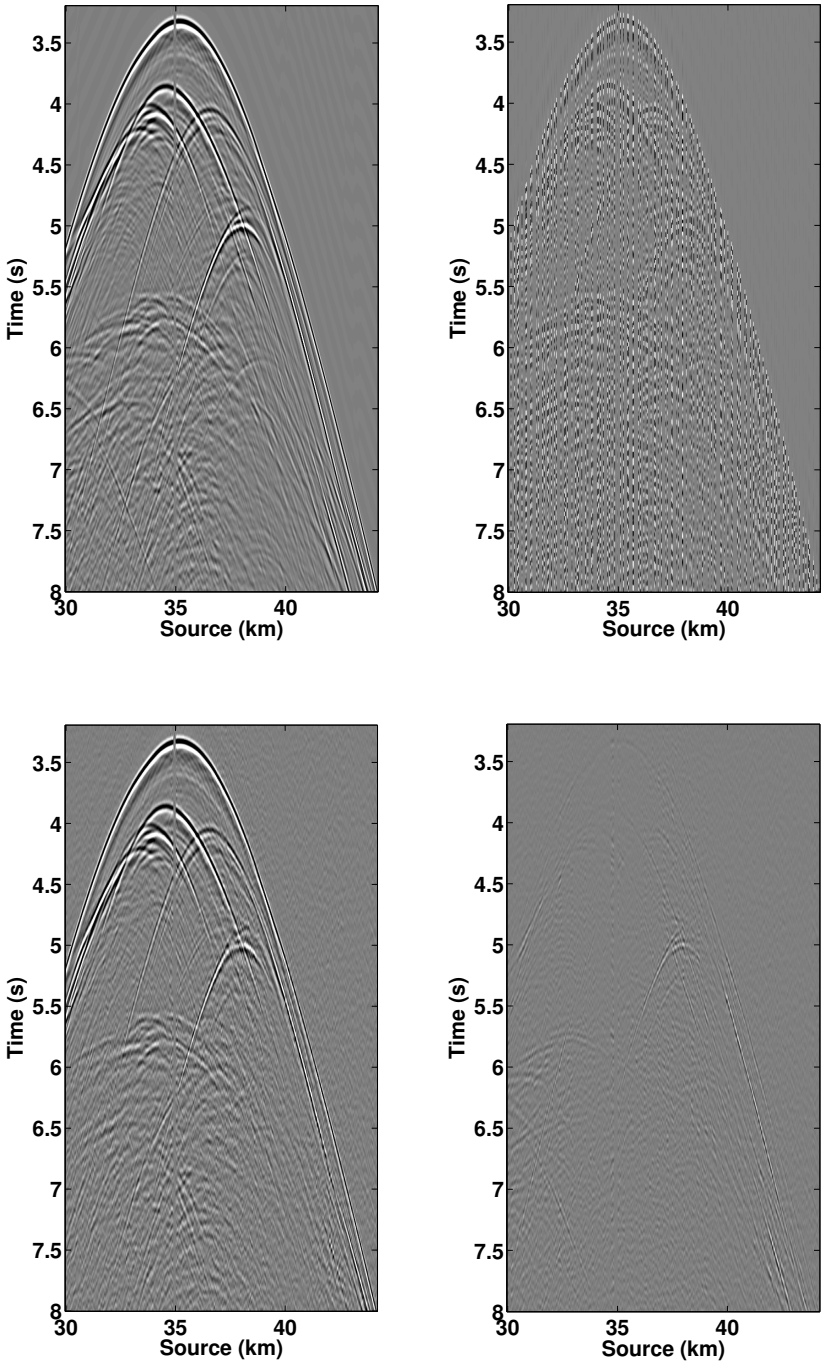


Figure 16: Gulf of Mexico data set, common receiver gather. *Top left* : Densly-sampled data. *Top right* : Uniformly-random subsampled data with 80% missing sources. *Bottom left* : Reconstruction results using rank-minimization based techniques (SNR = 7.8 dB). *Bottom Right* : Residual.

