Fast Methods for Rank Minimization with Applications to Seismic-Data Interpolation

Rajiv Kumar, Aleksandr Y. Aravkin, and Felix J. Herrmann. Seismic Laboratory of Imaging and Modeling - The University of British Columbia*

SUMMARY

Rank penalizing techniques are an important direction in seismic inverse problems, since they allow improved recovery by exploiting low-rank structure. A major downside of current state of the art techniques is their reliance on the SVD of seismic data structures, which can be prohibitively expensive. Fortunately, recent work allows us to circumvent this problem by working with matrix factorizations. We review a novel approach to rank penalization, and successfully apply it to the seismic interpolation problem by exploiting the low-rank structure of seismic data in the midpoint-offset domain. Experiments for the recovery of 2D monochromatic data matrices and seismic lines represented as 3D volumes support the feasibility and potential of the new approach.

INTRODUCTION

Missing data due to physical and/or budget constraints is a common issue in seismic data acquisition. One of the initial goals of processing is to spatially transform irregularly acquired data to regularly sampled data, which is required by subsequent processing steps, including removal of artifacts, improvement of spatial resolution, and key analysis, such as migration. As a result, recovery of missing traces has become a problem of paramount interest in seismic research. Methods to solve this problem rely on a wide variety of mathematical techniques. Some of the methods require transforming the data into different domains, such as the Fourier transform (Sacchi et al., 1998) and Curvelet transform (Herrmann and Hennenfent, 2008). Structure of the data in these domains can then be exploited to recover the missing traces. One example of special structure is sparsity (e.g. in the Fourier basis or Curvelet frame), and exploiting this structure in linear inverse problems has been extensively studied in the field of Compressive Sensing (CS) (Donoho, 2006). In seismic applications, ℓ_1-norm minimization has been successfully applied to reconstruction of missing traces.

Low rank of the seismic data matrix obtained by a particular 2D arrangement, e.g. source-receiver, is another form of structure that can be exploited for seismic interpolation. Rank penalization has recently been applied to seismic data regularization and denoising (Oropeza and Sacchi, 2011) using multichannel singular spectrum analysis (MSSA). Specifically, this was done by generalizing the concept of singular spectrum, and then using rank reduction via the SVD. This approach conceptually resembles a recent optimization approach to rank minimization, known as nuclear norm minimization (Candes and Plan, 2010), which also relies on the SVD.

While low rank is a key structure that should be exploited, there are two main challenges that must be addressed. First, in order to gain the full benefit of low-rank interpolation, we must find transformations or representations of seismic data where the corresponding seismic-data matrix has as low rank as possible, i.e. the singular values decay as quickly as possible. We take an important step in this direction by working with data in midpoint-offset coordinates, where we show empirically that the eigenvalue decay is much faster than for the source-receiver indexing. Second, the main computational challenge of rank minimization problem is the reliance on the SVD decomposition, which is prohibitively expensive for large matrices. Recent work (Lee et al., 2010) has demonstrated that rank penalization can be done very quickly, and we show that the new methodology is feasible for seismic data interpolation. The paper proceeds as follows. We review nuclear norm minimization, its connections to CS, and then describe the new max-norm penalty approach (Lee et al., 2010). We also briefly discuss the connections between low rank penalization and CS. We then present experimental results that recover missing traces using the max-norm penalty approach to rank penalization in the midpoint-offset domain for 2D monochromatic data matrices and seismic lines represented as 3D volumes from the Gulf of Suez (see Figure 1), and compare to results obtained by sparsity regularization using SPGℓ_1. A discussion section ends the paper.

