Three-dimensional Seismic Imaging & Sparse Inversion

Deborah Pereg

Supervised by Prof. Israel Cohen

Technion, Electrical Engineering Faculty

April 2021

Table of Contents

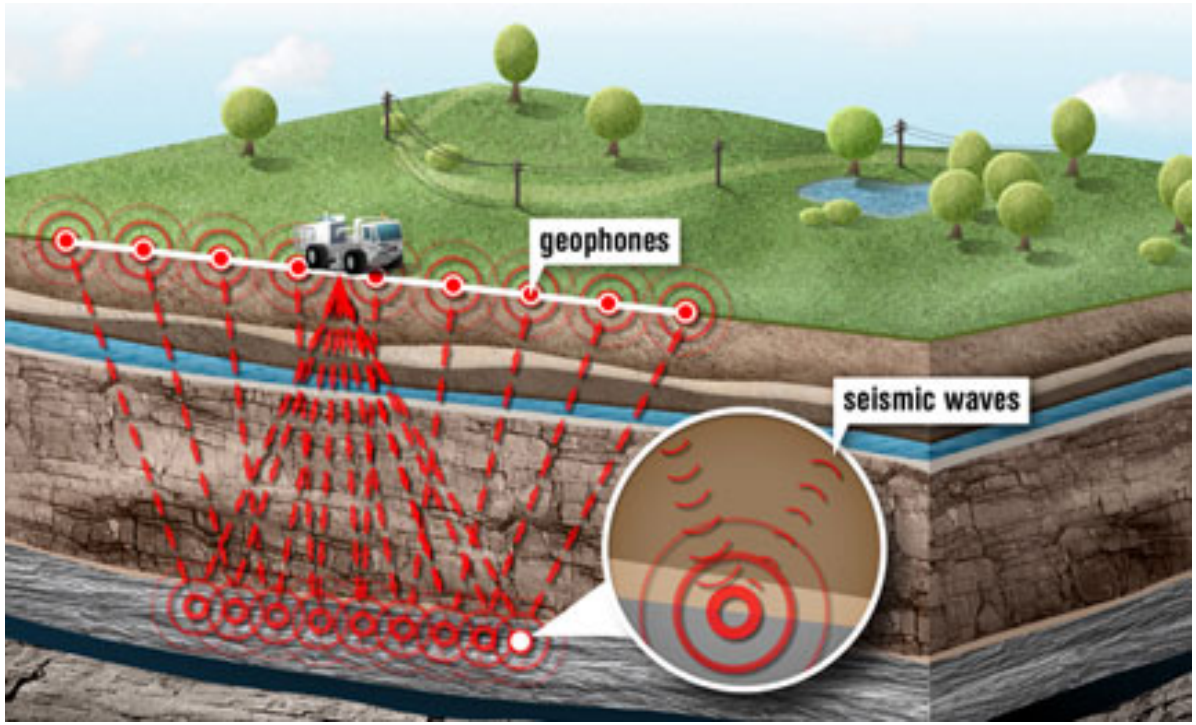
1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

See Into Material

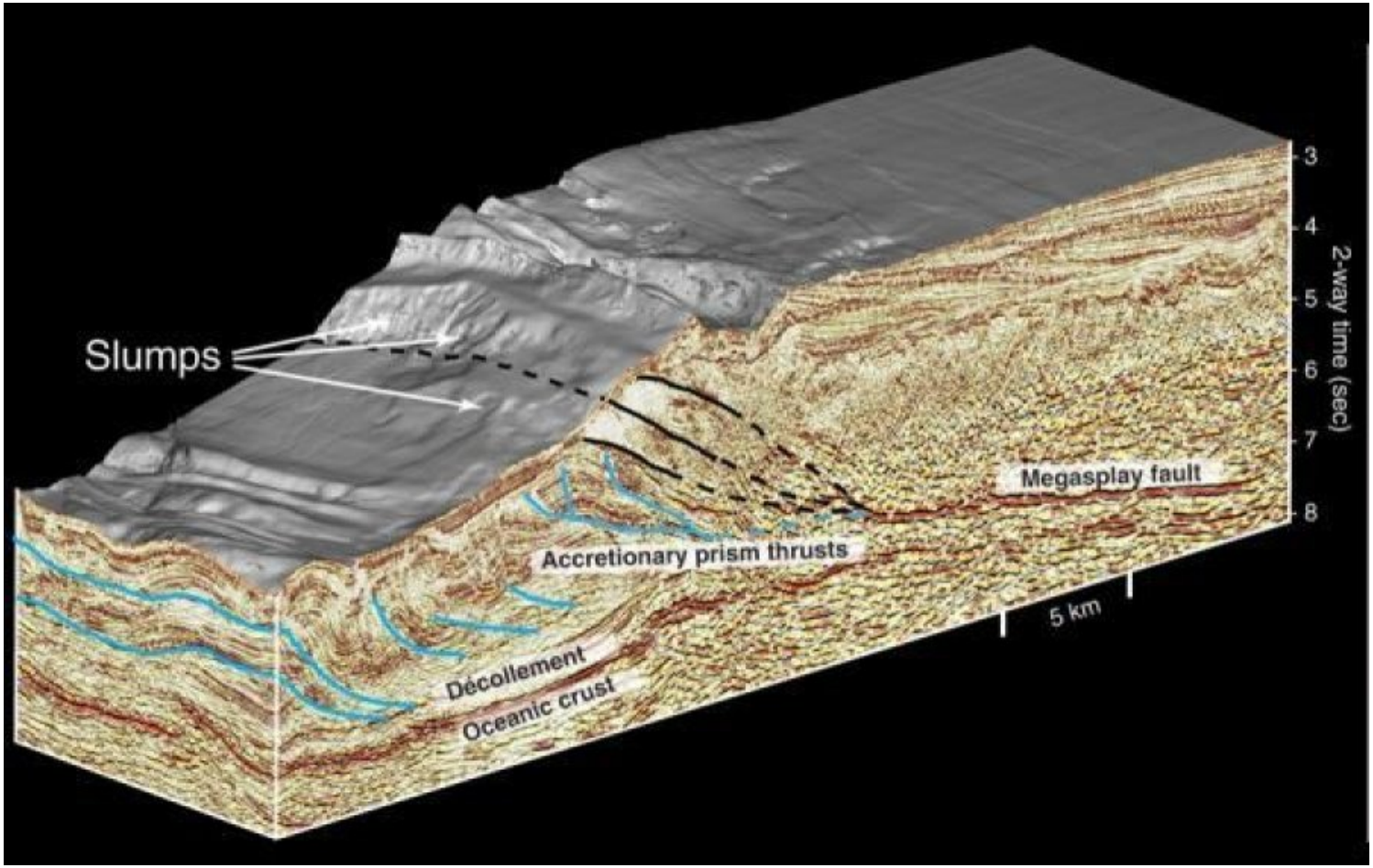
Develop simpler and efficient ways of processing of large amounts of (seismic) data, in pursuit of optimal methods for

- 3D imaging
- Super resolution
- Noise reduction

Exploration Seismology 3D imaging



Exploration Seismology 3D imaging



Three-Dimensional Splay Fault Geometry and Implications for Tsunami Generation.

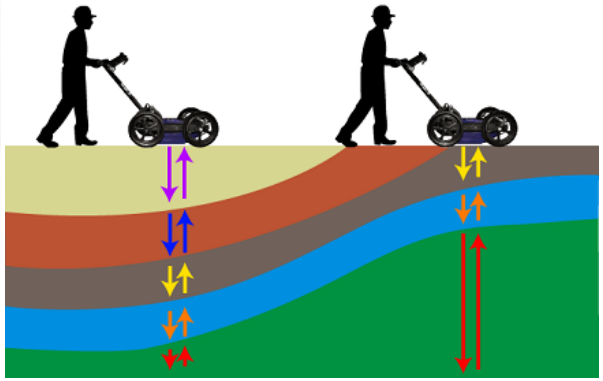
Exploration Seismology

Why?

- Mineral deposits
- Archaeological sites
- Geological information for engineering
- Geothermal energy surveys
- Risk assessment (tsunamis!)
- Underground activity detection (tunnels)

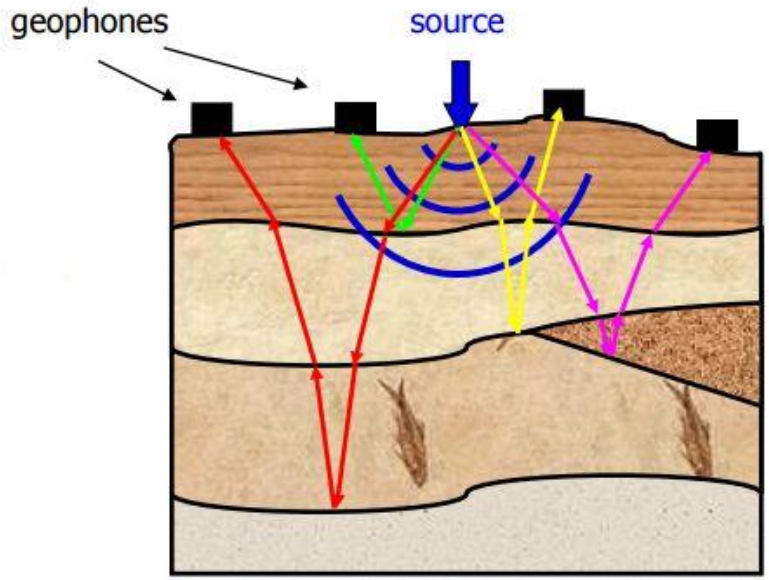
Related Applications

- Ultrasound Imaging
- GPR (Ground Penetrating Radar)
- Monitoring sensors



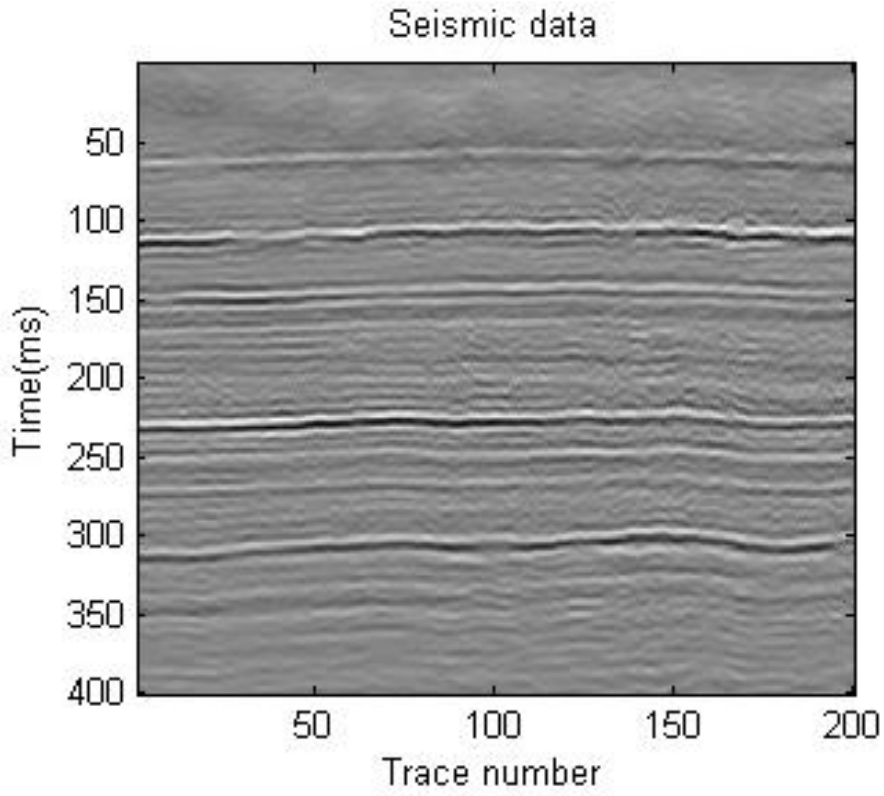
Exploration Seismology

- Acoustic pulse transmitted from the earth surface.
- Reflected pulses received by array of geophones.
- Reflections generated at layer boundaries.
- Depth of up to 10 km.