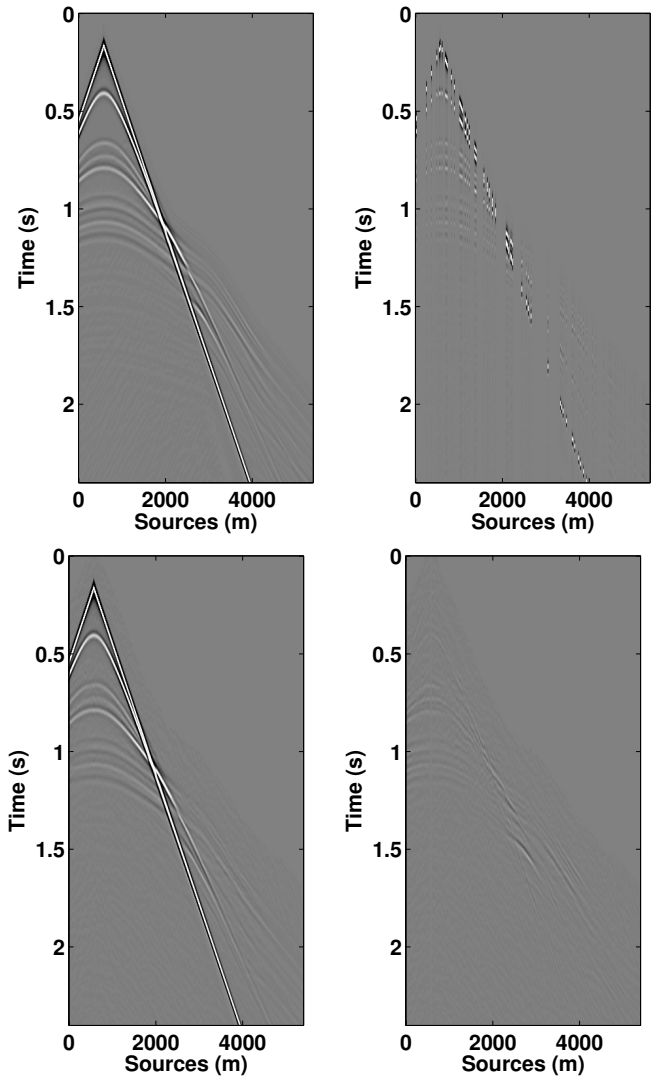
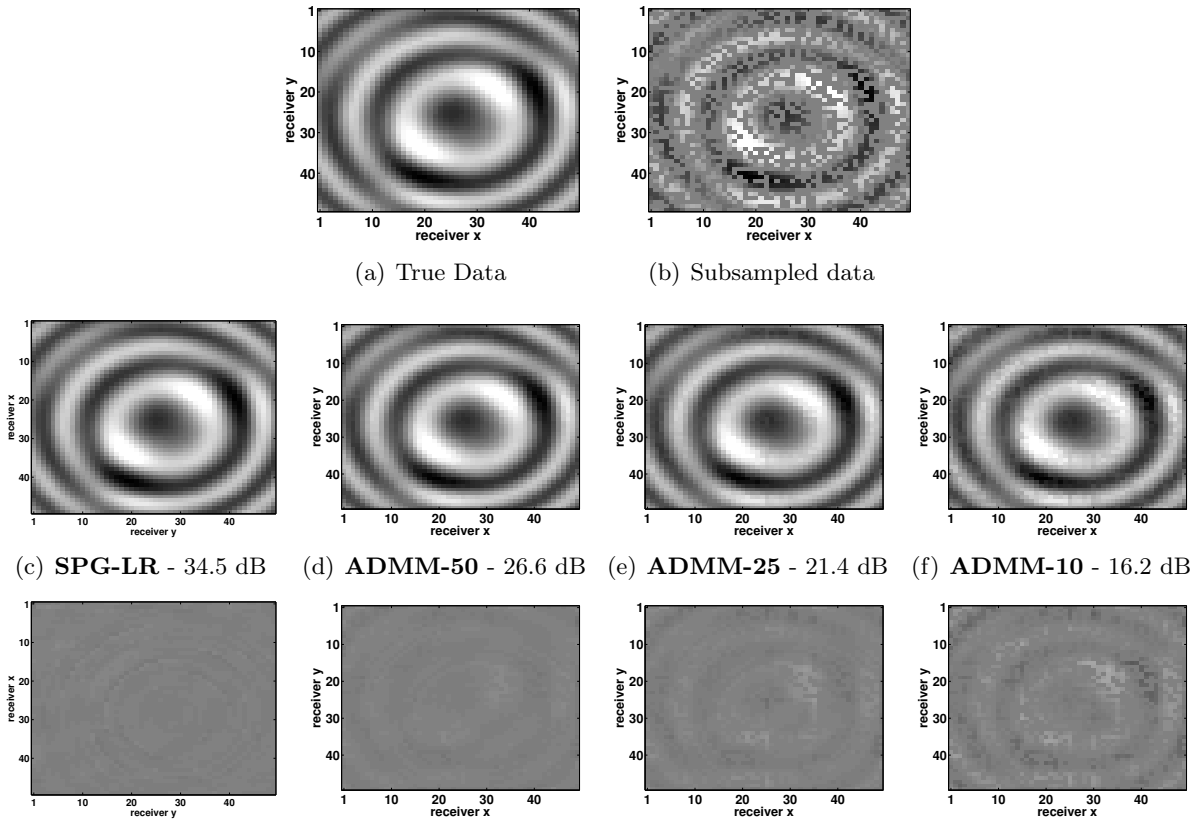


