Relaxation algorithms for matrix completion

with applications to seismic travel-time data interpolation

Robert Baraldi[1], **Carl Ulberg**[2], **Rajiv Kumar**[3], **Kenneth Creager**[2] **and Aleksandr Aravkin**[1]

[1]Department of Applied Mathematics, University of Washington
[2]Department of Earth and Space Sciences, University of Washington
[3]School of Earth and Atmospheric Sciences, Georgia Institute of Technology

Acknowledgments: The authors acknowledge support from the Department of Energy Computational Science Graduate Fellowship, which is provided under grant number DE-FG02-97ER25308.

Introduction

- Travel-time tomography is used to determine the underlying structure of the earth in exploratory and global seismology by solving a data fitting (inverse) problem.
- Data quality and availability is a key constraint, motivating denoising and interpolation techniques prior to inversion.
- We analyze data obtained by the Imaging Magma Under St. Helens (iMUSH) project, a multi-year effort to image and infer the architecture of the greater Mount St. Helens magmatic system [1].
- iMUSH data collected:
  - Active experiments initiate seismic tremors and record waves;
  - Passive experiments record seismic waves from earthquakes.
- Our contributions:
  - A new formulation that allows a target data misfit, smoothness, and low-rank regularization.
  - An efficient algorithm for the new formulation competitive with state-of-the-art for denoising and interpolation methods.

Mathematical Background

Formulations for Low-rank Interpolation

- Low-rank matrix completion estimates the missing data entries of $X \in \mathbb{R}^{m\times n}$ from observed entries.
- For $\mathcal{T} \subset \{1,\ldots,n\}\times\{1,\ldots,m\}$ the set of observed entries, we can describe the sampling operator $\mathcal{A}:\mathbb{R}^{n\times m}\to\mathbb{R}^{n\times m}$

$$\mathcal{A}(X) = \begin{cases} X_{ij}, & \text{if } (i,j)\in\mathcal{T}, \\ 0, & (i,j)\notin\mathcal{T} \end{cases}$$

- Observed data is $b = \mathcal{A}(X) + \epsilon$ for $\epsilon = \mathcal{A}(\epsilon) \in \mathbb{R}^{n\times m}$.
- Rank proxy is the nuclear norm $\|X\|_* = \sum_{j=1}^{\min(n,m)} \sigma_j(X)$
- The classic formulations that balance data fit with regularization are

$$\min_{X\in\mathbb{R}^{n\times m}} \|X\|_* + \tfrac{1}{\sigma}\|\mathcal{A}(X)-b\|_2 \quad (1)$$

$$\min_{X\in\mathbb{R}^{n\times m}} \|\mathcal{A}(X)-b\|_2 \quad \text{s.t.} \quad \|X\|_* \le \tau \quad (2)$$

$$\min_{X\in\mathbb{R}^{n\times m}} \|X\|_* \quad \text{s.t.} \quad \|\mathcal{A}(X)-b\|_2 \le \sigma. \quad (3)$$

are known as Tichonoff, Ivanov, and Morozov regularizations [2].
- The misfit-constraint variant (3) is most suited for situations where a good estimate of the 'noise floor' σ is available.

Smoothness Constraints and Factorized Formulation

- Smoothness/continuity (a local property) between gridpoints is desired in travel time tomography since geological structure of the crust exhibits the traits of approximately homogeneous media.
  - Implemented with penalty term $\frac{1}{2\gamma}\|\mathcal{L}(X)\|_2^2$, where $\mathcal{L}$ the discretization of the Laplacian operator
  - Smoothness and low-rank combine local and global information.
- An efficient alternative to the SVD needed in optimizing the convex $\|X\|_*$ is the matrix factorization formulation [3]
  - Explicitly decompose $X = LR^T$, with $L\in\mathbb{R}^{n\times k}, R\in\mathbb{R}^{m\times k}$.
  - From [4], the *variational* characterization

$$\|X\|_* = \inf_{L,R:X=LR^T} \frac{1}{2}(\|L\|_F^2 + \|R\|_F^2)$$

  lets us replace $\|X\|_*$ by $\frac{1}{2}(\|L\|_F^2+\|R\|_F^2)$ where $k \ll \min(n,m)$
  - Memory requirements reduce from mn to kn+km with no SVDs.
- Our goal here is to solve the factorized Morozov formulation

$$\min_{L,R} \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 + \frac{1}{2\gamma}\|\mathcal{L}(LR^T)\|_2^2 \quad \text{s.t.} \quad \|\mathcal{A}(LR^T)-b\|_2 \le \sigma \quad (4)$$

with both local and global structure, and a misfit target σ.