RANK PENALIZATION VS. SPARSITY

Seismic data has a lot of structure that can be exploited to recover missing traces. Compressibility (rapid decay of coefficients) in the curvelet frame is one example of this structure; see e.g.(Candes and Donoho, 2000; Herrmann, 2008). The key idea is that missing traces *increase* the number of nonzero coefficients in the curvelet representation, so solving an optimization problem that promotes curvelet sparsity recovers these traces. The optimization problem is given by

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}\,, \tag{1}$$

where $\mathbf{x}$ are the curvelet coefficients to recover, $\mathbf{b}$ are the available data, and $\mathbf{A} = \mathbf{RMS}^H$, with $\mathbf{S}^H$ the sparsifying synthesis matrix, $\mathbf{R}$ is restriction operator, and $\mathbf{M}$ the measurement matrix. The reconstruction is given by $\hat{\mathbf{f}} = \mathbf{S^H}\tilde{\mathbf{x}}$.

In addition to the structure in the representation, seismic data can be organized as a matrix, and in this case low-rank structure has been exploited for seismic applications (Oropeza and Sacchi, 2011). The key idea is very similar: additive noise and missing samples tend to increase the rank of the data *matrix*, and so penalizing rank recovers the missing traces and attenuates the noise.

Matrix completion using rank minimization has been formulated and studied by the same community that has worked on compressive sensing (see Candes and Plan (2010) and the references within). The optimization formulation is analogous to Eq. 1:

$$\tilde{\boldsymbol{Y}} = \arg\min_{\boldsymbol{Y}} \|\boldsymbol{Y}\|_* \text{ subject to } \mathscr{A}_\Omega(\boldsymbol{Y}) = \mathscr{A}_\Omega(\boldsymbol{X}) \tag{2}$$

where $\mathscr{A}_\Omega : \mathbb{R}^{mxn} \to \mathbb{R}^{mxn}$ is defined by

$$[\mathscr{A}(\boldsymbol{X})]_{ij} = \begin{cases} \boldsymbol{X}_{ij}, \text{(i,j)} \in \Omega \\ 0, \text{ otherwise,} \end{cases}$$

Ω is the subset of the complete set of entries $m \times n$, $\boldsymbol{Y} \in \mathbb{R}^{mxn}$ is the decision variable, and $\boldsymbol{X}$ are the available data. Just as the algorithm proposed by (Oropeza and Sacchi, 2011), optimization methods for problem 2 rely on the SVD, which is very expensive for seismic data. Recent work by (Lee et al., 2010) avoids the SVD completely, and provides simple first-order algorithms for solving surrogate rank-penalization problems. We review the approach below.

Matrix factorization and max-norm

As pointed out by (Recht and Ré, 2011), the nuclear norm can be represented as a penalty over all possible factorizations of $\boldsymbol{Y}$:

$$\|\boldsymbol{Y}\|_* = \inf\left\{\frac{1}{2}\|\boldsymbol{L}\|_F^2 + \frac{1}{2}\|\boldsymbol{R}\|_F^2 : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{R}^H\right\}, \quad (3)$$

where the infimum (minimum) is taken over all factorizations of $\boldsymbol{Y}$, and $\|\cdot\|_F$ is the Frobenius norm (sum of squares of all entries). It is well known that the nuclear norm promotes low-rank decomposition with factors in ℓ_2. In (Lee et al., 2010), the authors propose using the max-norm, given by

$$\|\boldsymbol{Y}\|_* = \inf\left\{\max\left\{\|\boldsymbol{L}\|_{2,\infty}^2, \|\boldsymbol{R}_{2,\infty}^2\right\} : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{R}^H\right\}, \quad (4)$$

where $\boldsymbol{L}_{2,\infty}^2$ is the maximum of the squares of the column norms of $\boldsymbol{L}$. This norm is closely related to the nuclear norm, with the difference that it promotes low-rank decomposition in ℓ_∞ factors rather than ℓ_2 factors (Lee et al., 2010). Working with the maximum formulation rather than the Frobenius formulation yields very simple and efficient first-order algorithms, as we discuss below, while still promoting low-rank structure.