Exploration Seismology

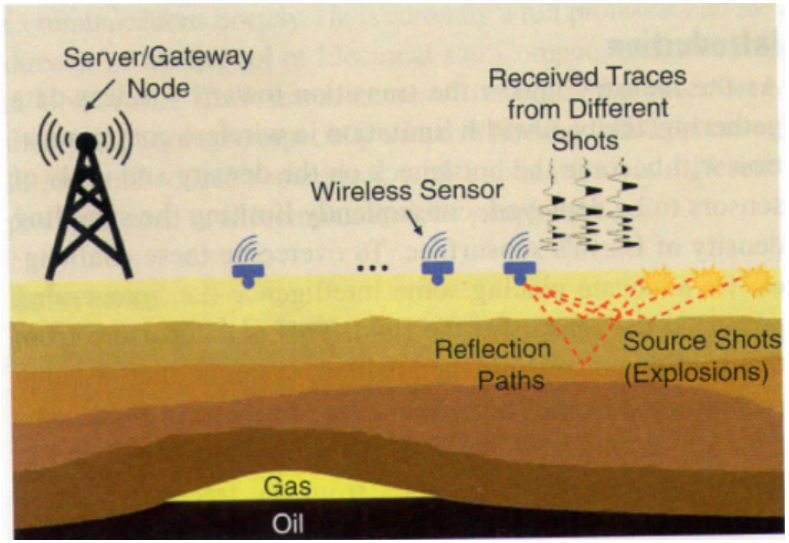
- Recorded data processed into seismic 3D data volume.
- 3 main stages:
  - Stacking *
  - Migration
  - Deconvolution

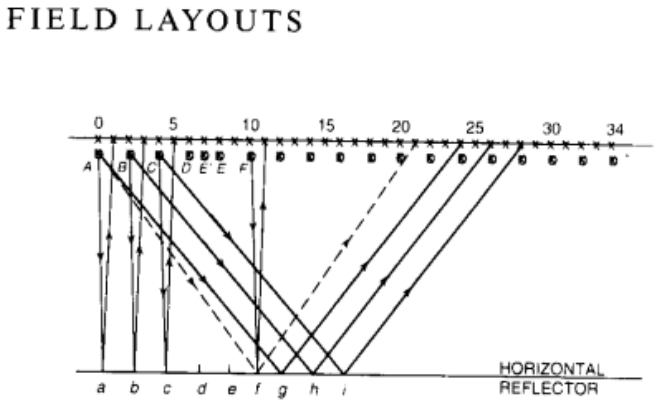


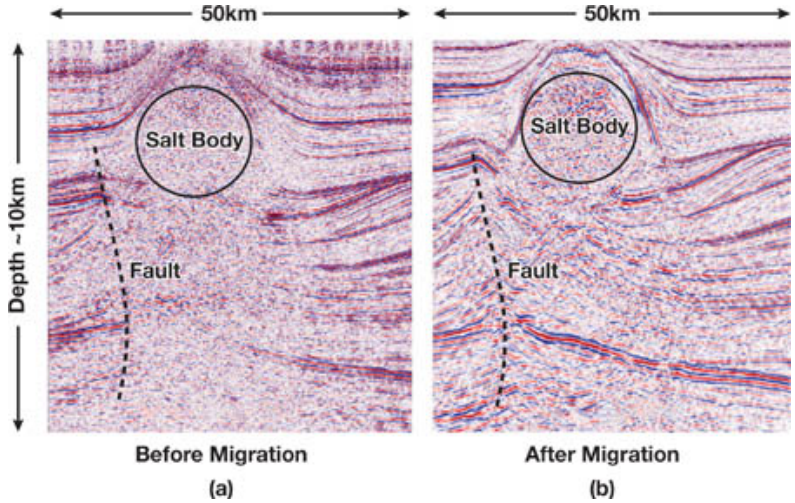
*Seismic data is courtesy of GeoEnergy Inc., Texas

Exploration Seismology

- Increase SNR
- Compensate for attenuation
- Relocate reflectors to their true location
- Focus the data



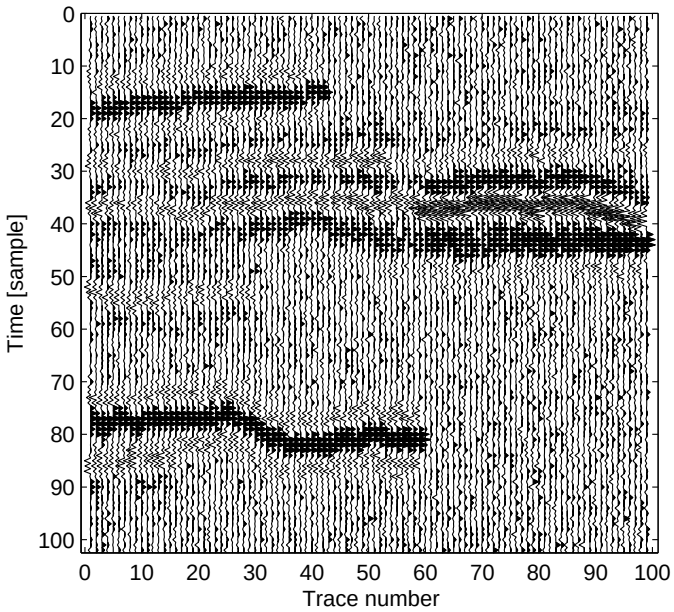




Seismic Deconvolution

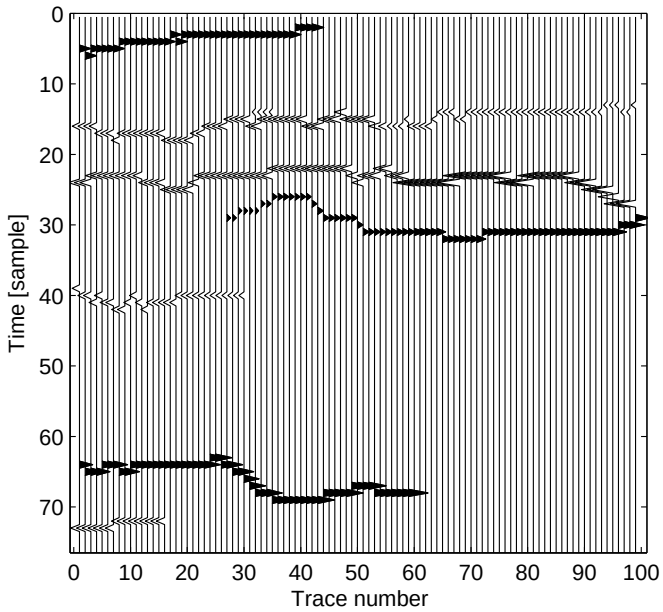
Final goal: Reveal ground layers structure.

Synthetic seismic data



$\Longrightarrow$

Reflectivity



Super Resolution Point of View

Aim to recover the fine-scale details, or equivalently, the high frequencies of the signal, even though they have been mathematically killed by the measurement process.