Relaxed Low-Rank & Smooth Inversion

- Our novel method for Problem (4) is a *relaxation* by way of [5].
  - We introduce an auxiliary variable $W \approx LR^T$:

$$\min_{L,R,W} \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 + \frac{1}{2\gamma}\|\mathcal{L}(W)\|_2^2 + \frac{1}{2\eta}\|W - LR^T\|_F^2 \quad \text{s.t.} \quad \|\mathcal{A}(W)-b\|_2 \le \sigma. \quad (5)$$

  - Problem (5) is a relaxation for Problem (4), as W approximates $X = LR^T$, with $\|W - LR^T\| = \mathcal{O}(\eta)$.
  - This preserves features of (4) and is solved via block-coordinate descent in Algorithm 1.

Algorithm 1 Block-Coordinate Descent for (5).

1: **Input:** W_0, L_0, R_0
2: **while** not converged **do**
3: $L_+ \leftarrow \left(I + \eta R^T R\right)^{-1}(\eta W R)$
4: $R_+ \leftarrow (\eta W^T L_+)\left(I + \eta L_+^T L_+\right)^{-1}$
5: $W_+ \leftarrow \arg\min_W \frac{1}{2\gamma}\|\mathcal{L}(W)\|_2^2 + \frac{1}{2\eta}\|W - L_{k+}R_+^T\|_F^2$ s.t. $\|\mathcal{A}(W)-b\|_2 \le \sigma$
6: **Output:** W, L, R

  - Step 6 is equivalent to the quadratic trust-region subproblem for $\sigma > 0$, and can be solved using an efficient root-finding method.
  - The $\sigma = 0$ case has a closed-form solution for W_+.

FISTA & L-BFGS for Low-Rank & Smooth Inversion

- A simple convex formulation that uses smoothness, data misfit, and a rank proxy (nuclear norm) is given by

$$\min_X \|\mathcal{A}(X)-b\|^2 + \frac{1}{2\gamma}\|\mathcal{L}(X)\|^2 + \tau\|X\|_*. \quad (6)$$

  - Solved via Fast Iterative Shrinkage-Thresholding Algorithm [7].
  - The step size α is the reciprocal of the largest singular value of $(\mathcal{A}^*\mathcal{A} + \gamma^{-1}\mathcal{L}^T\mathcal{L})$, and $S_{\alpha\tau}$ is the soft-thresholding operator:

$$S_{\alpha\tau}(\Sigma)_{ii} = \max(0, \Sigma_{ii} - \alpha\tau).$$

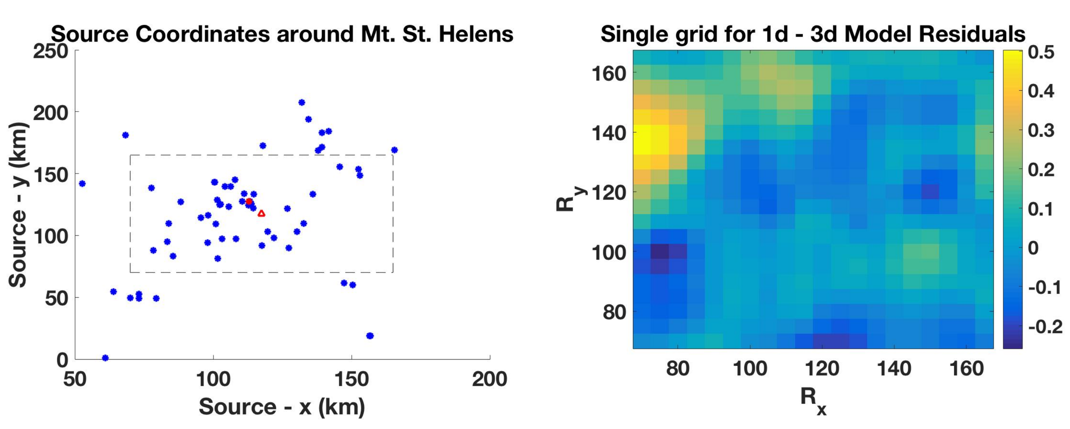
  - Need gradient of smooth terms: $\mathcal{A}$, $\mathcal{L}$ and their adjoints
  - Need prox operator of $\|\cdot\|_*$, which requires thresholding on singular values computed via SVD.
  - Prohibitively expensive as the dimensions of X grow.
- SVDs are avoided with variational factorization and L-BFGS [6]

