Seismic data interpolation and denoising using SVD-free low-rank matrix factorization

R. Kumar[1], A.Y. Aravkin[1,2], H. Mansour[1,3] , B. Recht[4] and F.J. Herrmann[1]
[1] Dept. of Earth and Ocean sciences, University of British Columbia, Vancouver, BC, Canada
[2] Dept. of Computer Science, University of British Columbia, Vancouver, BC, Canada
[3] Dept. of Mathematics, University of British Columbia, Vancouver, BC, Canada
[4] Dept. of Computer Science, University of Wisconsin, Madison, USA .

January 14, 2013

Abstract

Recent developments in rank optimization have allowed new approaches for seismic data interpolation and denoising. In this paper, we propose an approach for simultaneous seismic data interpolation and denoising using robust rank-regularized formulations. The proposed approach is suitable for large scale problems, since it avoids SVD computations by using factorized formulations. We illustrate the advantages of the new approach using a seismic line from Gulf of Suez and 5D synthetic seismic data to obtain high quality results for interpolation and denoising, a key application in exploration geophysics.

Introduction

In exploration seismology, extremely large amounts of data (often on the order of petabytes) must be acquired and processed in order to determine the structure of the subsurface. In many situations, only a subset of the complete data is acquired due to physical and/or budgetary constraints. Most of the time, acquired data is also contaminated with noise. Therefore, interpolation and noise attenuation are considered key pre-processing steps for removing artifacts and increasing the spatial resolution, and for migration. State of the art trace interpolation and denoising schemes transform data into sparsifying domains, for example using the Fourier (Sacchi et al. (1998)) and Curvelet (Herrmann and Hennenfent (2008)) transforms. The underlying sparse structure of the data can then be exploited to de-noise the data and recover missing traces. Moreover, when organized in a matrix, seismic data often exhibits low-rank structure, i.e. small number of nonzero singular values, or quickly decaying singular values. Therefore, it is possible to perform missing-trace interpolation and denoising via rank-minimization strategies. Oropeza and Sacchi (2011) identified that seismic temporal frequency slices organized in block Hankel matrices are low rank. Additive noise and missing traces increase the rank of the block Hankel matrix, and the authors use a low rank approximation of the Hankel matrix via the randomized singular value decomposition (Liberty et al. (2007); Halko et al. (2011)) to consequently de-noise and interpolate seismic temporal frequency slices. While this technique is effective in interpolating and denoising the seismic data, the approach requires embedding the data into an even larger space where each dimension of size n is mapped to a matrix of size $n \times n$. Another approach is to generalize rank minimization to tensor representations as proposed in another contribution by one of the authors to the proceedings of this conference.

Rank-minimization approaches in seismic data processing have two main challenges that must be addressed. The first challenge is to find a suitable transformation or representation where the fully sampled seismic-data matrix has low-rank structure, while the subsampled seismic-data matrix does not, i.e. missing data increases the rank or decreases the decay rate of the singular values. The second challenge is computational, since rank minimization problems rely on the singular value decomposition (SVD), which is prohibitively expensive for large matrices.

In this paper, we present a method that exploits the low-rank structure of seismic data, and solves the interpolation and denoising problem simultaneously. We first show that transforming the data into the midpoint-offset domain increases the decay rate of its singular values compared to the source-receiver domain. We then formulate an optimization problem that simultaneously minimizes the rank, and uses a robust misfit function to aid in denoising. Moreover, we avoid the prohibitive computational complexity of performing SVDs on large matrices and instead use a matrix factorization approach recently developed by Lee et al. (2010). Finally, we demonstrate the efficacy of the new method using seismic lines from the Gulf of Suez and on 5D synthetic seismic data generated by BG.