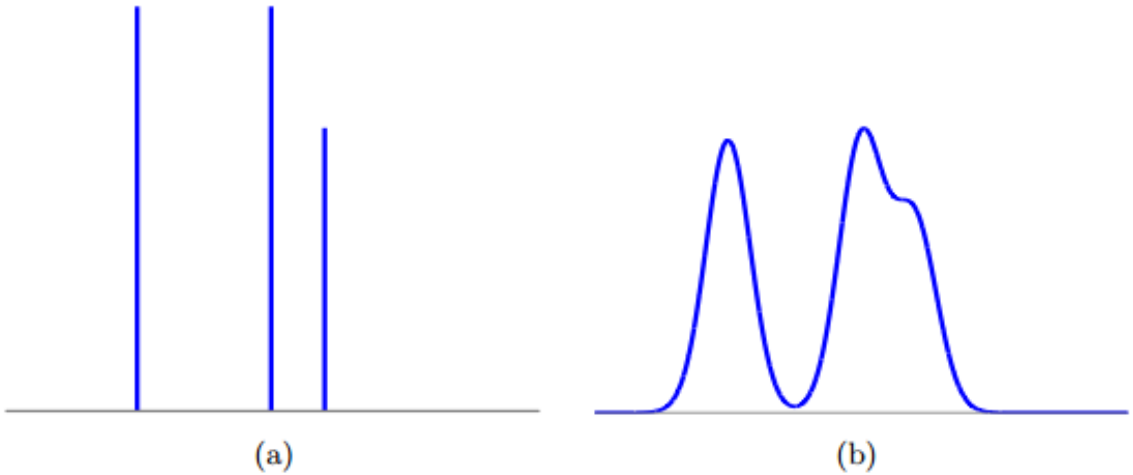
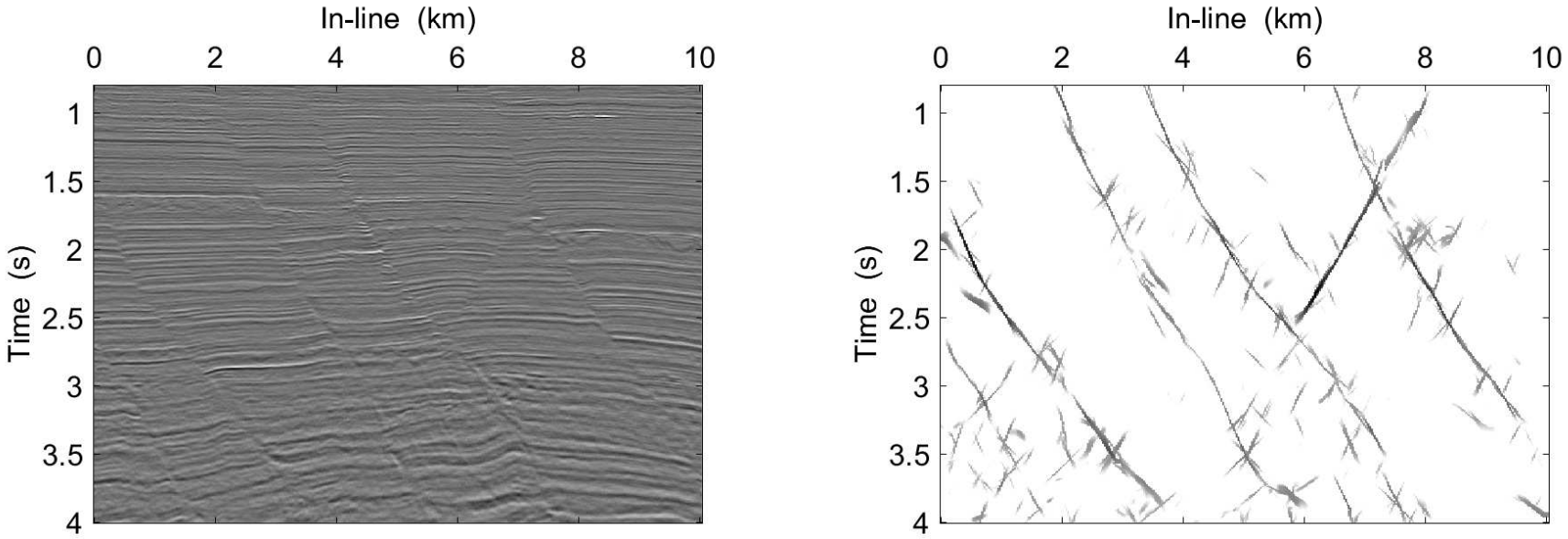


Illustration: (a) Highly resolved signal. (b) Low pass filtered signal.

Some Real Data Examples...

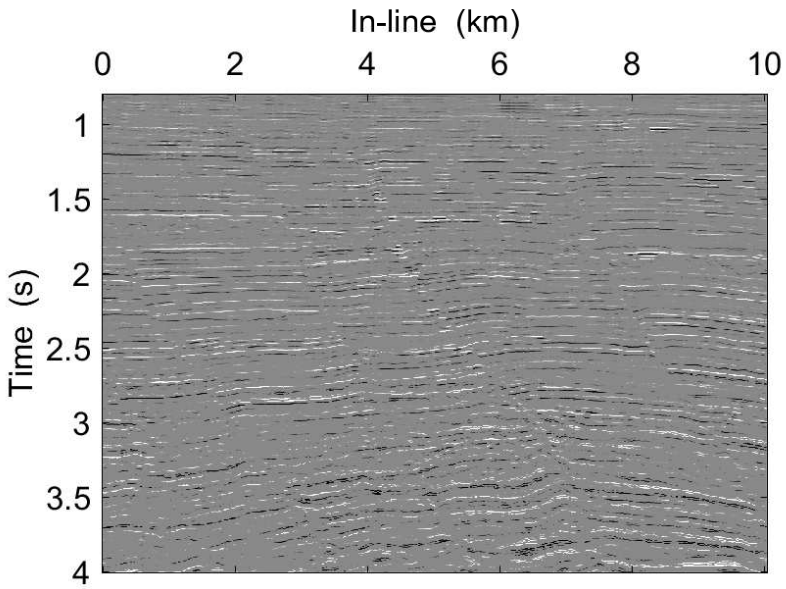


(a) (b)

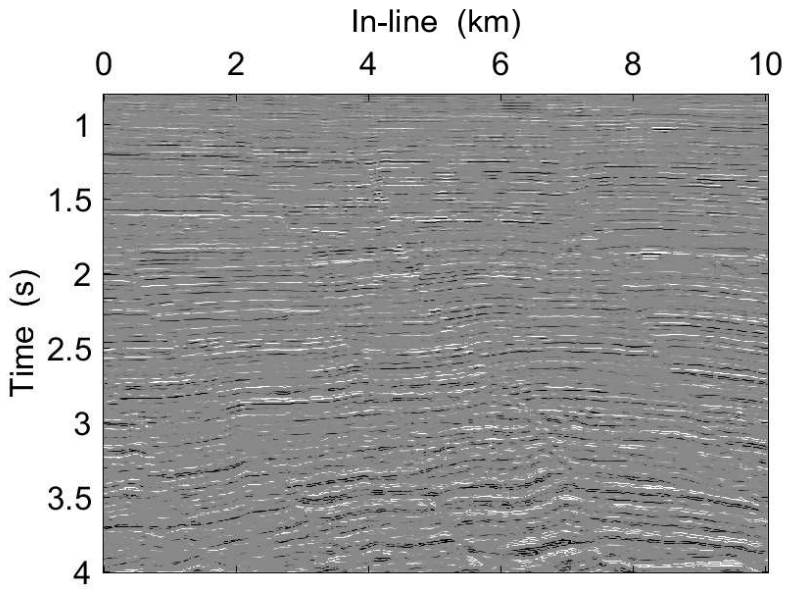
Vertical cross-section at $y = 1.3$km through:

(a) seismic data (b) skeletonized LFE (Local Fault Extraction) volume.

Real Data Inversion Results



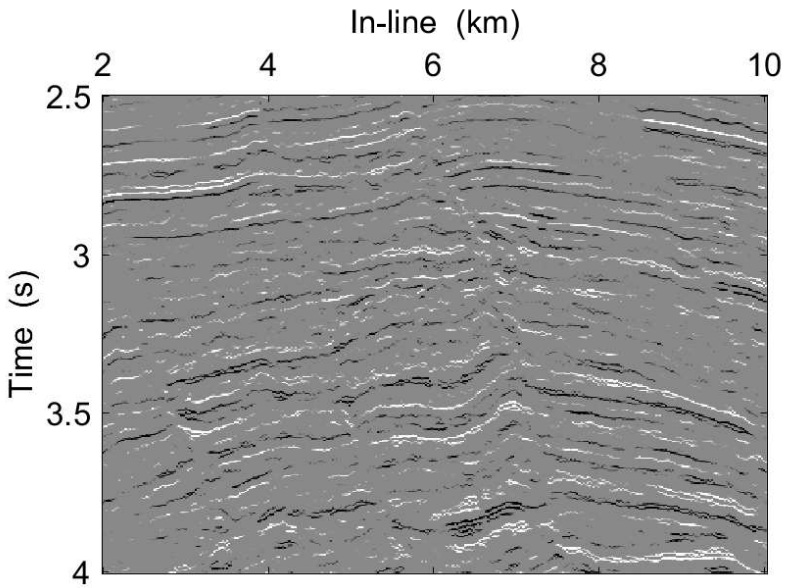
(a) single-channel



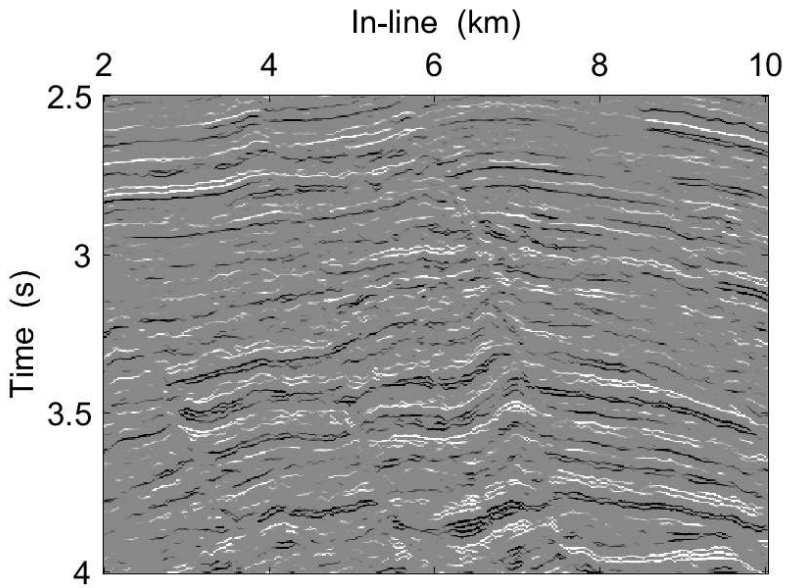
(b) multichannel

Real Data Inversion Results

Estimated reflectivity - zoom-in.



(a) single-channel



(b) multichannel

Research Overview

- Three-dimensional Sparse Seismic Deconvolution Based on Earth Q Model
- Sparse Seismic Deconvolution via Recurrent Neural Networks
- Prestack Time Migration Velocity Analysis using Recurrent Neural Networks
- Convolutional Sparse Coding Fast Approximation with Application to Seismic Reflectivity Estimation

Research Challenges

1. Develop mathematical model or hypothesis
2. Solve and test
3. Provide analytical guarantees
4. Real data inherent uncertainty
5. "**Big**" data
6. Few-shot learning issues

Research Guidelines

- Allow Error, Gain Speed

  Tradeoff between speed and accuracy

  When time is expensive, it might be worthwhile to allow inexact approximations.

- Global models suffer from the curse of dimensionality

  Attempt to propose methods specifically tailored to local invariant structure

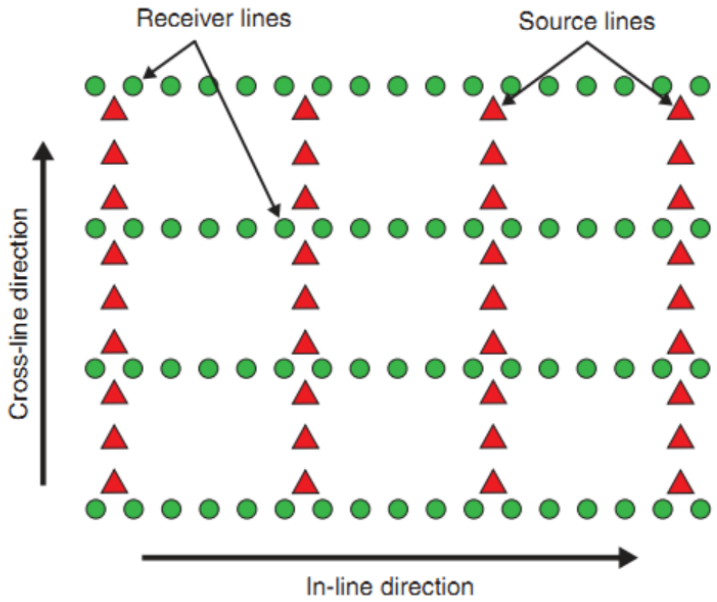
Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

Seismic Data Acquisition

Goal: Imaging of internal structure of the earth.