The advantage of the representations 3 and 4 is the ability to work directly with the factors $\boldsymbol{L},\boldsymbol{R}$ rather than with the full matrix. This requires deciding ahead of time how many columns to use in the decision variables—one could always use the full system size, but it is much better to use a smaller number of columns if possible because of memory costs. Rather than solving 2, we can then solve either one of the following formulations:

$$\min_{\boldsymbol{L},\boldsymbol{R}} \frac{1}{2}\|\mathscr{A}_\Omega(\boldsymbol{L}\boldsymbol{R}^H - \boldsymbol{X})\|_F^2 + \lambda r(\boldsymbol{L},\boldsymbol{R}) \quad (5)$$

or

$$\min_{\boldsymbol{L},\boldsymbol{R}} \frac{1}{2}\|\mathscr{A}_\Omega(\boldsymbol{L}\boldsymbol{R}^H - \boldsymbol{X})\|_F^2 \text{ subject to } r(\boldsymbol{L},\boldsymbol{R}) \leq B, \quad (6)$$

where $r(\boldsymbol{L},\boldsymbol{R})$ is given by $\frac{1}{2}\|\boldsymbol{L}\|_F^2 + \frac{1}{2}\|\boldsymbol{R}\|_F^2$ or $\max\left\{\|\boldsymbol{L}\|_{2,\infty}^2, \|\boldsymbol{R}_{2,\infty}^2\right\}$, depending on the type of penalty we wish to use. (Lee et al., 2010) show that using the constrained formulation 6 in conjunction with the max-norm penalty gives a very simple projected-gradient method, and we use this combination for all of the experiments.

The projected-gradient method for the factor formulation holds a huge advantage over SVD-based methods, and is worth explaining. The iteration is given by

$$\begin{bmatrix}\boldsymbol{L}\\ \boldsymbol{R}\end{bmatrix}^{k+1} = \mathscr{P}_B\left(\begin{bmatrix}\boldsymbol{L}^k - \tau\mathscr{A}_\Omega(\boldsymbol{L}^k(\boldsymbol{R}^k)^H - \boldsymbol{X})\boldsymbol{R}^k\\ \boldsymbol{R}^k - \tau\mathscr{A}_\Omega(\boldsymbol{L}^k(\boldsymbol{R}^k)^H - \boldsymbol{X})^H\boldsymbol{L}^k\end{bmatrix}\right), \quad (7)$$

where the projection $\mathscr{P}_B$ simply re-scales the rows of its arguments whose norms exceed $\sqrt{B}$ in such a way that they are then equal to $\sqrt{B}$.

To apply the method, a practitioner must choose both the rank of the decision variables $\boldsymbol{L},\boldsymbol{R}$ and the regularization parameter B. In our experiments, we show the results across a range of decision-variable ranks, and for several values of the regularization parameter, to give a sense of how these values affect the results.

RANK IN THE MIDPONT-OFFSET REPRESENTATION

In the CS framework, to achieve an effective inversion scheme one must first find a representation that satisfies two criteria:

- The signal of interest is sparse/compressible in the representation and
- Missing traces make the signal much less sparse/compressible in the representation.

In this setting, sparsity promoting schemes can improve recovery. Analogously, for the low rank case, we must find ways to organize the data matrix so that

- The true data is low rank or has quickly decaying eigenvalues and
- Missing traces increase the rank or make the eigenvalues decay much less quickly.

Note first that neither missing traces nor random source superpositions in the source-receiver (s-r) domain can increase the rank of the matrix (see Figure 2). This follows from the fact that for any matrices $\mathbf{A},\mathbf{B}$, the range of $\mathbf{AB}$ is contained in the range of $\mathbf{A}$, and both acquisition types can be represented this way. Fortunately, when we work with data in the midpoint-offset (m-h) domain, missing traces do indeed increase the effective rank (see Figure 2). The conversion from (s-r) to (m-h) is a coordinate transformation, with the midpoint $m = 0.5(s+r)$ and half-offset $h = 0.5(s-r)$. Moreover, for full data, the decay of the eigenvalues is very fast, which satisfies the first condition of recoverability.