Regularized Matrix Factorization

Let X_0 be a matrix in $\mathbb{C}^{n\times m}$ and let $\mathscr{A}$ be a linear measurement operator that maps from $\mathbb{C}^{n\times m} \rightarrow \mathbb{C}^p$. Recht et al. (2010) showed that under certain general conditions on the operator $\mathscr{A}$, the solution to the rank minimization problem can be found by solving the following nuclear norm minimization problem:

$$\min_X ||X||_* \quad \text{s.t. } \|\mathscr{A}(X) - b\|_2 \leq \varepsilon. \qquad (\text{BPDN}_\sigma)$$

where $b = \mathscr{A}(X_0) + e$ is a set of noisy measurements, e is a random noise vector, $\|X\|_* = \|\sigma\|_1$, and σ is the vector of singular values. The least-squares misfit is a good choice when the noise vector e is effectively modeled by the Gaussian distribution. However, in most practical situations, the measurement noise typically has many outliers, and it does not obey the Gaussian assumption. Aravkin et al. (2012b) studied the effect of using robust misfit functions for recovering non-negative undersampled sparse signals from noisy measurements with large outliers. Their study showed that Huber and Student's t misfits worked well in this setting, with Student's t performing as well or better than Huber. For large scale seismic inverse problems, Aravkin et al. (2012a) showed that the Student's t penalty function outperformed the least-squares and Huber penalty when in the presence of extreme data contamination. Motivated by these results, we consider the following formulation of robust nuclear norm minimization

$$\min_X ||X||_* \quad \text{s.t. } \rho(\mathscr{A}(X)-b)\le\varepsilon, \tag{rBPDN$_\sigma$}$$

where ε is an estimate of the noise level, and ρ is a robust penalty. While the methodology we propose works with multiple penalties, including Huber and hybrid penalties (proposed in Bube and Langan (1997)), we focus on the Student's t penalty

$$\rho(r)=\log(1+r^2/\nu)\,, \tag{1}$$

where ν is the degrees of freedom parameter that controls the sensitivity of ρ to outlier noise (for large ν, the student's t distribution approaches the Gaussian distribution). In order to efficiently solve (rBPDN$_\sigma$), we use an extension of the SPGℓ_1 solver (Berg and Friedlander, 2008) developed for the (rBPDN$_\sigma$) problem in Aravkin et al. (2012b). The SPGℓ_1 algorithm finds the solution to the (rBPDN$_\sigma$) by solving a sequence of robust LASSO subproblems

$$\min_X \rho(\mathscr{A}(X)-b) \quad \text{s.t. } ||X||_*\le\tau, \tag{rLASSO$_\tau$}$$

where τ is updated by traversing the Pareto curve. Solving each robust LASSO subproblem requires a projection onto the nuclear norm ball $\|X\|_*\le\tau$ in every iteration by performing a singular value decomposition and then thresholding the singular values. In the case of large scale seismic problems, it becomes prohibitive to carry out such a large number of SVDs. Instead, we propose to adopt a recent factorization-based approach to nuclear norm minimization introduced by (Rennie and Srebro; Lee et al. (2010); Recht and Ré (2011)). The factorization approach parametrizes the matrix $X\in\mathbb{C}^{n\times m}$ as the product of two low rank factors $L\in\mathbb{C}^{n\times k}$ and $R\in\mathbb{C}^{m\times k}$, such that,

$$X = LR^H, \tag{2}$$

where the superscript H indicates the Hermitian transpose. The optimization framework can then be carried out using the factors L and R instead of X, thereby significantly reducing the size of the decision variable from nm to $k(n+m)$ when $k\ll m,n$. In (Rennie and Srebro), it was shown that the nuclear norm obeys the relationship

$$\|X\|_* = \|LR^T\|_* \le \frac{1}{2}\left\|\begin{bmatrix}L\\R\end{bmatrix}\right\|_F^2 =: \Phi(L,R). \tag{3}$$

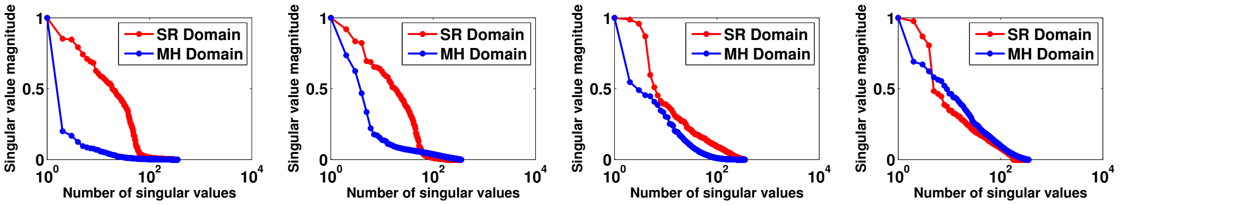
where $\|\cdot\|_F^2$ is Frobenius norm of the matrix (sum of the squared entires). Consequently, the robust LASSO subproblem can be replaced by

$$\min_{L,R} \rho(\mathscr{A}(LR^T)-b) \quad \text{s.t. } \Phi(L,R)\le\tau\,. \tag{4}$$

where the projection onto $\Phi(L,R)\le\tau$ is easily achieved by multiplying each factor L and R by the scalar $2\tau/\Phi(L,R)$. By (3), we are guaranteed that $\|LR^T\|_*\le\tau$ for any solution of (4).