- Acoustic pulse transmitted into the ground.
- Reflected energy received by geophones and processed into a 3D image.



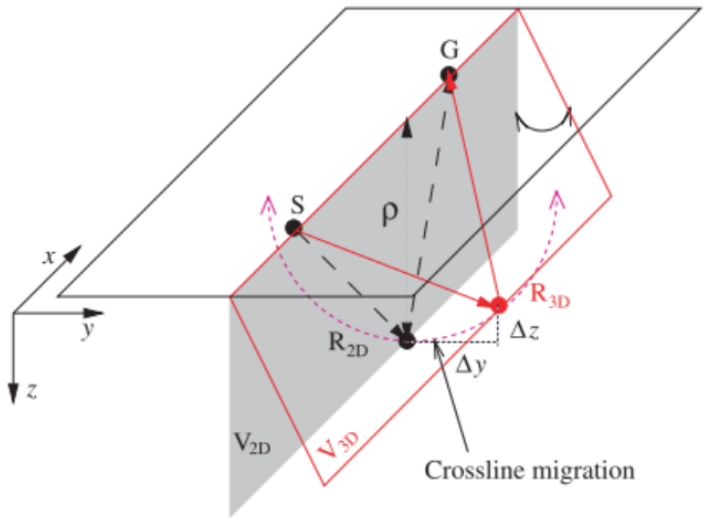
3D Data Geometries

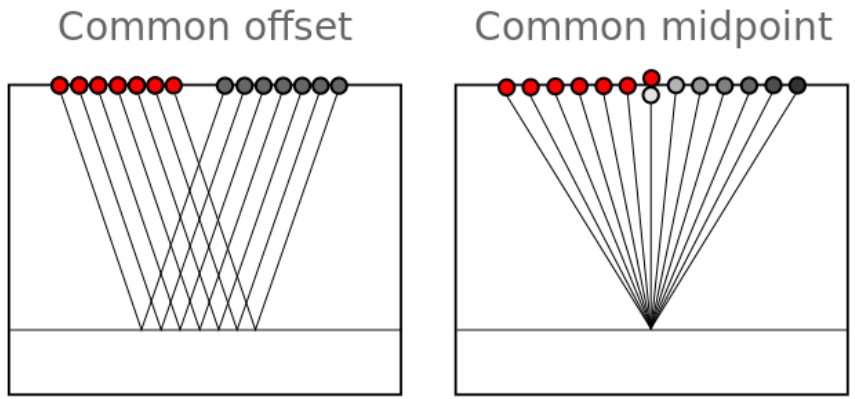
Field coordinates: Source location - $\mathbf{s} = (x_s, y_s)$.

Receiver location - $\mathbf{g} = (x_g, y_g)$.

Midpoint-offset coordinates $(\mathbf{m}, \mathbf{h})$

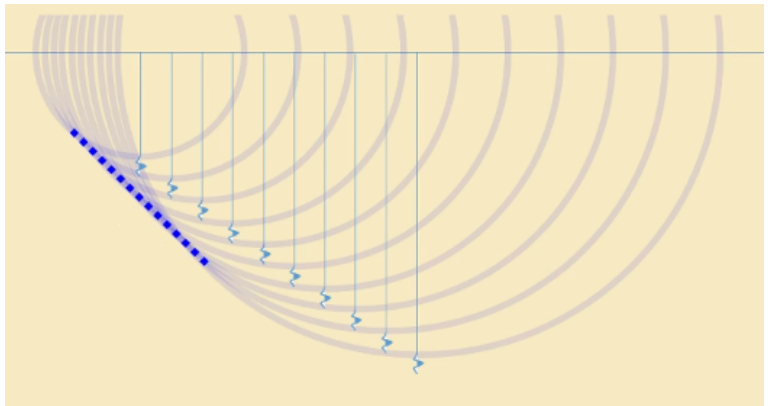
$$\left(\mathbf{m} = \frac{\mathbf{g} + \mathbf{s}}{2}, \mathbf{h} = \frac{\mathbf{g} - \mathbf{s}}{2}\right).$$





Migration

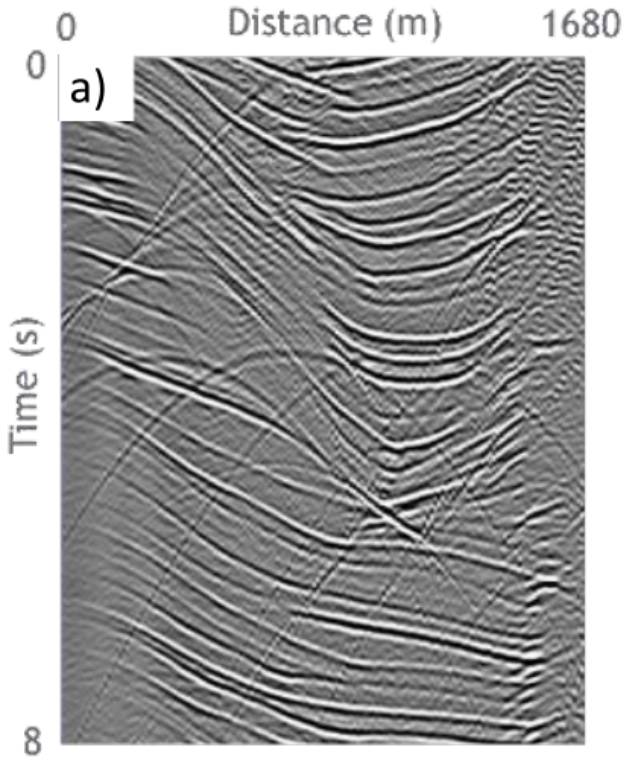
- Relocate reflectors to their true location.
- Correct position of reflectors and focus the data.
- Find true reflecting surfaces, given vertically plotted points.



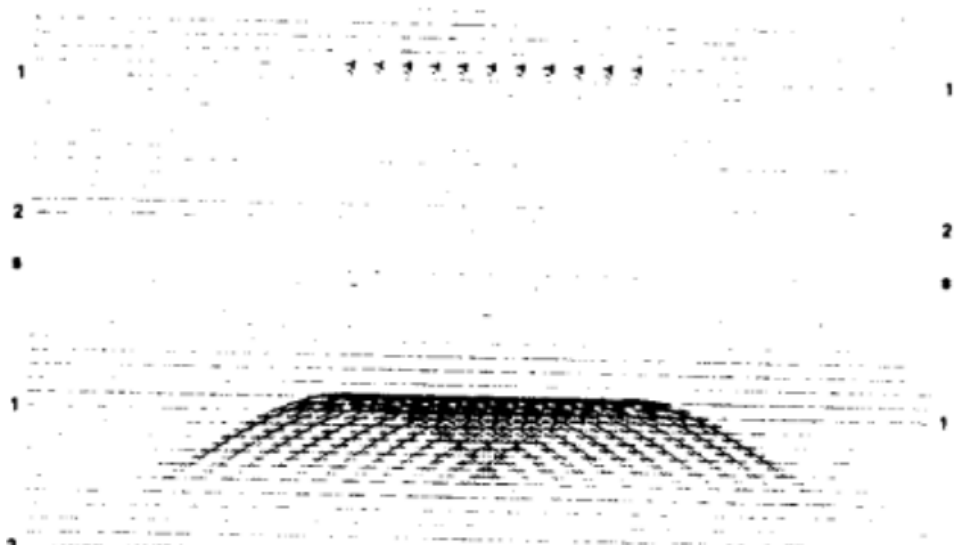
(a) (b)

Illustration of ideal seismic experiment: (a) coincide source-receiver ; (b) "spraying" energy of each trace, semi-circles constructively interfere to plot reflector in correct position.

Migration reflections as diffractions

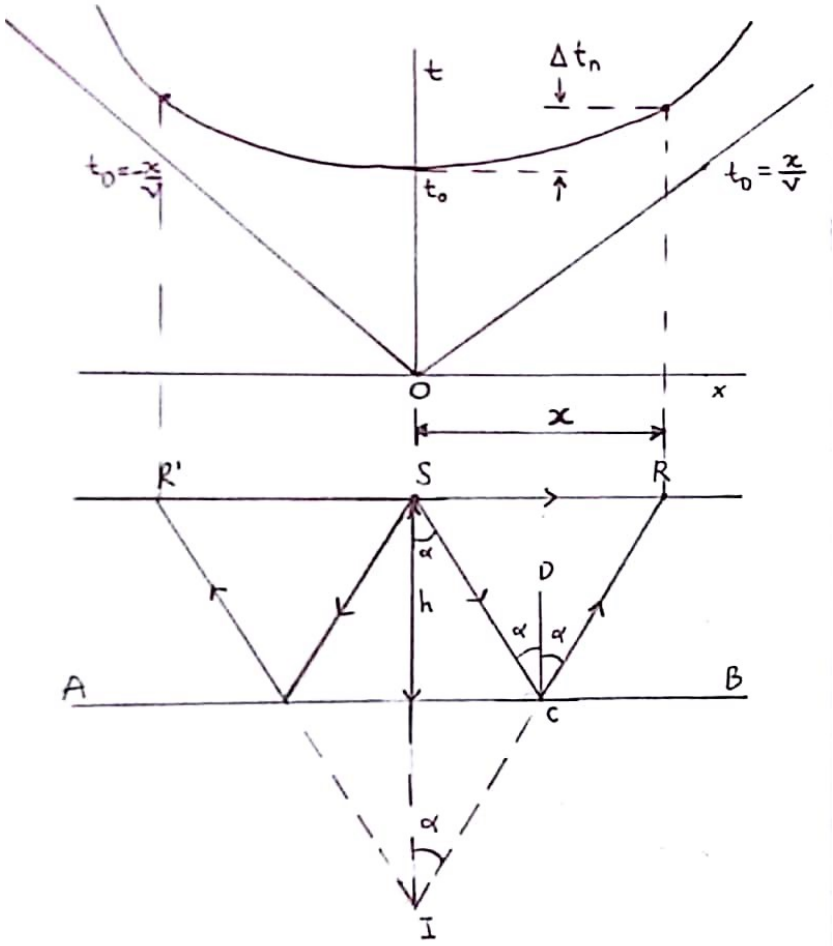


(a) Synthetic data



(b) Diffracting points.