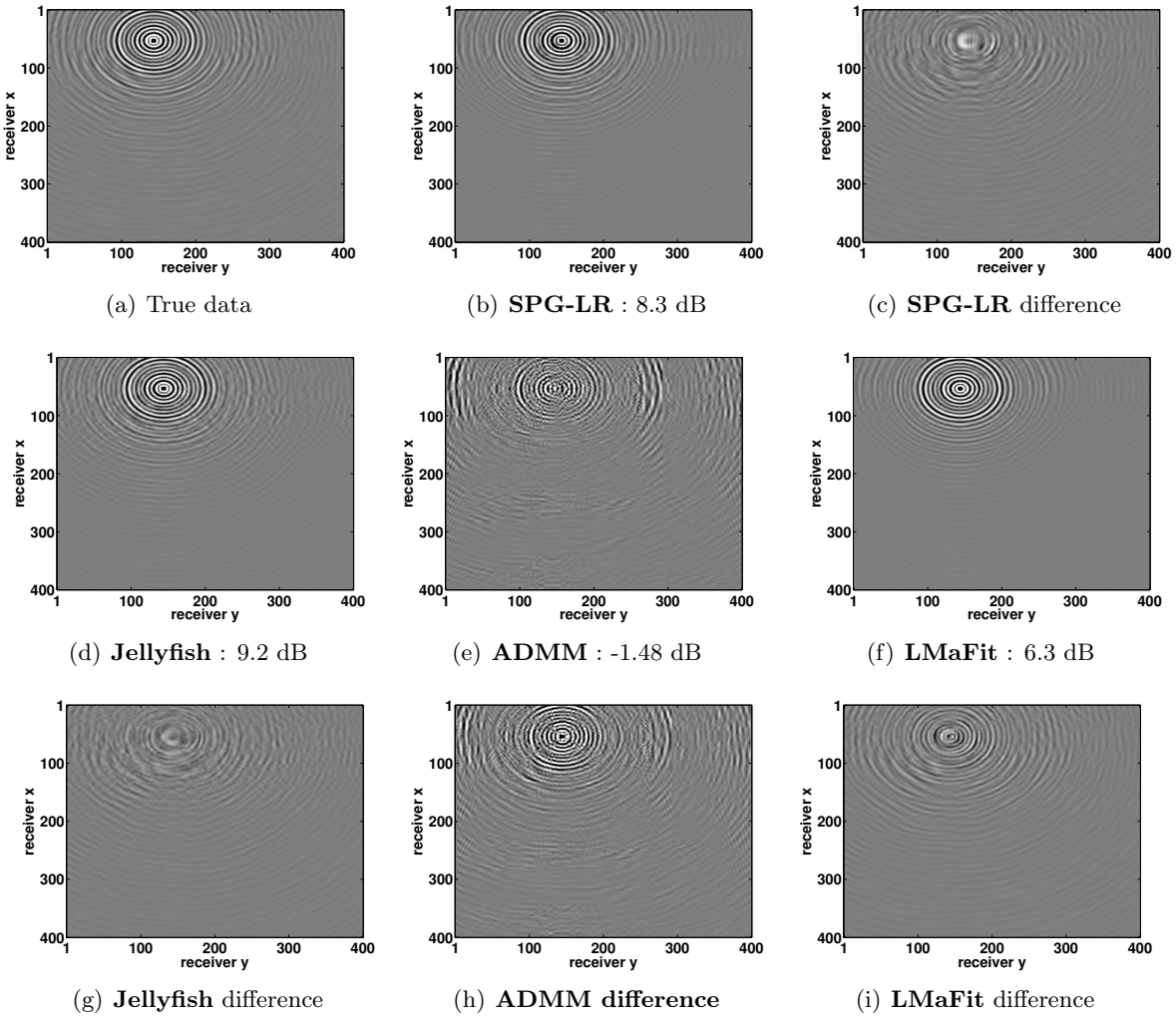
Figure 17: Missing-trace interpolation (80% sub-sampling) in case of geological structures with a fault. *Top left*: Fully sampled data. *Top right*: 80% sub-sampled data. *Bottom left*: after interpolation (SNR = 23 dB). *Bottom right*: difference.



