A practical workflow for land seismic wavefield recovery with weighted matrix factorization

Yijun Zhang and Felix J. Herrmann

SEG 2021
Motivation

Fully sample data is a prerequisite to
- multiple removal, migration & FWI

Seismic data is collected randomly along the spatial coordinate to
- shorten the acquisition time, reduces cost
- shift the burden from field acquisition to data processing

Matrix completion for low-to-mid frequencies
- exploits low-rank structure for recovery, simple & computationally cheap

Weighted matrix completion for higher frequencies
- gives good results for higher frequency slices on marine data
- application to land data is hampered by ground roll

Blind study on 3D SEAM Barrett dataset

**Data dimension:** $80 \times 21 \times 641 \times 641 \times 667$

$$\left(n_{sx} \times n_{sy} \times n_{rx} \times n_{ry} \times n_t\right)$$

**Receiver sampling interval:** 12.5 m

**Source sampling interval:** 25 m in the shot line direction and 100 m in the perpendicular direction

**Time sampling interval:** 0.006 s

**Subset of dataset**
- benchmark for land data
- contains realistic surface waves

https://www.researchgate.net/publication/
Acquisition geometry w/ ~ 75% missing receivers, 21 source lines

A subset consists of 21 source lines (red lines in the center area)

Each source line contains 80 sources

The $8 \times 8$ km receiver aperture moves with the source location

- neighboring shots share most randomly sampled receivers (black dots in the figure)
- some drop-off and others add (red and blue rectangles in the figure)
Impact of ground roll

Ground roll corresponds to Rayleigh-type surface waves
  - slow & aliased
  - strong amplitude

Dominate the reconstruction at the expense of weaker body waves.

Observed data in time domain (one shot) w/ ~ 75% missing receivers
Reconstructed data in time domain (one shot)
via weighted matrix completion w/o proposed workflow
Main research questions

How can we use weighted matrix factorization on land data while avoiding the impact of ground roll?

Answer:

› *Reconstruct the body and surface (ground roll) waves separately.*

Why separation?

Answer:

› *Reduce the effect of strong aliased ground roll on body wave reconstruction.*
› *The ground roll can be separated from body waves, at least in an approximate sense.*
Wavefield reconstruction via matrix completion

Successful reconstruction schemes

- exploit structure: Low rank/fast decay of singular values in “transform domain”
- sample randomly: Increase rank in “transform domain”
- optimization via rank-minimization (matrix factorization)

Weighted matrix completion

- further improves the wavefield recovery at higher frequencies by introducing matrix weights
Weighted matrix completion

Variational formulation:

\[
\min_X \frac{\|QXW\|_*}{\|QXW\|_*} \quad \text{subject to} \quad \| \mathcal{A}(X) - B \|_F \leq \epsilon
\]

where

\[
Q = w_1 U U^H + U^\perp U^\perp H, \quad W = w_2 V V^H + V^\perp V^\perp H, \quad B, X \in \mathbb{C}^{(N_s \times N_{sx}) \times (N_r \times N_{ry})}.
\]

- \( U, V \) column and row subspaces (prior information) derived from neighboring frequencies
- \( w_1, w_2 \in (0, 1] \) similarity of the prior information and to-be-recovered data
- Small values for scalars indicate more confidence in the prior information

Expensive for large scale
Weighted matrix factorization

Weighted matrix factorization:

\[
\underset{L,R}{\text{minimize}} \quad \frac{1}{2} \left\| \begin{bmatrix} QL \\ WR \end{bmatrix} \right\|_F^2 \quad \text{subject to} \quad \left\| \mathcal{A}(LR^H) - B \right\|_F \leq \epsilon
\]

where

\[
X = LR^H, \quad L \in \mathbb{C}^{(N_s \times N_r) \times r}, \quad R \in \mathbb{C}^{(N_y \times N_r) \times r}.
\]

Incorporating weighting matrices into the objective function

- complicates optimization
- increases the computational cost of minimization
Variational formulation:

\[
\begin{align*}
\text{minimize} & \quad \| \tilde{X} \|_\ast \\
\text{subject to} & \quad \| \mathcal{A}(Q^{-1}\tilde{X}W^{-1}) - B \|_F \leq \epsilon
\end{align*}
\]

where

\[
\tilde{X} = QXW, \quad Q^{-1} = \frac{1}{w_1} \begin{bmatrix} U \end{bmatrix}^H + \begin{bmatrix} U \perp \end{bmatrix}^H, \quad W^{-1} = \frac{1}{w_2} \begin{bmatrix} V \end{bmatrix}^H + \begin{bmatrix} V \perp \end{bmatrix}^H.
\]

Weighted matrix factorization:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \left\| \begin{bmatrix} \tilde{L} \\ \tilde{R} \end{bmatrix} \right\|_F^2 \\
\text{subject to} & \quad \| \mathcal{A}(Q^{-1}\tilde{L}\tilde{R}^HW^{-1}) - B \|_F \leq \epsilon
\end{align*}
\]

where

\[
\tilde{X} = \tilde{L}\tilde{R}^H
\]

The original solution can be recovered by \( X = Q^{-1}\tilde{X}W^{-1} \).
Proposed workflow

1. Observed data
   - Linear shift

2. Observed data with linear shift
   - Apply a taper

3. Observed ground roll estimation with linear shift
   - Remove the linear shift

4. Observed ground roll estimation
   - Apply parallel weighted matrix factorization
   - Apply steps 1, 2, and 3 on recovery

5. With only observed positions

6. Observed body waves without ground roll estimation
   - Apply parallel weighted matrix factorization

7. Recovered body waves
   - Denoising

Body wave recovery

Recovered ground roll estimation

Output reconstruction
Ground roll estimation

A. Observed data
B. The output after aligning the ground roll
  ‣ by using a linear shift
C. The output after applying a smooth taper
D. The ground roll estimation
  ‣ by undoing the linear shift
Ground roll recovery

Observed ground roll  
Recovered ground roll
Body wave recovery

Observed body wave

Recovered body wave
Final reconstructed result in time domain
Conclusions

We mitigate the effects of strongly aliased ground roll by employing the proposed separation.

Furthermore, the proposed workflow successfully recovers body waves (reflections and diffractions).
Acknowledgement

This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center.

We would like to express our gratitude to Georgia Institute of Technology for funding this study.

We would also like to thank Occidental for providing the dataset, as well as Klaas Koster for assisting us with the blind test.

Thank you for your attention!!