Traveltime curve for a horizontal reflector



The traveltime curve is a hyperbola

A reflecting bed AB, at depth h, below source S and a geophone R.

$$V^2 t^2 = x^2 + 4h^2$$

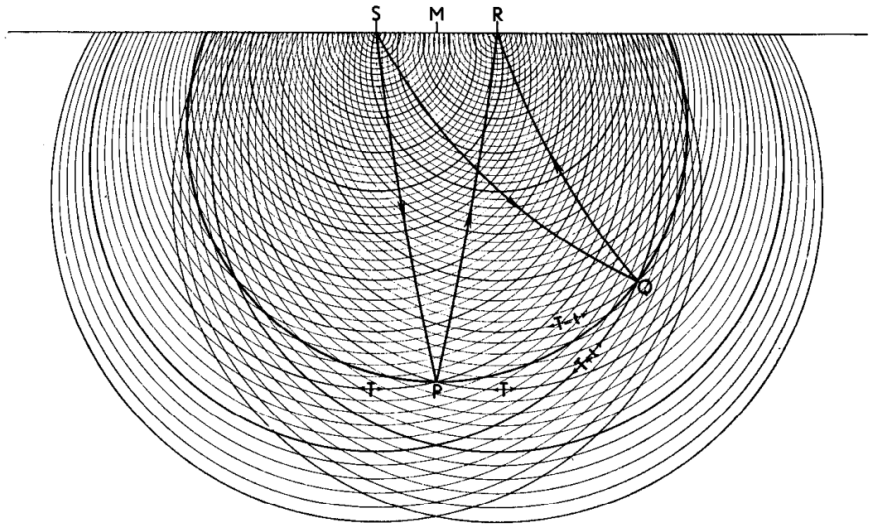
$$\frac{V^2 t^2}{4h^2} - \frac{x^2}{4h^2} = 1$$

Kirchhoff Migration

Start with the scalar wave equation

$$\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2}\right) = \frac{1}{v^2(x,y,z)} \frac{\partial^2 P}{\partial t^2}$$

$P(x, y, z; t)$ is the pressure wavefield propagating in a medium with velocity $v(x, y, z)$.

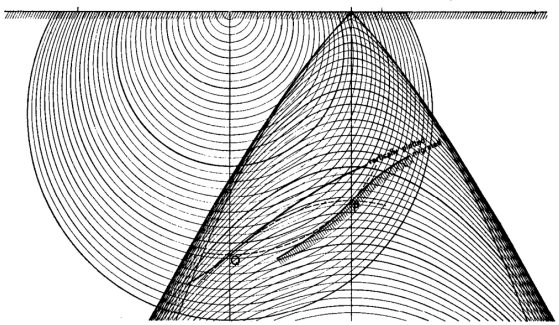


Kirchhoff Migration

$$\nabla^2 P = \frac{1}{v^2(x,y,z)} \frac{\partial^2 P}{\partial t^2}$$

Homogenous wave equation, inhomogeneous boundary conditions. No real sources in the subsurface image, only reflectors and scatters.

- $P(x_0, y_0, z = 0; t_0)$ - the observed seismic data.
- $U(x, y, z) = P(x, y, z; t = 0)$ - the migrated image.



Kirchhoff Migration

The discrete form of Kirchhoff migration is given by

$$U(\mathbf{r}) \approx \sum_{i \in \Omega_r} W(\mathbf{r}, \mathbf{m}_i, \mathbf{h}_i) D(t = t_D(\mathbf{r}, \mathbf{m}_i, \mathbf{h}_i), \mathbf{m}_i, \mathbf{h}_i)$$

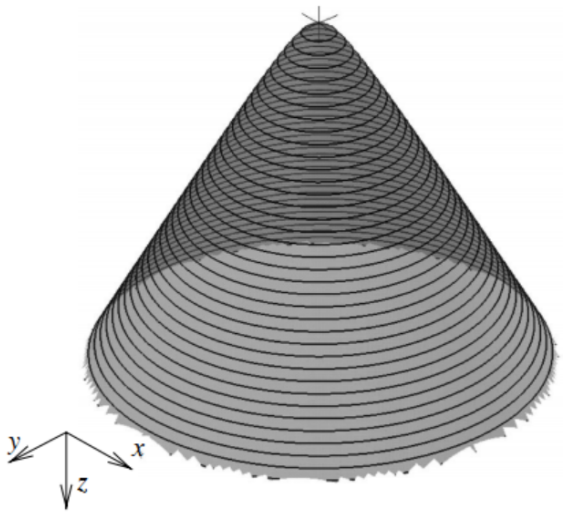
- $W(\mathbf{r}, \mathbf{m}_i, \mathbf{h}_i)$ - known weighting factors. $\mathbf{r} = (x, y, z)$
- The summation domain Ω_r - a pre-determined aperture.
- $t_D(\mathbf{r}, \mathbf{m}_i, \mathbf{h}_i)$ - total time delay of reflections propagation from source to image point and back to the surface, to be recorded at the receiver position.

Summation Surfaces - Gathering Methods

Assuming constant velocity, the time delays define a family of summation surfaces

$$t_D = t_s + t_g = \frac{\sqrt{z^2 + |(x,y) - \mathbf{s}|^2}}{V} + \frac{\sqrt{z^2 + |(x,y) - \mathbf{g}|^2}}{V}.$$

$$t_D = \frac{\sqrt{z^2 + |(x,y) - \mathbf{m} + \mathbf{h}|^2}}{V} + \frac{\sqrt{z^2 + |(x,y) - \mathbf{m} - \mathbf{h}|^2}}{V}.$$



Zero-offset summation surface defined in the data space.

Non constant velocity

- * When velocity is changing slowly in the horizontal direction, time-delay functions can be approximated by

$$t_D = \frac{\sqrt{z^2 + |(x,y) - \mathbf{m} + \mathbf{h}|^2}}{V_{\text{rms}}(x,y,z)} + \frac{\sqrt{z^2 + |(x,y) - \mathbf{m} - \mathbf{h}|^2}}{V_{\text{rms}}(x,y,z)}.$$

$V_{\text{rms}}(x,y,z)$ is the average velocity to the point (x,y,z).

- **

Our Goal

Migration Velocity Analysis

Find $V_{\mathrm{rms}}(x, y, z)$

in order to perform imaging, by computing Kirchhoff migration

$$U(\mathbf{r}) \approx \sum_{i \in \Omega_r} W(\mathbf{r}, \mathbf{m}_i, \mathbf{h}_i) D(t = t_D(\mathbf{r}, \mathbf{m}_i, \mathbf{h}_i), \mathbf{m}_i, \mathbf{h}_i)$$

Ill posed inverse problem

Migration velocity Analysis

Do data hold all necessary information

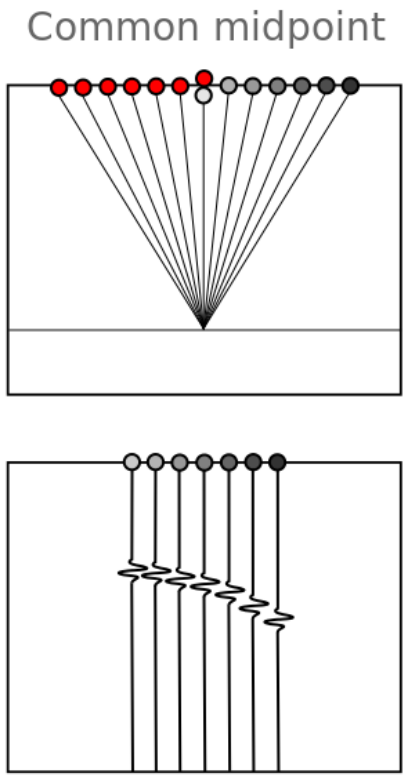
to compute a velocity function that varies with

depth and along the horizontal direction

???

Principles of Velocity Analysis

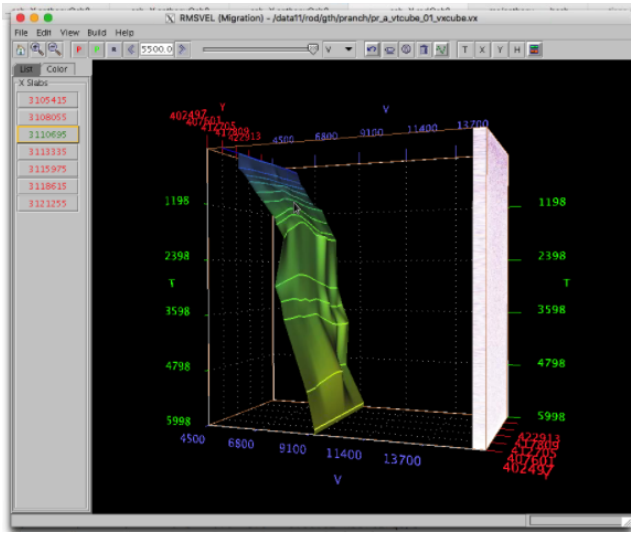
- The simplest method is measure the moveout along the offset axis.
- Flatness of corrected gathers is measured by a coherency measure (such as semblance).
- Velocities corresponding to maxima are picked, either by hand or automatically.