Our approach is to reconstruct the data matrix $\mathbf{X}$ by inverting the transform-sampling operator defined by

$$\mathscr{A} = \mathbf{R}\mathbf{M}\mathscr{S}^H,$$

where $\mathbf{R}$ is a restriction matrix acting on the measurement matrix $\mathbf{M}$, and $\mathscr{S}^H$ is the matrix-valued operator which transforms the data into the (m-h) domain. We then solve problem 6, with $r(\boldsymbol{L},\boldsymbol{R}) = \max\left\{\|\boldsymbol{L}\|_{2,\infty}^2, \|\boldsymbol{R}_{2,\infty}^2\right\}$. To obtain the final result in the (s-r) domain, we simply calculate $\hat{\mathbf{f}} = \mathscr{S}^H(\boldsymbol{L}\boldsymbol{R}^H)$.

EXPERIMENTAL SETUP

We evaluated the efficacy of the max-norm approach by implementing it on two acquisition scenarios: a) Sequential Source, and b) Simultaneous Source. Mathematically, both approaches can be encoded by appropriately defining the measurement matrix $\mathbf{M}$. We used seismic lines represented as 3D volumes from the Gulf of Suez data (see Figure 1) to study the performance. The data contains 354×354 traces. The temporal length of the data was 4s with sampling interval of 0.004s, leading to maximum frequency of 125 Hz. For better visualization, we only show the 0-2s portion of the data. The seismic

band of 7-85Hz preserve most of the energy. We implemented all of the experiments in this band. We performed a two-part experiment for each scenario a) and b) above. First, we recovered low (10Hz) and high (70Hz) frequency 2D slices in (m-h) domain and compared it with the curvelet sparsity promoting l_1 norm minimization using SPGℓ_1 (van den Berg and Friedlander, 2008). We also obtained the optimal value of the regularization parameter and a reasonable range for the rank of the data matrix in (m-h) domain. We used these values in the second part of the experiment to implement the proposed approach on a complete 3D data set.

SCENARIO A. SEQUENTIAL SOURCE ACQUISITION

In this approach, we activate impulsive sources sequentially and acquire data for each source for a length of time t, known as *listening time*. The measurement operator $\mathbf{M}$ is an identity matrix with dimension $N = N_t \times N_s$, where N_t is the number of time samples, and N_s is the number of shots. The restriction operator is given by the Kronecker product $\mathbf{R} = \mathbf{R}_{n_s} \otimes \mathbf{I}_{N_t}$ that selects $n_s \leq N_s$ source indices uniformly randomly without replacement from $[1, \ldots, N_s]$. Note that while the restriction operator $\mathbf{R}$ can act on source and receiver coordinates simultaneously, in our experiments we restrict along the source coordinate only. Figures 3a and 3d show the low (10 Hz) and high (70 Hz) frequency slice in (s-r) domain for subsampling ratio $\delta = 0.5$. Figures 3b and 3e show the regularization result using the max-norm formulation. Figures 3c and 3f show the corresponding recoveries using SPGℓ_1. The SNR results are comparable. Figure 4 shows the plot of SNR versus Rank and Residual versus Rank for 10 Hz and 70 Hz frequency slices for several values of the regularization parameter. These plots show the effects of the regularization parameter and the rank on the quality of the recovery, and on the data misfit. We used the figure to obtain a range of rank values along with a good value of the regularization parameter for the 3D experiment. Figures 5a and 5c show a compressively sampled data slice of the 3D cube with 50% missing shots. Figures 5b and 5d show the inversion result for the data slice using the proposed max-norm approach. It is clear that missing traces were recovered with very low reconstruction error.