(a) True Data (b) Subsampled data

(c) **SPG-LR** - 34.5 dB (d) **ADMM-50** - 26.6 dB (e) **ADMM-25** - 21.4 dB (f) **ADMM-10** - 16.2 dB

Figure 18: ADMM data fit + recovery quality (SNR) for single reflector data, common receiver gather. Middle row: recovered slices, bottom row: residuals corresponding to each method in the middle row. Tensor-based windowing appears to visibly degrade the results, even with overlap.

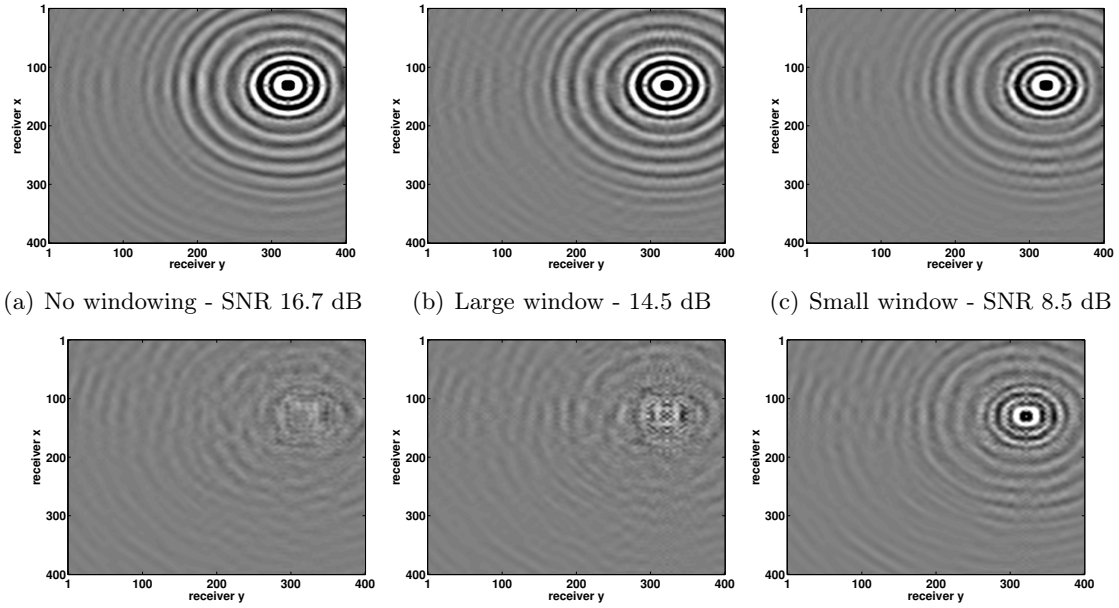


(a) True data (b) **SPG-LR** : 8.3 dB (c) **SPG-LR** difference

(d) **Jellyfish** : 9.2 dB (e) **ADMM** : -1.48 dB (f) **LMaFit** : 6.3 dB

(g) **Jellyfish** difference (h) **ADMM difference** (i) **LMaFit** difference

Figure 19: BG 5-D seismic data, 12.3 Hz, 75% missing sources. Middle row: interpolation results, bottom row: residuals.



(a) No windowing - SNR 16.7 dB (b) Large window - 14.5 dB (c) Small window - SNR 8.5 dB

Figure 20: BG 5D seismic data, 4.68 Hz, Comparison of interpolation results with and without windowing using Jellyfish for 75% missing sources. Top row: interpolation results for differing window sizes, bottom row: residuals.

LIST OF TABLES

Table 1: Curvelet versus matrix completion (MC). **Real data** results for completing a frequency slice of size 401×401 with 50% and 75% missing sources. *Left*: 10 Hz (low frequency), *right*: 60 Hz (high frequency). SNR, computational time, and number of iterations are shown for varying levels of $\eta = 0.08, 0.1$.

		Curvelets		**MC**	
η		0.08	0.1	0.08	0.1
50%	SNR (dB)	18.2	17.3	**18.6**	17.7
