

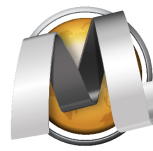
# Rank minimization via alternating optimization: seismic data interpolation

Oscar Lopez, Rajiv Kumar and Felix J. Herrmann

Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

## Abstract

Low-rank matrix completion techniques have recently become an effective tool for seismic trace interpolation problems. In this talk, we consider an alternating optimization scheme for nuclear norm minimization and discuss the applications to large scale wave field reconstruction. By adopting a factorization approach to the rank minimization problem we write our low-rank matrix in bi-linear form, and modify this workflow by alternating our optimization to handle a single matrix factor at a time. This allows for a more tractable procedure that can robustly handle large scale, highly oscillatory and critically subsampled seismic data sets. We demonstrate the potential of this approach with several numerical experiments on a seismic line from the Nelson 2D data set and a frequency slice from the Gulf of Mexico data set.



## Introduction

Densely sampled seismic data is crucial for accurate inversion and imaging procedures such as Full Waveform Inversion, Reverse Time Migration and multiple removal methods (EPSI, SRME) where otherwise subsampled sources and receivers would result in unwanted image artifacts. To this end, various missing trace interpolation methodologies have been proposed that exploit some low dimensional structure of seismic data such as sparsity (Herrmann and Hennenfent (2008), Mansour et al. (2013)) and low-rank (Yang et al. (2012), Aravkin et al. (2014)), two key themes of compressive sensing literature. Thus, low-rank matrix completion techniques (Candès and Plan (2009), Candès and Recht (2009)) have recently become relevant in seismic trace interpolation problems. In work by Aravkin et al. (2014), the authors utilize the low-rank structure of seismic data in midpoint-offset domain to allow for successful seismic data reconstruction. Furthermore, by allowing the data matrix to be factorized in bi-linear form as implemented by Recht et al. (2010), these authors propose LR-BPDN a rank penalizing formulation that is suitable for large scale seismic interpolation problems.

While this approach yields state of the art results, it is subject to suboptimal performance when faced with increasingly complicated problems due to large scale, oscillatory data, or highly subsampled data. Since such cases encompass a large part of practical interest we see that we would greatly benefit from a more robust optimization procedure. The goal of this talk is to improve LR-BPDN in these aspects by proposing a modified optimization scheme, we consider a rank minimization work flow that alternates the optimization procedure between the bi-linear matrix factors. In this way we tackle one matrix factor at a time, which allows for a more tractable procedure that can overcome these impediments without increasing time complexity. Experiments on the Nelson 2D seismic line and Gulf of Mexico 7 Hz frequency slice demonstrate the potential of the new approach.

## Matrix Completion: Large Scale Implementation

The goal of low-rank matrix completion is to accurately estimate the unobserved entries of a given incomplete data matrix from the observed entries and the prior knowledge that the matrix has few nonzero or quickly decaying singular values, i.e. the matrix exhibits low-rank structure. If we let  $\Omega$  be the set of observed entries, we can define our sampling operator  $P_\Omega$  by its action as:

$$P_\Omega(X) = \begin{cases} X_{ij} & \text{if } (i, j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

where  $X \in \mathbb{C}^{n \times m}$ . If  $X$  is our true matrix of interest, then our observations are given by  $b = P_\Omega(X) + n$ , where  $n$  encompasses our corruption by noise. As introduced by Fazel (2002) and shown by Recht et al. (2010) we can approximately solve the rank minimization problem by instead minimizing the best convex approximation of the rank function, the nuclear norm. This convex relaxation gives us our data estimate  $X^\sharp$  as the argument output of the optimization procedure:

$$\min_{Y \in \mathbb{C}^{n \times m}} \|Y\|_* \quad \text{s.t.} \quad \|P_\Omega(Y) - b\|_F \leq \varepsilon \quad (1)$$

where  $\|n\|_F \leq \varepsilon$  is the noise level in our observations,  $\|Y\|_*$  is the nuclear norm of the matrix  $Y$  and  $\|X\|_F$  denotes the Frobenius norm of the matrix  $X$ . This procedure exploits the knowledge that  $X$  has low-rank structure while making sure we agree with our observations up to some known noise level.