SCENARIO B: SIMULTANEOUS SOURCE ACQUISITION

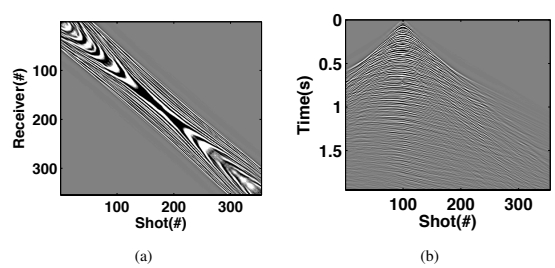
In this scenario, several *simultaneous shots* are obtained by activating all sources simultaneously with several realizations of random weights. The main goal is to recover the Green's function for the full acquisition using only simultaneous data. The measurement operator $\mathbf{M}$ is given by $\mathbf{G}_{N_s} \otimes \mathbf{I}_{N_t}$ where $\mathbf{G}_{N_s}$ is a Gaussian matrix $N_s \times N_s$ with $i.i.d$ entries. The restriction operator $\mathbf{R}$ is same as defined previously in the sequential scenario. In this case, we are only using 50% of simultaneous measurements for the recovery. Figures 6a and 6d show simultaneous shots uniformly selected at random with subsampling ratio $\delta = 0.5$ for low (10Hz) and high (70Hz) frequency slices. Figures 6b and 6e show the recovery using the max-norm formulation. Figures 6c and 6f show the corresponding recovery using SPGℓ_1. Again, the results are comparable. Figures 7a and 7b show the SNR versus Rank and Residual versus Rank for 10 Hz and 70 Hz frequency slices, for several values of the regularization parameter. As in the previous scenario, we used these plots to obtain a good value of regularization parameter, as well as a range of rank values for the complete 3D data experiment. Figures 8a and 8c show data slices of compressively sampled simultaneous shots with $\delta = 0.5$ uniformly selected at random from the 3D cube. Figures 8b and 8d show the recovery of the separated Green's function. We can clearly see that we recovered the Green's function with low reconstruction error.

DISCUSSION

We have presented a new method for the recovery of missing traces using rank penalization. The method relies on a novel recently proposed optimization formulation that works directly with matrix factors. The specific version we worked with fits the data subject to a constraint on the maximum squared column norms of the data, and can be solved in a straightforward way with a projected-gradient algorithm. The technique is very promising compared to current state of the art rank penalty methods, which rely on the SVD procedure. The experimental results on 2D and 3D data slices are comparable to those achieved using sparsity regularization with SPGℓ_1, and demonstrate the potential benefit of the methodology for seismic data interpolation.

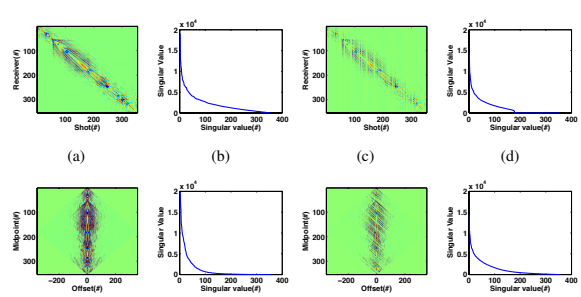
ACKNOWLEDGMENTS

We would like to thank Ben Recht for providing code of Max-Norm, also the authors of CurveLab, and SPGℓ_1. This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.



(a) (b)

Figure 1: Fully sampled sequential source acquisition data from Gulf of Suez with 354 shots, 354 receivers and 500 time samples. (a) Time slice at 1s. (b) Common receiver gather corresponding to y = 100.



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

(e) (f) (g) (h)

Figure 2: Singular value decay of data, displaying real part of frequency slice (70Hz). (a) Fully sampled data in (s-r) domain. (b) Singular value decay of data in (a). (c) Decimated Data with 50% of missing shots in (s-r) domain. (d) Singular value decay of data in (c).(e) Fully sampled data in (m-h) domain. (f) Singular value decay of data in (e). (g) Decimated Data with 50% of missing shots in (m-h) domain. (f) Singular value decay of data in (e).