Principles of Velocity Analysis

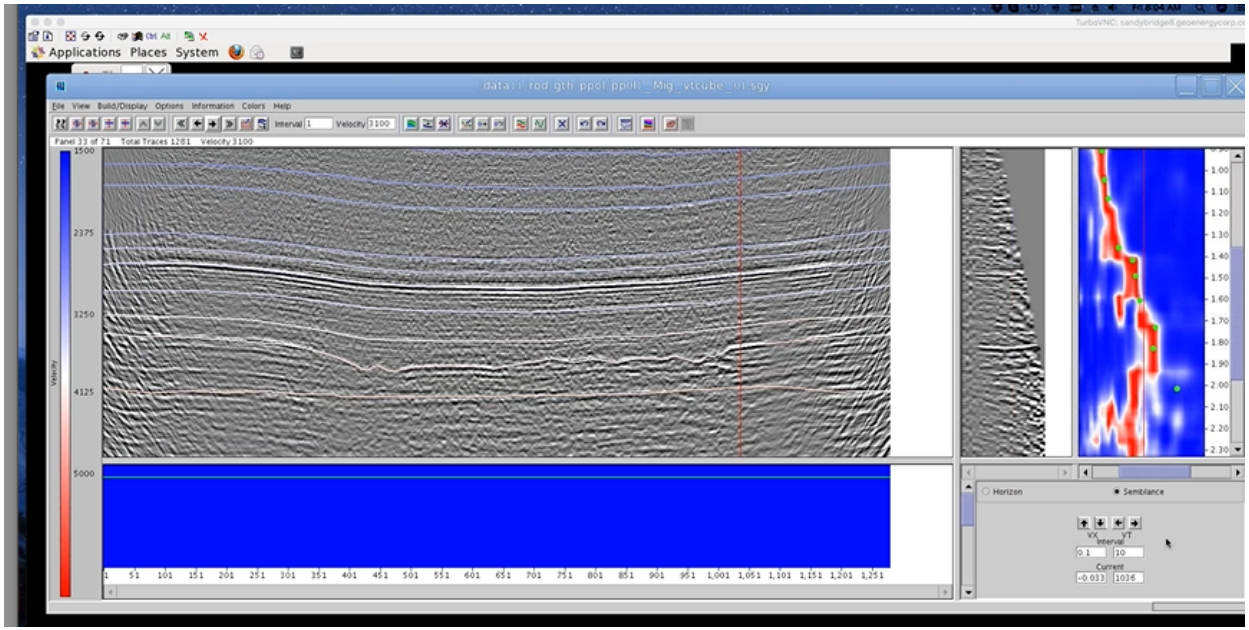
Start with unmigrated data.

- Generate Constant Velocity Migration (CVM) panels and extract time slices from the CVMs.
- Form a super image - a volume of migrated data of midpoint, time, and migration velocity axes.



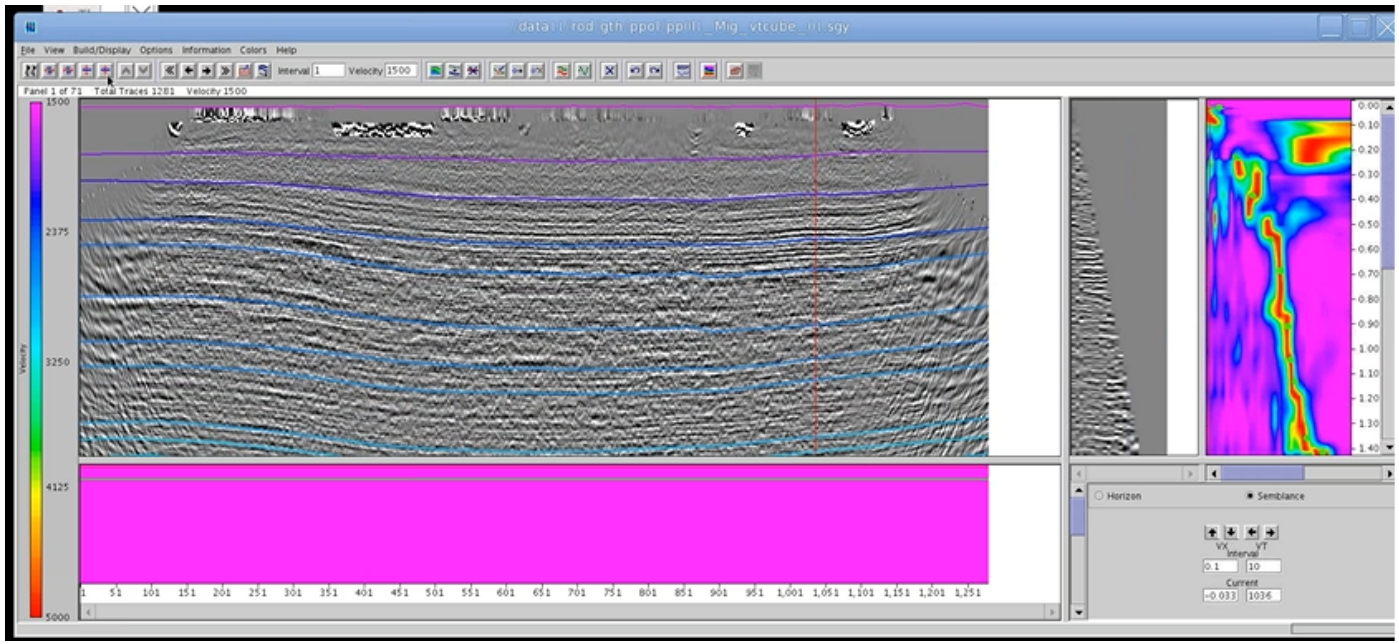
Principles of Velocity Analysis

Then, for each time slab and each midpoint we pick the maximum event associated with a velocity value.



Principles of Velocity Analysis

This is commonly done by picking structurally consistent horizon strands of the highest amplitude along the midpoint direction from the CVM panels.



Migration Velocity Analysis

MVA is an iterative process of two stages:

1. Data is imaged by prestack migration.
2. Velocity function is updated on the basis of the migration results. The accuracy of the velocity function is determined by measuring the focusing of reflections in the migrated data.

Existing Methods Drawbacks

- Expensive
  - Demands a lot of human manual tedious work
  - Computationally expensive
  - Time consuming (takes weeks...)
- Inaccurate
  - Can't handle complex velocity changes
  - Limited in terms of handling steep dips
  - Limited in terms of handling noise and multiples
  - Errors in this stage result in critical deterioration of the produced subsurface image

Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

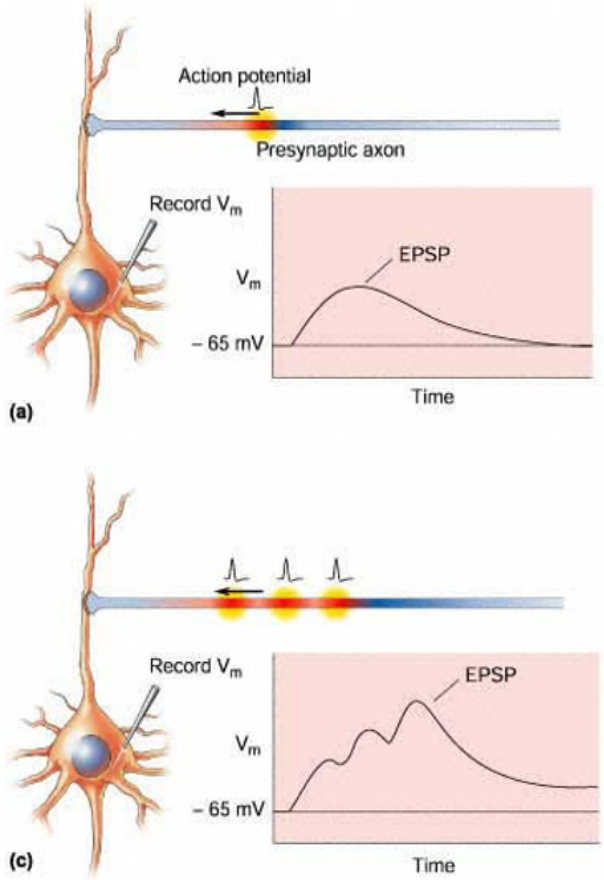
PTSM Velocity Analysis using RNN

Our contribution

- Develop an automatic, fast and efficient way to estimate velocity for prestack time-migration.
- Verify capability of the proposed automatic algorithm to extract all the information needed from raw seismic data to produce an actual true image of the subsurface.

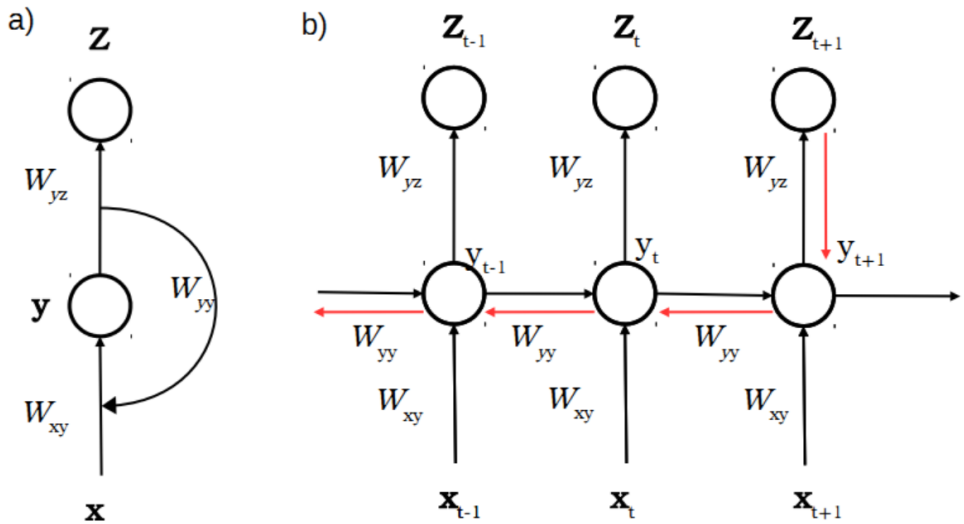
Neural Nets in Exploration Seismology

- First arrival picking (Murat et al., 1992; McCormack et al.,1993, Hollander 2018)
- Stacking velocity estimation (Calderón-Maciás et al., 1998; Fish et al., 1994; Schmidt et al., 1992; Biswas et al., 2018) *
- Seismic interpretation (Araya-Polo et al., 2017; Zhang et al.,2014)



Automated MVA

1. Use velocities picked manually to train the data.
2. RNN estimates V_{rms} for the entire volume.
3. Use the velocities to migrate the data and verify the performance of the algorithm.