Many implementations to solve (1), such as singular value projection and singular value thresholding (Jain et al. (2010), Cai et al. (2010)) require numerically expensive SVD computations of our data matrix at each iteration and furthermore require full matrix storage. Since these issues can make (1) impractical for large data matrices, Recht et al. (2010) and Aravkin et al. (2014) propose a matrix factorization approach to overcome this impediment. By choosing  $r \ll \min(n, m)$ , the authors write the matrix in bi-linear form  $X = LR^H$ , where  $L \in \mathbb{C}^{n \times r}$ ,  $R \in \mathbb{C}^{m \times r}$  and  $R^H$  denotes the Hermitian transpose of  $R$ . As shown by Srebro (2004) and Recht et al. (2010), the fact that  $\|X\|_* = \min_{\{L, R: X=LR^H\}} \frac{1}{2}(\|L\|_F^2 + \|R\|_F^2)$



allows us to solve (1) by instead solving:

$$\min_{L \in \mathbb{C}^{n \times r}, R \in \mathbb{C}^{m \times r}} \frac{1}{2} (\|L\|_F^2 + \|R\|_F^2) \quad \text{s.t.} \quad \|P_\Omega(LR^H) - b\|_F \leq \varepsilon. \quad (2)$$

If we choose our factorization parameter,  $r$ , significantly smaller than the ambient dimensions  $n$  and  $m$ , we have reduced our memory requirements from  $mn$  to  $rn + rm$  while implementing a rank penalization scheme that avoids computing SVD of our potentially large data matrix.

In this talk we adopt formulation (2) and consider the following algorithm for this purpose:

---

**Algorithm 1** Rank Minimization via Alternating Optimization

---

**Require:**  $P_\Omega, b, r, T$

- 1: **initialize:**  $L^0$  to be the top- $r$  singular vectors of  $b$
  - 2: **for**  $t = 0$  to  $T - 1$  **do**
  - 3:  $R^{t+1} \leftarrow \min_{\{R \in \mathbb{C}^{m \times r}\}} \frac{1}{2} \|R\|_F^2 \quad \text{s.t.} \quad \|P_\Omega(L^t R^H) - b\|_F \leq \varepsilon$
  - 4:  $L^{t+1} \leftarrow \min_{\{L \in \mathbb{C}^{n \times r}\}} \frac{1}{2} \|L\|_F^2 \quad \text{s.t.} \quad \|P_\Omega(L(R^{t+1})^H) - b\|_F \leq \varepsilon$
  - 5: **end for**
  - 6: **return**  $(X^\sharp = L^T (R^T)^H)$
- 

The key component of Algorithm 1, is to allow our optimization to focus on a single matrix factor at a time in steps 2 – 5. As we will see in the experiments section, this choice of implementation results in a favorable output for cases where our data matrix is large, highly oscillatory or critically subsampled.

Similar alternating techniques have been proposed and discussed by Jain et al. (2013) and Xu and Yin (2013). However the approach considered by Jain et al. (2013) does not penalize the rank and while Algorithm 1 is a consequence of the work in Xu and Yin (2013), the authors do not provide a specific implementation to achieve steps 3 and 4. For this purpose we utilize a generalization of the SPGL<sub>1</sub> solver (Berg and Friedlander (2008)) which at step  $t$  given  $L^t$  solves step 3 of Algorithm 1 by solving a sequence of LASSO subproblems:

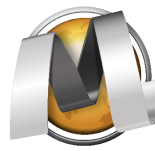
$$\min_{\{R \in \mathbb{C}^{m \times r}\}} \|P_\Omega(L^t R^H) - b\|_F \quad \text{s.t.} \quad \frac{1}{2} \|R\|_F^2 \leq \tau$$

Where  $\tau$  is a regularization parameter updated appropriately by SPGL<sub>1</sub> so that the LASSO optimization procedure is equivalent to solving step 3. Solving such LASSO subproblems requires us to minimize a convex function ( $\|P_\Omega(L^t R^H) - b\|_F$  as a function of  $R$  only) and project onto  $\frac{1}{2} \|R\|_F^2 \leq \tau$ , which is easily achieved by appropriate scalar multiplication. We thus end up with a sequence of tractable convex problems that do not require expensive SVD computations and effectively penalize the rank. Solving step 4 of Algorithm 1 can be done in a similar manner.