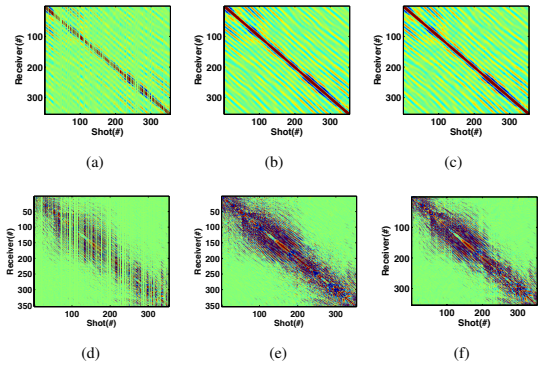


Figure 3: Recovery from compressively sampled 2D frequency slices with $\delta = 0.5$ in sequential source experiment. (a,d) Initial data prior to regularization for low (10Hz) and high (70Hz) frequency. (b,e) Regularization of incomplete data using the Max-Norm formulation, SNR = 18.6/7.4 db for low/high frequency. (c,f) Regularization of incomplete data using SPGℓ_1, SNR = 17.1/7.9 db for low/high frequency.

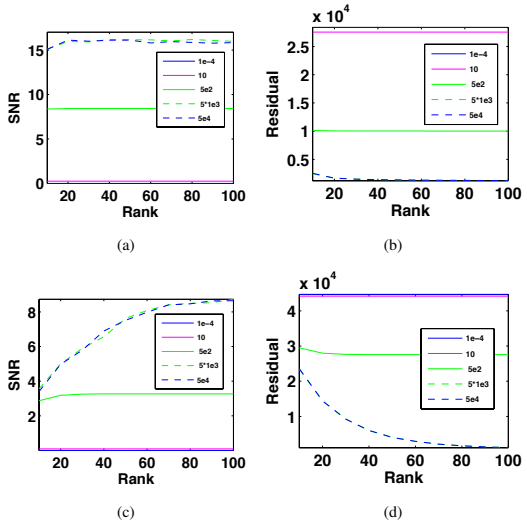


Figure 4: Signal-to-noise (SNRs) ratio and Residual as a function of rank in sequential source. (a,c) The SNRs are plotted as a function of rank for low (10Hz) and high (70Hz) frequency. (b,d) Residual are plotted as a function of rank for low (10Hz) and high (70hz) frequency.

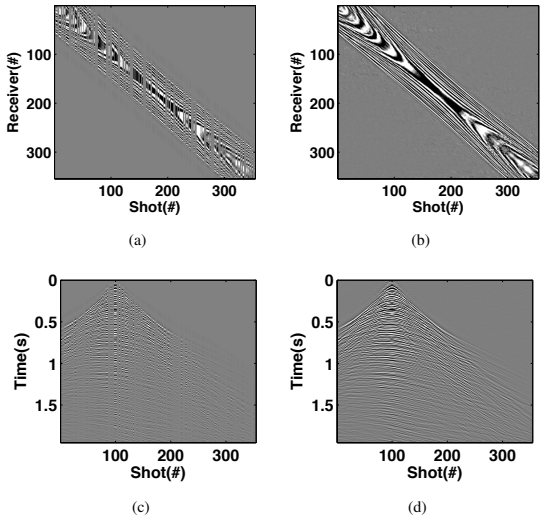


Figure 5: Recovery from compressively sampled 3D data with $\delta = 0.5$ in sequential source experiment . (a,c) Intial data prior to regularization. (b,d) Regularization of incomplete data using the Max-Norm formulation (SNR= 15.5 db)