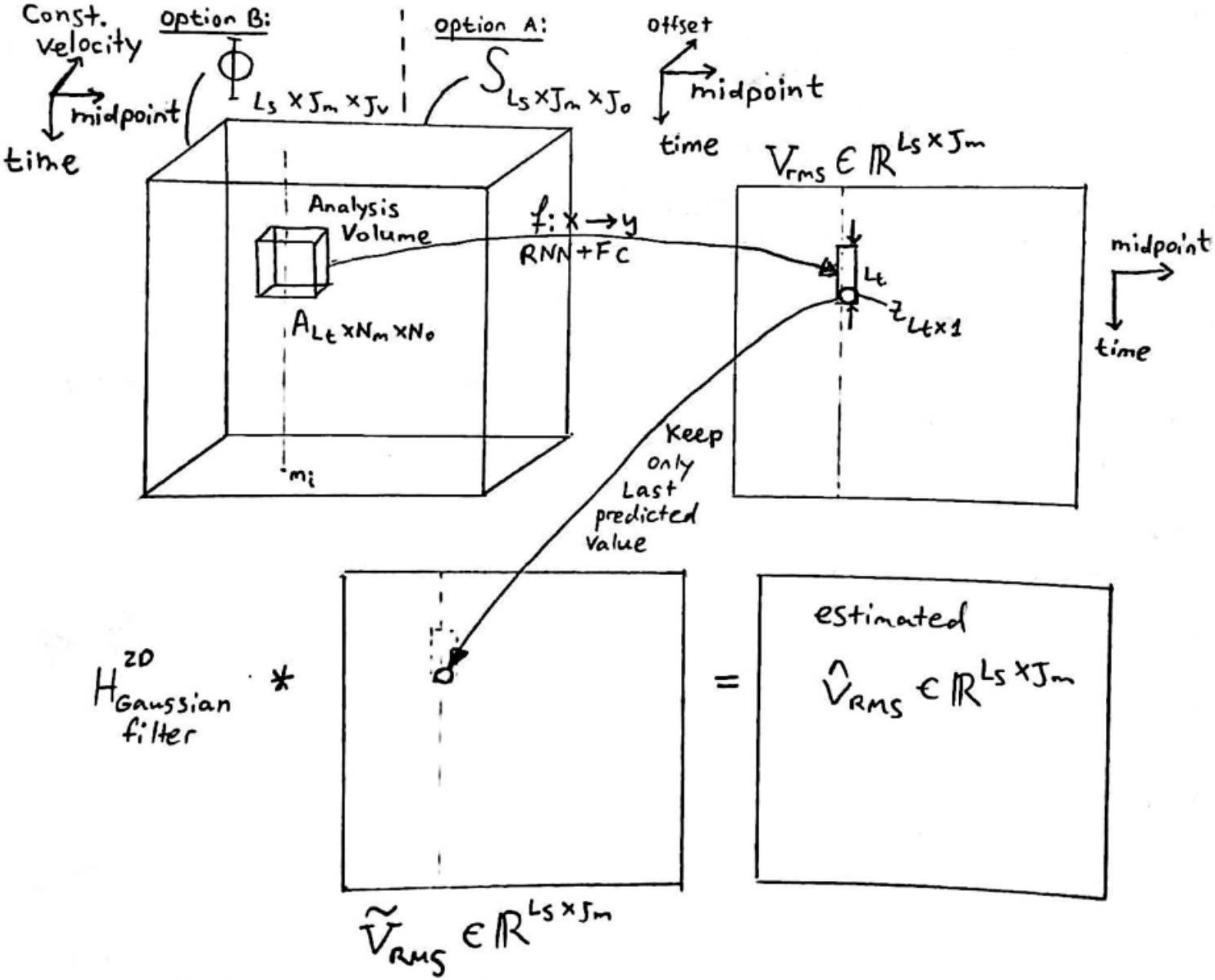
A simple example of a RNN: (a) The recursive form of RNN; (b) The extended form of RNN in time.

Automated MVA

RNN learns a mapping from the known input-output patterns $f : \mathbf{x} \rightarrow \mathbf{y}$. Suppose a given observation sequence $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t)$, and an expected output vector $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t)$. At time t, every neuron receives both the input vector $\mathbf{x}_t$ and the previous output $\mathbf{y}_{t-1}$, which can formulated as

$$\mathbf{y}_t = \phi(\mathbf{W}_{xy}^T \mathbf{x}_t + \mathbf{W}_{yy}^T \mathbf{y}_{t-1} + \mathbf{b}),$$

where ϕ is the activation function, $\mathbf{W}_{xy}$ and $\mathbf{W}_{yy}$ are the weight matrices, and $\mathbf{b}$ is the bias vector.

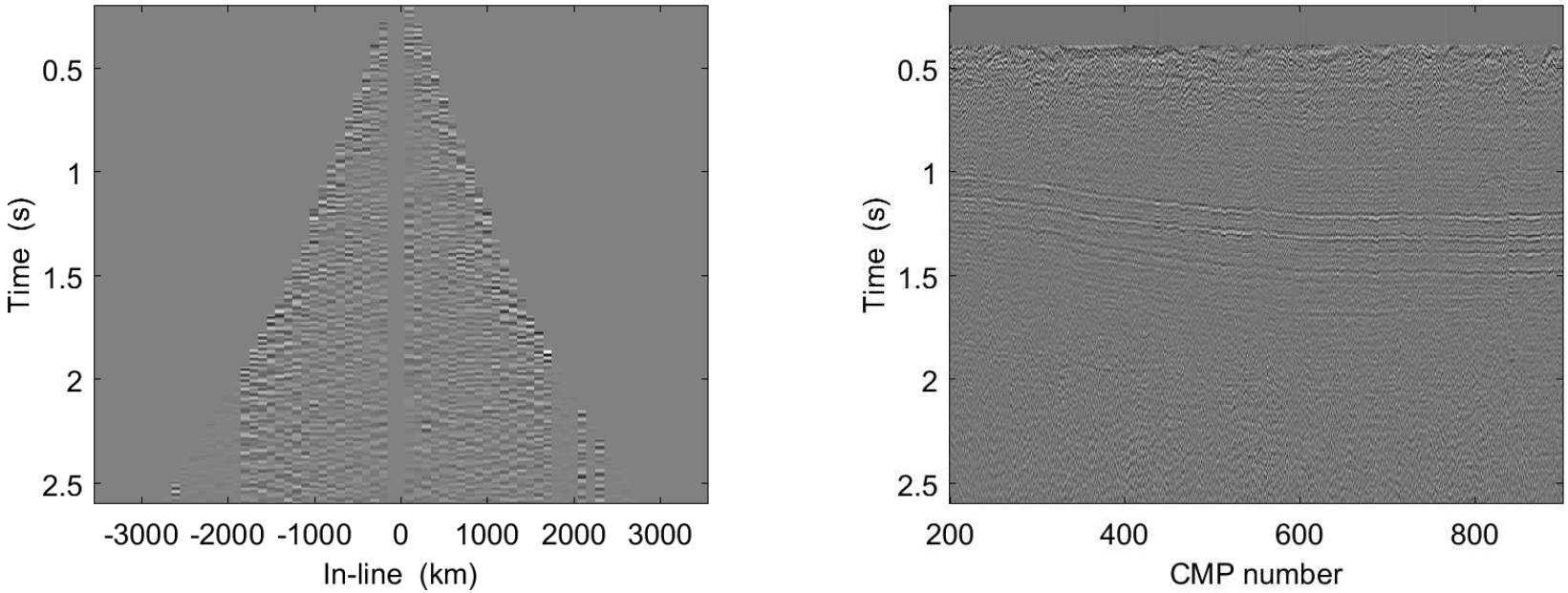

Const. velocity
midpoint
time
option B:
$L_S \times J_m \times J_v$
option A:
$S_{L_S \times J_m \times J_o}$
offset
midpoint
time
$V_{rms} \in \mathbb{R}^{L_S \times J_m}$
Analysis Volume
$A_{L_t \times N_m \times N_o}$
m_i
$f: x \to y$
RNN+FC
L_t
$z_{L_t \times 1}$
midpoint
time
Keep only Last predicted value
$H^{2D}_{Gaussian\ filter}$ *
$\tilde{V}_{RMS} \in \mathbb{R}^{L_S \times J_m}$
=
estimated
$\hat{V}_{RMS} \in \mathbb{R}^{L_S \times J_m}$

PTSM Velocity Analysis using RNN

Why RNN?

- RNN remembers both in space and time:
  - Signal travels in both directions.
  - Output of current state depends on output of previous states.
  - Estimation of velocity at each point considers neighboring points.
- Hypothesis: Data in same feature space and have the same distribution.

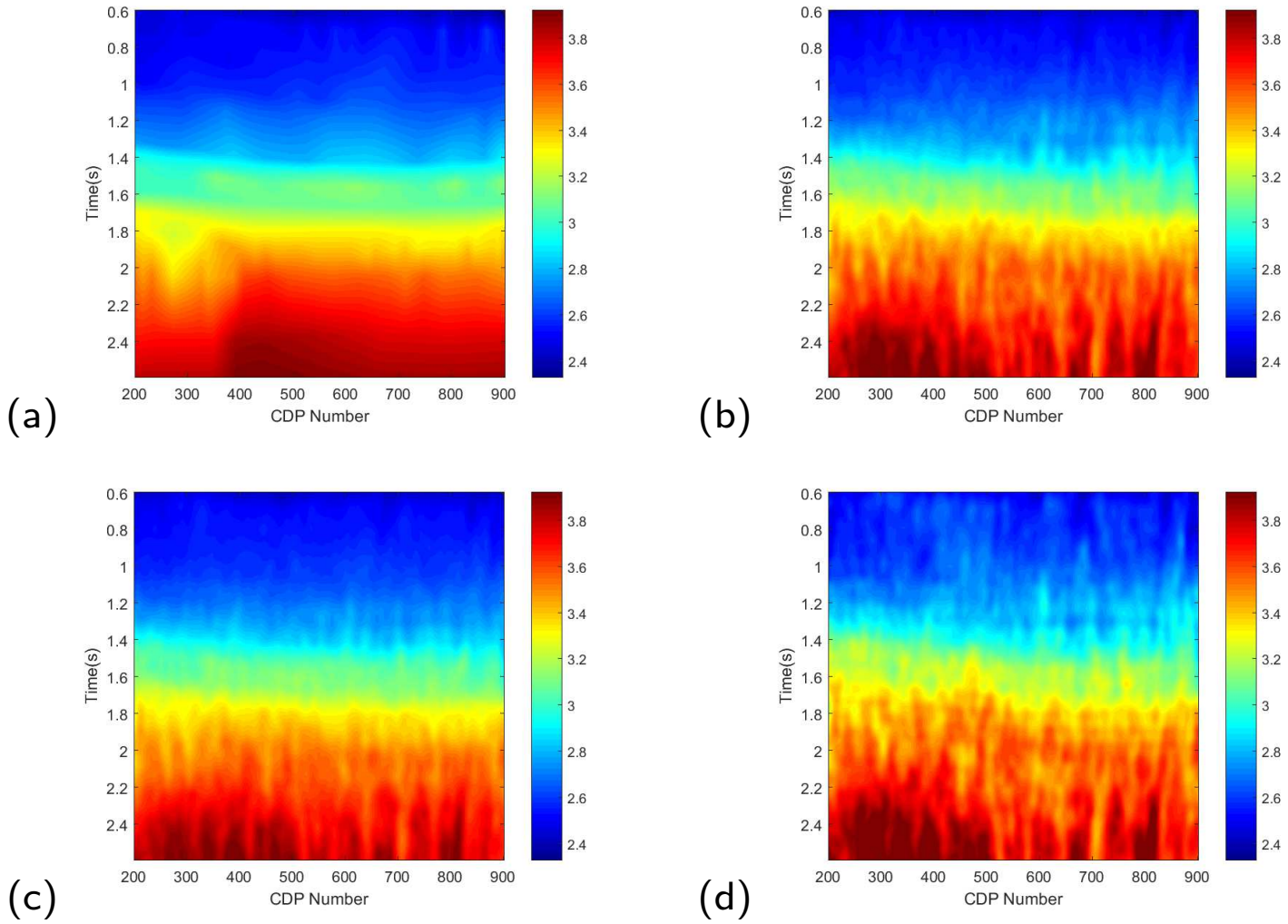
Real Data Results



(a) (b)