### Seismic Data Reconstruction via Rank Minimization: Experiments

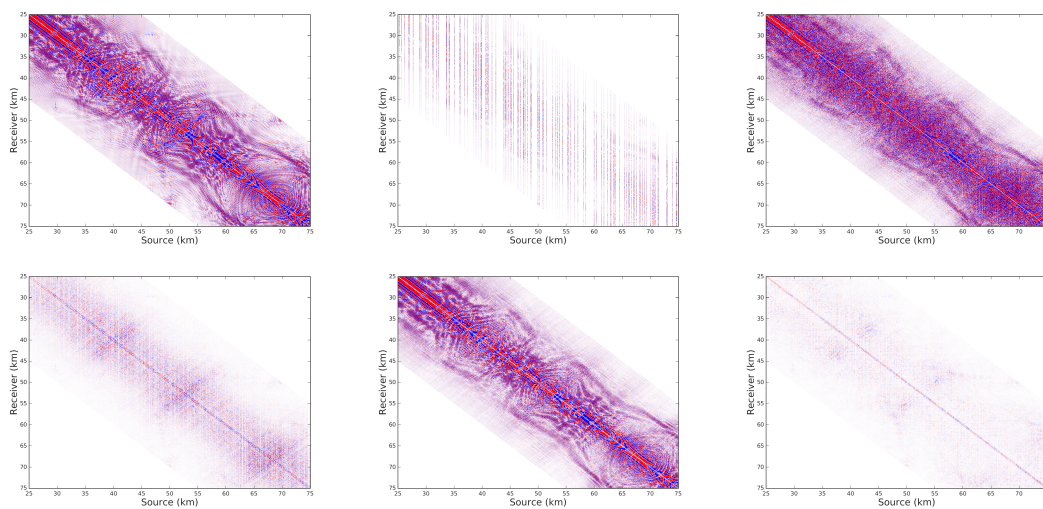
We allow our monochromatic frequency slice,  $X \in \mathbb{C}^{n \times m}$ , to be organized as a matrix in midpoint-offset (m-h) domain. As Aravkin et al. (2014) discuss, implementing (2) in m-h domain allows us to achieve low-rank structure and furthermore results in successful low-rank matrix completion since in m-h domain missing sources and receivers result in missing matrix entries whose structure disrupts the low-rank properties of our complete matrix. Furthermore we simulate the effect of missing traces by applying a subsampling mask that removes a desired percentage of sources in a jittered random manner. As mentioned by Kumar et al. (2014) jittered subsampling allows for randomness in our sampling scheme, a key component in compressive sensing applications, while controlling the allowed gap sizes of our missing data, an important consideration for interpolation techniques.

We offer two experiments to demonstrate our interpolation scheme, where for the sake of comparison



we also implement LR-BPDN proposed by Aravkin et al. (2014). The first is conducted on a single 7 Hz frequency slice extracted from the Gulf of Mexico dataset with 4001 sources and 4001 receivers. We apply a sampling mask that removes 90% of sources in a jittered random manner. For Algorithm 1 we choose a total of  $T = 10$  alternations with 15 iterations of  $SPG_{l_1}$  per alternation. For fairness of comparison we run LR-BPDN with 300 iterations of  $SPG_{l_1}$  and choose factorization parameter  $r = 80$  for both methods. The results (Figure 1) demonstrate a drastic benefit offered by our approach. Results of LR-BPDN introduce more noise than signal in the reconstruction with an SNR of -9 dB, whereas the results of Algorithm 1 give output with SNR of 6.7 dB. This critical difference in results shows how our proposed method does indeed improve results in large scale and highly subsampled cases.

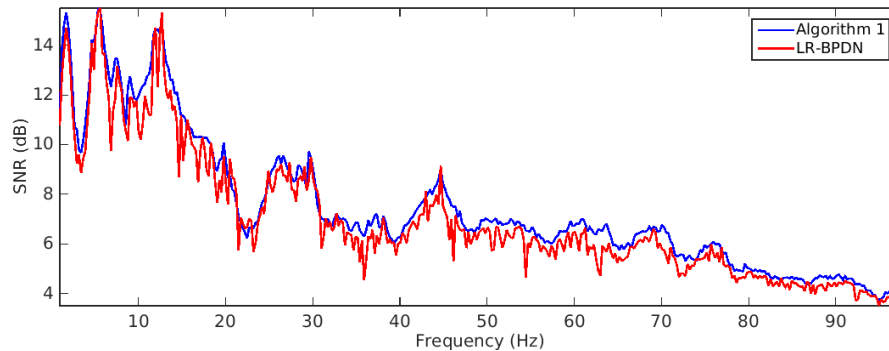
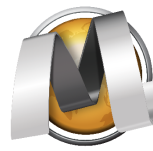
The second experiment is conducted on a seismic line from the Nelson dataset with 401 sources, 401 receivers and 1024 time samples where we interpolate in the 1-97 Hz frequency band. For this experiment we subsample by removing 80% of sources in a jittered random manner. In this case we choose  $T = 6$  alternations with 15 iterations of  $SPG_{l_1}$  per alternation for Algorithm 1, 180 iterations of  $SPG_{l_1}$  for LR-BPDN and choose the same factorization parameters adjusted from low to high frequency. We observe in Figure 2 that implementation of Algorithm 1 offers an average improvement in SNR by a factor of .5 dB when compared to LR-BPDN. We further notice that reconstruction improvements are most apparent in the higher frequency slices (50-70 Hz), which enforces our previous statement that Algorithm 1 is most beneficial in cases where our data is highly oscillatory and critically subsampled. We remind the reader that while this might not seem like a critical improvement, we are not increasing the time complexity of LR-BPDN.