	time (s)	1249	1020	15	**10**
	iterations	123	103	191	124
75%	SNR (dB)	**13.5**	13.2	13.0	13.3
	time (s)	1637	1410	8.5	**8**
	iterations	162	119	105	104

		Curvelets		**MC**	
η		0.08	0.1	0.08	0.1
50%	SNR (dB)	10.5	10.4	**12.5**	12.4
	time (s)	1930	1549	19	**13**
	iteration	186	152	169	118
75%	SNR (dB)	6.0	5.9	6.9	**7.0**
	time (s)	3149	1952	15	**10**
	iteration	284	187	152	105

Method	SNR	Solve time	Parameter selection time	Total time
SPG-LR	25.5	0.9	N/A	0.9
ADMM - 50	20.8	87.4	320	407.4
ADMM - 25	16.8	4.4	16.4	20.8
ADMM - 10	10.9	0.1	0.33	0.43

Table 2: Single reflector data results. The recovery quality (in dB) and the computational time (in minutes) is reported for each method. The quality suffers significantly as the window size decreases due to the smaller redundancy of the input data, as discussed previously.

Frequency	Missing sources	SPG-LR		Jellyfish		ADMM		LmaFit	
		SNR	Time	SNR	Time	SNR	Time	SNR	Time
4.68 Hz	75%	15.9	84	**16.34**	**36**	0.86	1510	14.7	204
	50%	**20.75**	96	19.81	**82**	3.95	1510	17.5	91
	25%	**21.47**	114	19.64	124	9.17	1510	18.9	**66**
7.34 Hz	75%	11.2	84	**11.99**	**52**	0.39	1512	10.7	183
	50%	**15.2**	126	15.05	146	1.71	1512	14.1	**37**
	25%	**16.3**	138	15.31	195	4.66	1512	14.3	**21**
12.3 Hz	75%	7.3	324	**9.34**	**223**	0.06	2840	8.1	814
	50%	**12.6**	438	12.12	706	0.21	2840	11.1	**72**
	25%	**14.02**	450	12.90	1295	0.42	2840	11.3	**58**

Table 3: 3D seismic data results. The recovery quality (in dB) and the computational time (in minutes) is reported for each method.