$$\min_{L,R} \|\mathcal{A}(LR^T)-b\|^2 + \frac{1}{2\gamma}\|\mathcal{L}(LR^T)\|^2 + \frac{\tau}{2}\|L\|_F^2 + \frac{\tau}{2}\|R\|_F^2. \quad (7)$$

  - Smooth with respect to the decision variables L and R

Tensor Formulation

- We matricize data by tensoring the receiver grid for each source
  - Each receiver grid is a 2D, 95 kilometer mesh with 5 kilometer spacing centered on the mountain.
  - We focus on synthetic residuals Figure 1(b) of the nonlinear 3D modeled data relative to the linear 1D model.

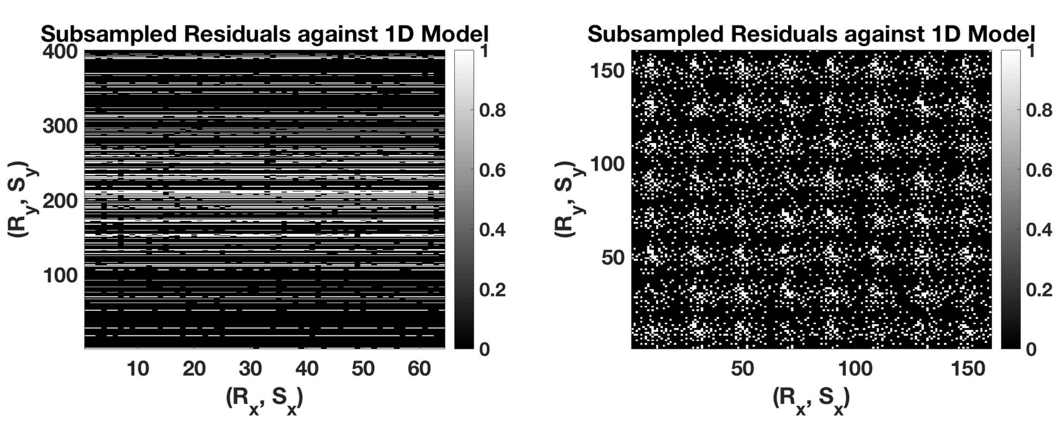


(a) Spatial locations of the sources ('*') around Mt St Helens (triangle). The synthetic receiver grid to the right of this figure is taken from the source with the red '*', seen just to the left of the mountain.

(b) Example receiver grid for a single source. This shows the residual difference between the 1D and 3D models. Note that this source has the most data points out of all sources.

Figure 1: Source and receiver grid positions.

- Each source i at (S_{x_i}, S_{y_i}) has an associated receiver grid of observations $(R_x, R_y) \in (70,165)\times(70,165)km^2$.
  - Data is recorded in 4D tensor format with dimension (R_x, R_y, S_x, S_y), where $R_x = 20, R_y = 20, S_x = 8, S_y = 8$
  - Two possible tensor formulations: Figures 2(a) and 2(b).
  - The second exhibits global low-rank structure.
- Sources are organized from high-energy to low-energy