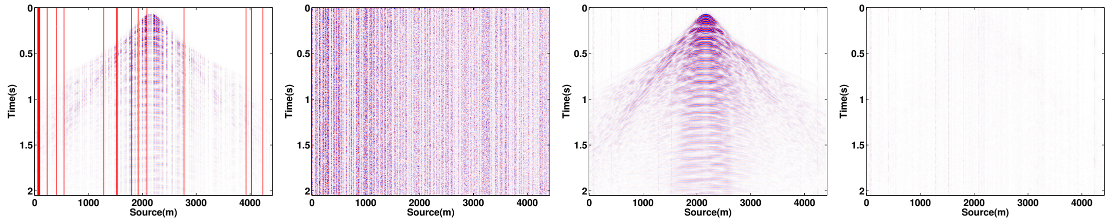
Seismic data interpolation and denoising

We implement the proposed formulation for two different acquisition examples. The first example is a seismic line from Gulf of Suez with $N_s = 354$ sources, $N_r = 354$ receivers, $N_t = 1024$ time samples and a sampling interval of 0.004s. Most of the energy of the seismic line is concentrated in the 12-60Hz frequency band. In order to interpolate and denoise simultaneously, we apply a sub-sampling mask that randomly removes 50% of the shots, and replace another 10% of the shots with large random errors, whose amplitudes are three times the maximum amplitude present in the data. In this example, we use the transformation from the source-receiver (s-r) domain to the midpoint-offset (m-h) domain.



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

Figure 1 *Singular value decay of fully sampled (a) low frequency slice at 12 Hz and (c) high frequency slice at 60 Hz in (s-r) and (m-h) domains. Singular value decay of 50% subsampled (b) low frequency slice at 12 Hz and (d) high frequency data at 60 Hz in (s-r) and (m-h) domains.*

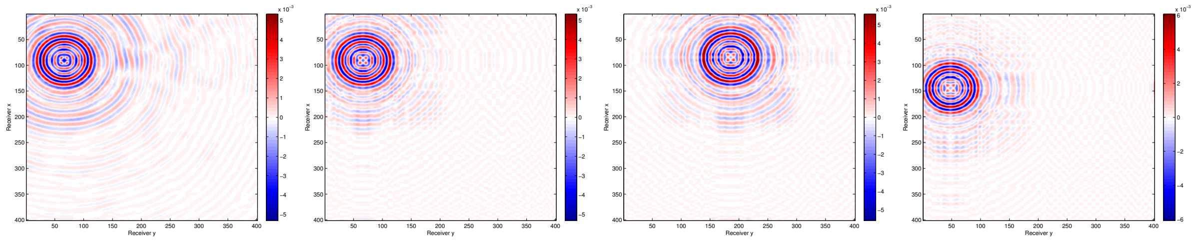


(a) (b) (c) (d)

Figure 2 *Comparison of interpolation and denoising results for the Student's t and least-squares misfit function. (a) 50% subsampled common receiver gather with another 10 % replaced by large errors. (b) Recovery result using the least-squares misfit function. (c,d) Recovery and residual results using the student's t misfit function with a SNR of 17.2 dB.*

To show that the midpoint-offset transformation is a good choice, we plot the singular values decay of monochromatic frequency slices at 12Hz and 60Hz in the source-receiver and midpoint-offset domain. Notice that the decay of singular values in both frequency slices are faster in the midpoint-offset domain. Sub-sampling does not noticeably change the decay of singular value in the (s-r) domain; but destroys the fast decay of singular values in the (m-h) domain, an essential feature for interpolation and denoising using nuclear-norm minimization. In this example, we use the student's t misfit function and implement the proposed formulation in the frequency domain, where we work with monochromatic frequency slices and adjust the rank and ν parameter while going from low to high frequency slices. Figure 2 compares the recovery results with and without using a robust penalty function. We can clearly see that results with Student's t are better than those with least-squares. In each example, we use 300 iterations of SPGℓ_1 for all frequency slices.