**Figure 1** Comparison of interpolation results on a 7 Hz frequency slice shown in source-receiver domain from the Gulf Of Mexico dataset. (top left) Ground truth. (top middle) 90% jitter random subsampled ground truth data. (top right) Recovery via LR-BPDN with SNR of -9 dB. (bottom left) Difference plot of LR-BPDN recovery. (bottom middle) Recovery via Algorithm 1 with SNR of 6.7 dB. (bottom right) Difference plot of Algorithm 1 recovery.

## Conclusions

We proposed a modified optimization workflow for seismic data interpolation via low-rank matrix completion. In order to implement the nuclear norm minimization problem (1), we proceed as Aravkin et al. (2014) but offer a different rank minimization approach. By allowing our optimization scheme to alternate between matrix factors and implementing the Pareto curve approach ( $SPG_{l_1}$ ) we are left with a sequence of tractable convex problems that do not require expensive SVD computations and effectively penalize the rank. This results in a more robust procedure, where the experimental results clearly highlight the reconstruction SNR improvements in cases where our data matrix is large, highly oscillatory



**Figure 2** Plot of frequency versus recovery SNR from implementation of Algorithm 1 and LR-BPDN on the seismic line from the Nelson dataset for 1-97 Hz frequency band.

and critically subsampled when compared to the LR-BPDN implementation.

### Acknowledgements

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, Sub Salt Solutions, WesternGeco, and Woodside.

### References

- Aleksandr Y. Aravkin, Rajiv Kumar, Hassan Mansour, Ben Recht, and Felix J. Herrmann [2014], Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation. *SIAM Journal on Scientific Computing*, **36**, S237-S266.
- Benjamin Recht, Maryam Fazel, and Pabla A. Parrilo [2010], Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, **52**, 471-501.
- Emmanuel J. Candès, and Benjamin Recht [2009], Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, **9**, 717-772.
- Emmanuel J. Candès, and Yaniv Plan [2009], Matrix Completion With Noise. *Proceedings of the IEEE*, **98**, 925 - 936.
- Ewout van den Berg and Michael P. Friedlander [2008], Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31(2)**, 890-912.
- Felix J. Herrmann and Gilles Hennenfent [2008], Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, **173**, 233-248.
- Hassan Mansour, Felix J. Herrmann, and Ozgur Yilmaz [2013], Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization. *Geophysics*, **78**, V193-V206.
- Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen [2010], A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, **20**, 1956-1982.
- Maryam Fazel [2002], Matrix rank minimization with applications. *Electrical Engineering Department Stanford University*, **54**, 1-130.
- Nathan Srebro [2004], Learning with Matrix Factorizations. PhD Thesis, *Massachusetts Institute of Technology*.
- Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi [2013], Low-rank matrix completion using alternating minimization. *STOC '13 Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 665-674.
- Prateek Jain, Raghu Meka, and Inderjit S. Dhillon [2010], Guaranteed Rank Minimization via Singular Value Projection. *Advances in Neural Information Processing Systems*, **23**, 937-945.
- Rajiv Kumar, Aleksandr Y. Aravkin, Ernie Esser, Hassan Mansour, and Felix J. Herrmann [2014], SVD-free low-rank matrix factorization : wavefield reconstruction via jittered subsampling and reciprocity. *EAGE*.
- Yangyang Xu and Wotao Yin [2013], A block coordinate descent method for regularized multi-convex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences*, **6(3)**, 1758-1789.
- Yi Yang, Jianwei Ma, and Stanley Osher [2012], Seismic data reconstruction via matrix completion. *UCLA CAM Report*, 12-14.