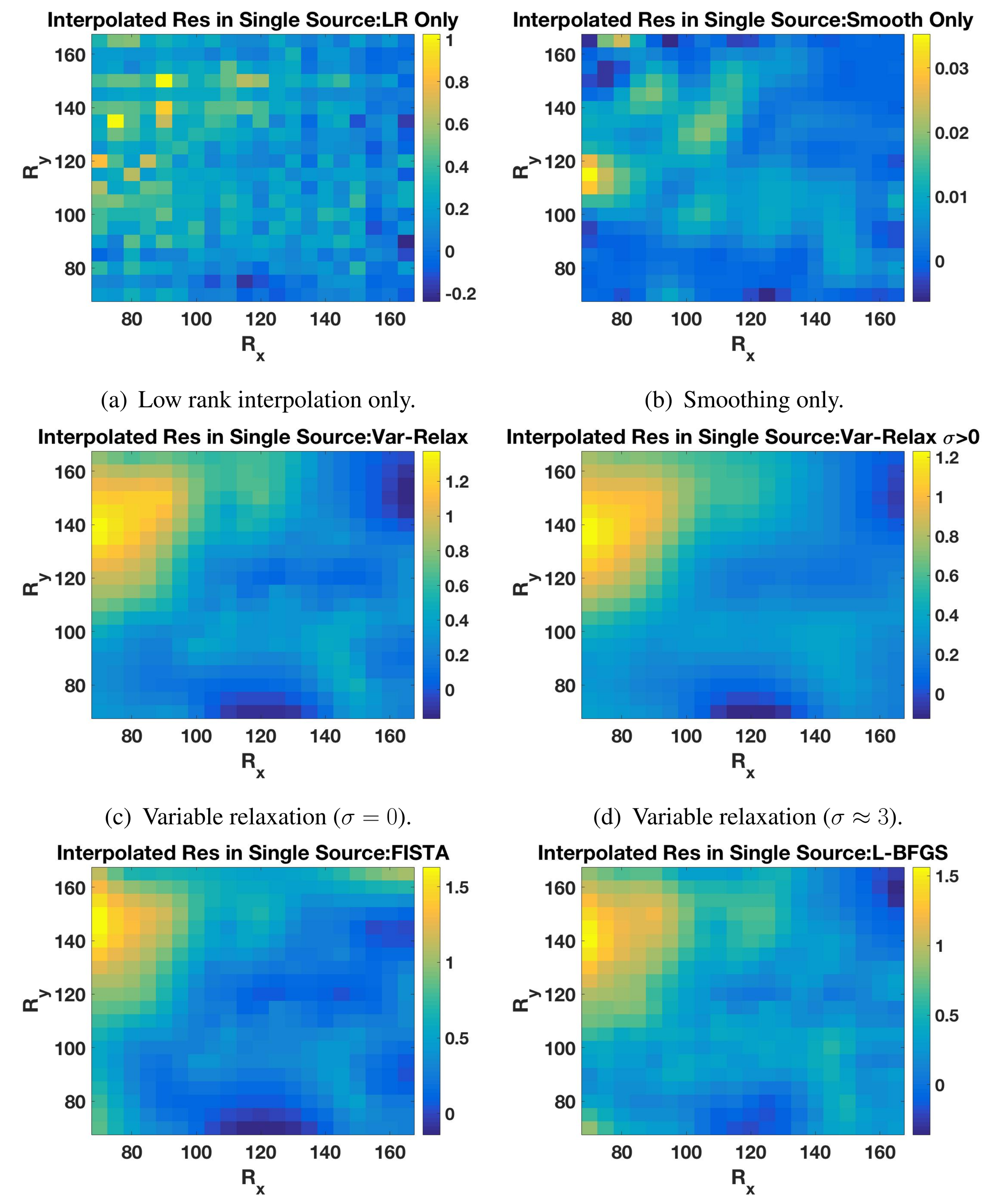
(a) Binary subsampled matricization $(R_x, R_y)\times(S_xS_y)$; the missing data (zeros) go across the rows: $X\in\mathbb{R}^{400\times 64}$.

(b) Binary subsampled matricization $(R_x\times R_y)\times(S_xS_y)$; missing entries (zeros) are are interwoven throughout the matrix: $X\in\mathbb{R}^{160\times 160}$.

Figure 2: Two possible low-rank formulations.

Results

- Subsampling rate is 15%, with $k = 40$ (for the LR^T formulation).
- The convergence criteria is set to 1e-10.
- We test the relaxation formulation against FISTA and L-BFGS for both $\sigma = 0$ and $\sigma = \sqrt{\sum_{i=1}^{\mathcal{T}_{obs}} 0.5^2} \approx 3$.
- Results, with low-rank and smoothing alone, are presented in Figure 3 (for single sources only) and in Table 1.