In the second acquisition example, we implement the proposed formulation on the 5D synthetic seismic data provided by BG. We select one 4D monochromatic frequency slice at 7 Hz. Each monochromatic frequency slice has 400×400 receivers spaced by 25m and 200 sources out of 4624 spaced by 150m. We are missing 97% of data in this case. In this example the transform domain is simply the permutation of source and receivers coordinates where matricization of each 4D monochromatic frequency slices is done as per (SourceX, ReceiverX) and (SourceY, ReceiverY) coordinates. We observed the same behaviour of singular value decay as in Figure 1, when we matricized each 4D monochromatic frequency slices as [(SourceX,ReceiverX),(SourceY,ReceiverY)] instead of [(SourceX,SourceY), (ReceiverX,ReceiverY)] coordinates. We use rank 10 for the interpolation, and run the experiment for 500 iterations. Figure 3a and 3b shows the interpolation results at one location of the acquisition grid where a shot is recorded, while Figures 3c and 3d show the interpolation results at the grid location where no reference shot is present. It is very clear that we can perform interpolation and denoising simultaneously with low reconstruction error.



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

Figure 3 *Interpolation results for a single monochromatic frequency slice at 7Hz from 5D synthetic seismic data. (a,b) Original and recovery of a common shot gather with a SNR of 8.8 dB at the location where shot is recorded. (c,d) Interpolation of common shot gathers at the location where no reference shot is present.*

Conclusions

We have presented a new method for seismic data interpolation and denoising using robust nuclear-norm minimization. The method combines the Pareto curve approach for optimizing rBPDN$_\sigma$ formulations with the SVD-free matrix factorization methods. The resulting formulation can be solved in a straightforward way using SPGℓ_1. The technique is very promising, since it is SVD-free and therefore may be used for very large-scale systems. The experimental results on 3D and 5D seismic data set demonstrate the potential benefit of the methodology.

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, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

References

Aravkin, A., Friedlander, M.P., Herrmann, F. and van Leeuwen, T. [2012a] Robust inversion, dimensionality reduction, and randomized sampling. *Mathematical Programming*, **134**(1), 101–125.

Aravkin, A., Burke, J. and Friedlander, M. [2012b] Variational properties of value functions. *submitted to SIAM J. Opt., arXiv:1211.3724*.

Berg, E.v. and Friedlander, M.P. [2008] Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31**(2), 890–912.

Bube, K.P. and Langan, R.T. [1997] Hybrid ℓ_1/ℓ_2 minimization with applications to tomography. *Geophysics*, **62**(4), 1183–1195.

C. Da Silva, F.H. [2013] Hierarchical tucker tensor optimization - applications to 4d seismic data interpolation. *75th EAGE Conference, Extended Abstracts*, vol. Submitted.

Halko, N., Martinsson, P.G. and Tropp, J.A. [2011] Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, **53**(2), 217–288.

Herrmann, F.J. and Hennenfent, G. [2008] Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, **173**(1), 233–248, ISSN 1365-246X.

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

Liberty, E., Woolfe, F., Martinsson, P.G., Rokhlin, V. and Tygert, M. [2007] Randomized algorithms for the low-rank approximation of matrices. **104**(51), 20167–20172, doi:10.1073/pnas.0709640104.

Oropeza, V. and Sacchi, M. [2011] Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, **76**(3), V25–V32.

Recht, B., Fazel, M. and Parrilo, P. [2010] Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Review*, **52**(3), 471–501.

Recht, B. and Ré, C. [2011] Parallel stochastic gradient algorithms for large-scale matrix completion. *Optimization Online*.

Rennie, J.D.M. and Srebro, N. [????] *ICML '05 Proceedings of the 22nd international conference on Machine learning*.

Sacchi, M., Ulrych, T. and Walker, C. [1998] Interpolation and extrapolation using a high-resolution discrete fourier transform. *Signal Processing, IEEE Transactions on*, **46**(1), 31 –38.