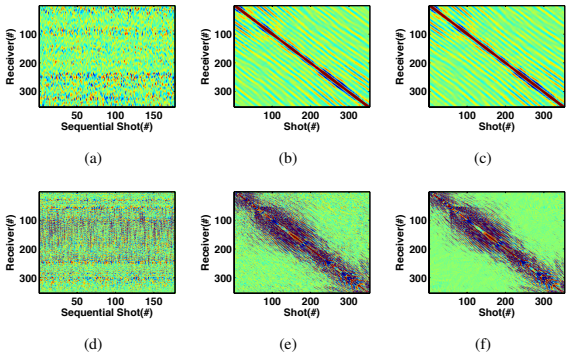


Figure 6: Recovery from compressively sampled 2D frequency slices with $\delta = 0.5$ in simultaneous soruce experiment. (a,d) Intial data with simultaneous shot uniformly selected at random for low(10Hz) and high(70Hz) frequency. (b,e) Recovery using the Max-Norm formulation, SNR = 23.81/11.1 db for low/high frequency. (c,f) Recovery using SPGℓ_1, SNR = 23.1/14.5 db for low/high frequency.

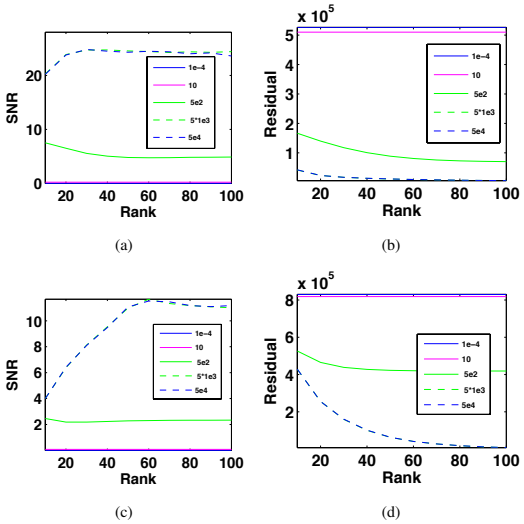


Figure 7: Signal-to-noise (SNRs) ratio and Residual as a function of rank in simultanoeus source. (a,c) The SNRs are plotted as a function of rank for low (10Hz) and high (70Hz) frequency. (b,d) Residual are plotted as a function of rank for low (10Hz) and high (70hz) frequency.

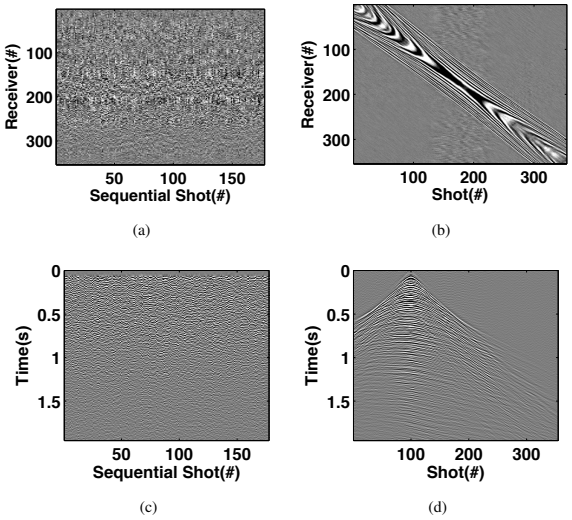


Figure 8: Recovery from compressively sampled 3D data with $\delta = 0.5$ simultaneous shots uniformly selected at random. (a,c) Intial data for random simultaneous shot. (b,d) Recovered Green function using the Max-Norm formulation (SNR=14.3db)

REFERENCES

Candes, E., and Y. Plan, 2010, Matrix completion with noise: Proceedings of the IEEE, **98**, 925 –936.

Candes, E. J., and D. L. Donoho, 2000, Curvelets - a surprisingly effective nonadaptive representation for objects with edges, *in* Curves and Surfaces: Vanderbilt University Press, Nashville, TN.

Donoho, D. L., 2006, Compressed sensing: **52**, 1289–1306.

Herrmann, F. J., 2008, Compressive sampling: a new paradigm for seismic data acquistion and processing?: Presented at the ION.

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

Lee, J., B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp, 2010, Practical large-scale optimization for max-norm regularization: Analysis, **23**, 1–9.

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

Recht, B., and C. Ré, 2011, Parallel Stochastic Gradient Algorithms for Large-Scale Matrix Completion: submitted.

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.

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.