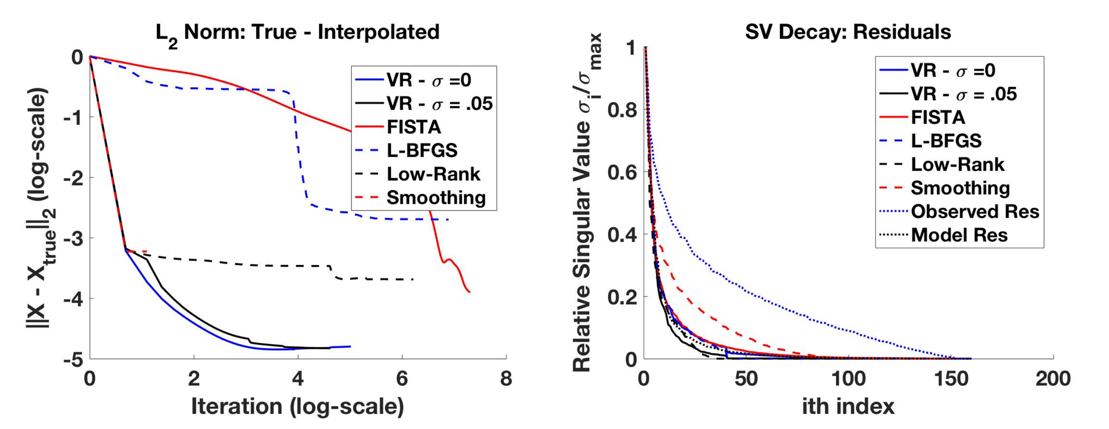


(a) Low rank interpolation only.
(b) Smoothing only.
(c) Variable relaxation ($\sigma = 0$).
(d) Variable relaxation ($\sigma \approx 3$).
(e) FISTA.
(f) L-BFGS.

Figure 3: Residual results for the different algorithms on only one source.

Alg	Time (s)	RMS (obs)	RMS (int)
Combined - VP ($\sigma = 0$)	**23.16**	**0.09**	**0.109**
Combined - VP ($\bar{\sigma} = .05$)	**18.11**	**0.06**	**0.102**
FISTA	21.02	0.09	0.120
L-BFGS	172.28	0.09	0.139
Smooth only	5.44	0.24	0.266
Low-rank only	2.31	0.08	0.213

Table 1: Different formulations for the same model residual dataset with 0% cross-validation. Here we have the combined formulation interpolating on model residuals only. $\gamma = 7.7e5,\ \eta = 1.0,\ \% = 9.38$. For l_2, $\gamma = 0,\ \eta = .5$.



(a) Objective value decay
(b) SVD decay for interpolated matrices.

Figure 4: Convergence information for different algorithms.

Conclusions & Future Directions

Our new variable relaxation algorithm can both solve

- Data misfit constraint with $\sigma > 0$
- Combine local (smoothness) & global (low-rank) information
- Converge quickly to local minima

Future directions are to interpolate with actual data as well as provide a metric for estimating uncertainties, with the goal of providing new data-stations to the PDE-constrained optimization routine.

References

[1] Ulberg C W, Creager K, Moran S C, Abers G A, Crosbie K, Crosson R S, Denlinger R P, Thelen W A, Kiser E, Levander A and Bachmann O 2017 Imaging Seismic Zones and Magma beneath Mount St. Helens with the iMUSH Broadband Array

[2] Oneto L, Ridella S and Anguita D 2016 *Machine Learning* **103** 103–136

[3] Kumar R, Lpez O, Davis D, Aravkin A Y and Herrmann F J 2017 *IEEE Transactions on Computational Imaging* **3** 264–274

[4] Recht B, Fazel M and Parrilo P A 2010 *SIAM Rev.* **52** 471–501 ISSN 0036-1445 URL http://dx.doi.org/10.1137/070697835

[5] Zheng P and Aravkin A 2018 *arXiv preprint arXiv:1802.02654*

[6] Nocedal J and Wright S J 2000 *Numerical Optimization* (Springer)

[7] Beck A and Teboulle M 2009 *SIAM Journal on Imaging Sciences* **2** 183–202 (*Preprint* https://doi.org/10.1137/080716542) URL https://doi.org/10.1137/080716542