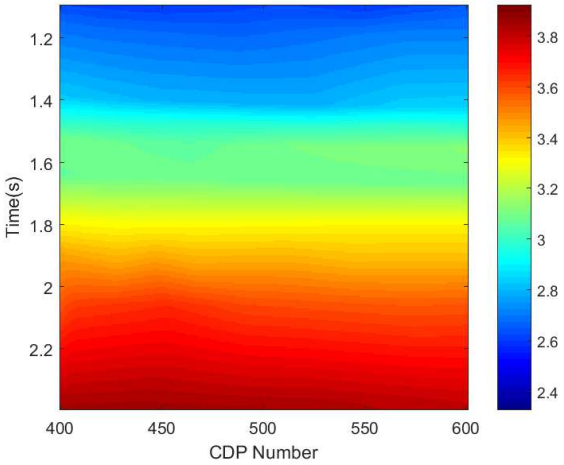
(a) CMP gather with 71 offsets; (b) common offset gather with 700 CMP traces.

Real Data Results - Migration Velocities

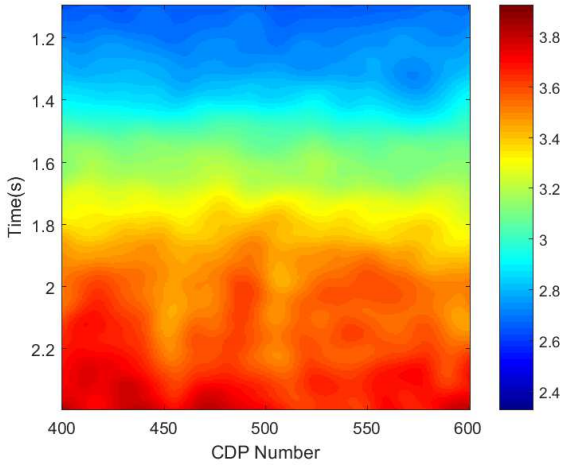


(a) manually picked velocities - ground truth; (b) predicted, $N_m = 1$, with 25% training; (c) predicted, $N_m = 1$ with 50% training; (d) predicted, $N_m = 3$, with 25% training.

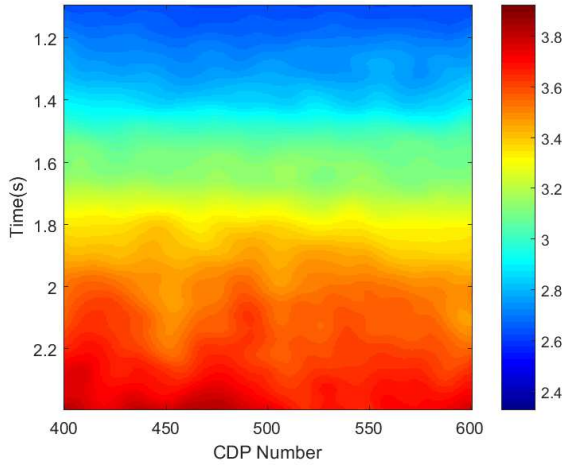
Real Data Results



(a)



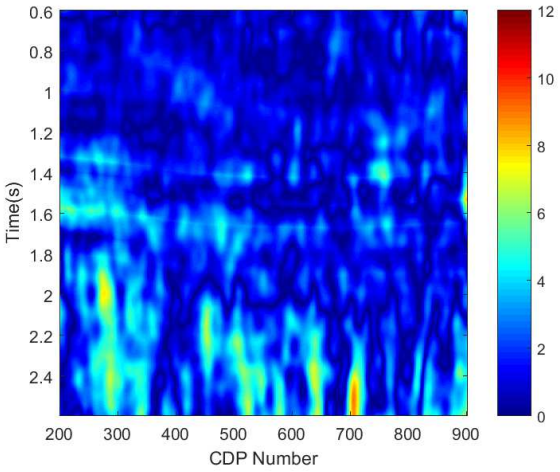
(b)



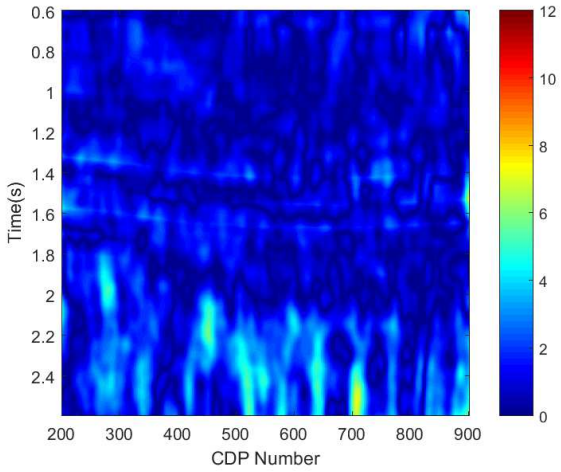
(c)

Migration velocities - zoom in: (a) picked velocities; (b) predicted, $N_m = 1$, with 25% training; (c) predicted single channel velocities, $N_m = 1$ with 50% training;

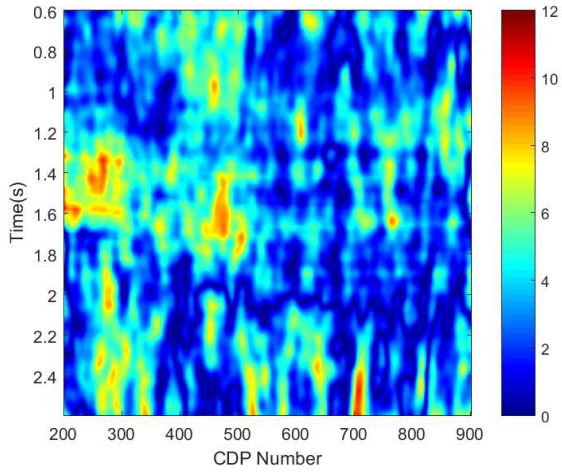
Real Data Results



(a)



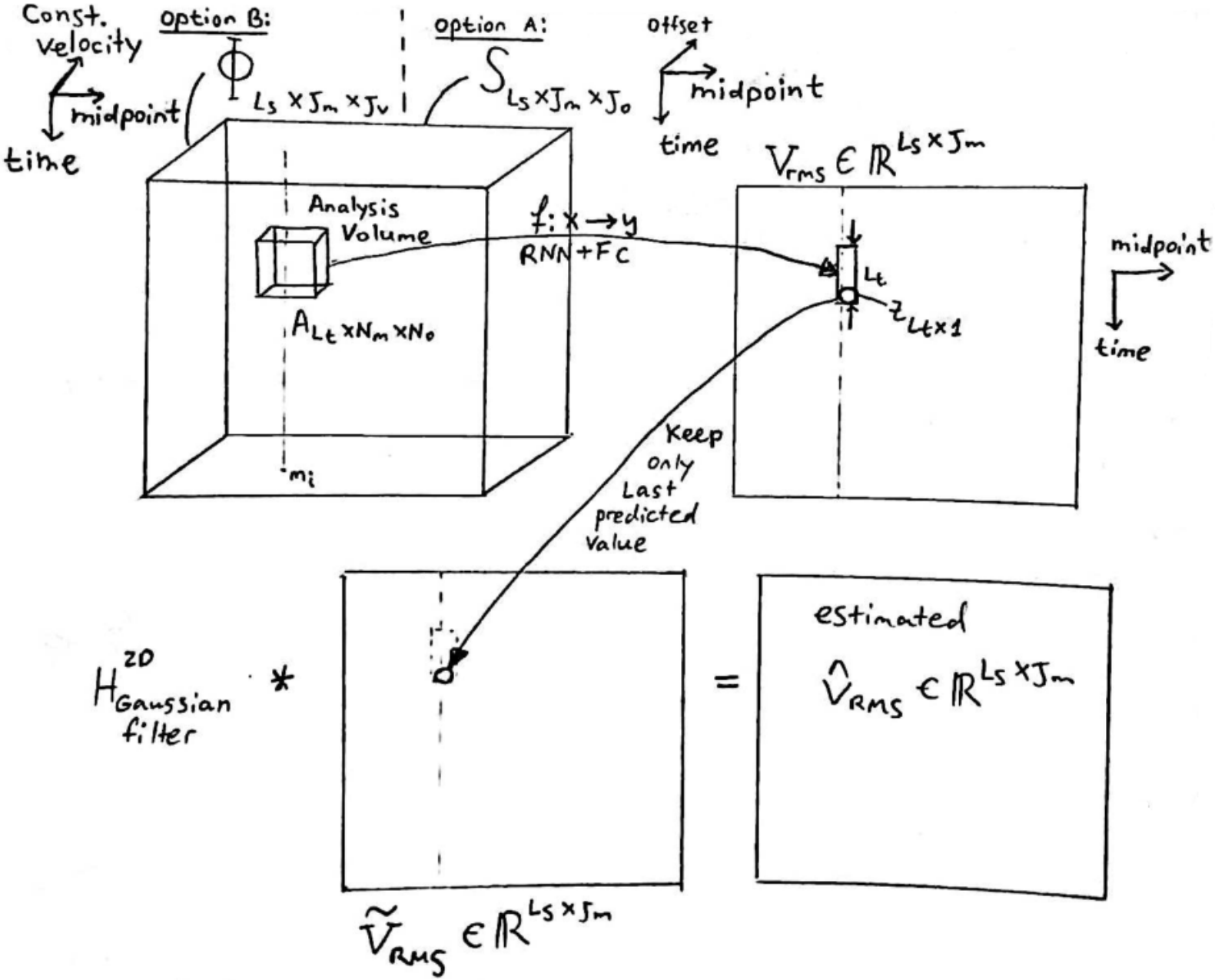
(b)



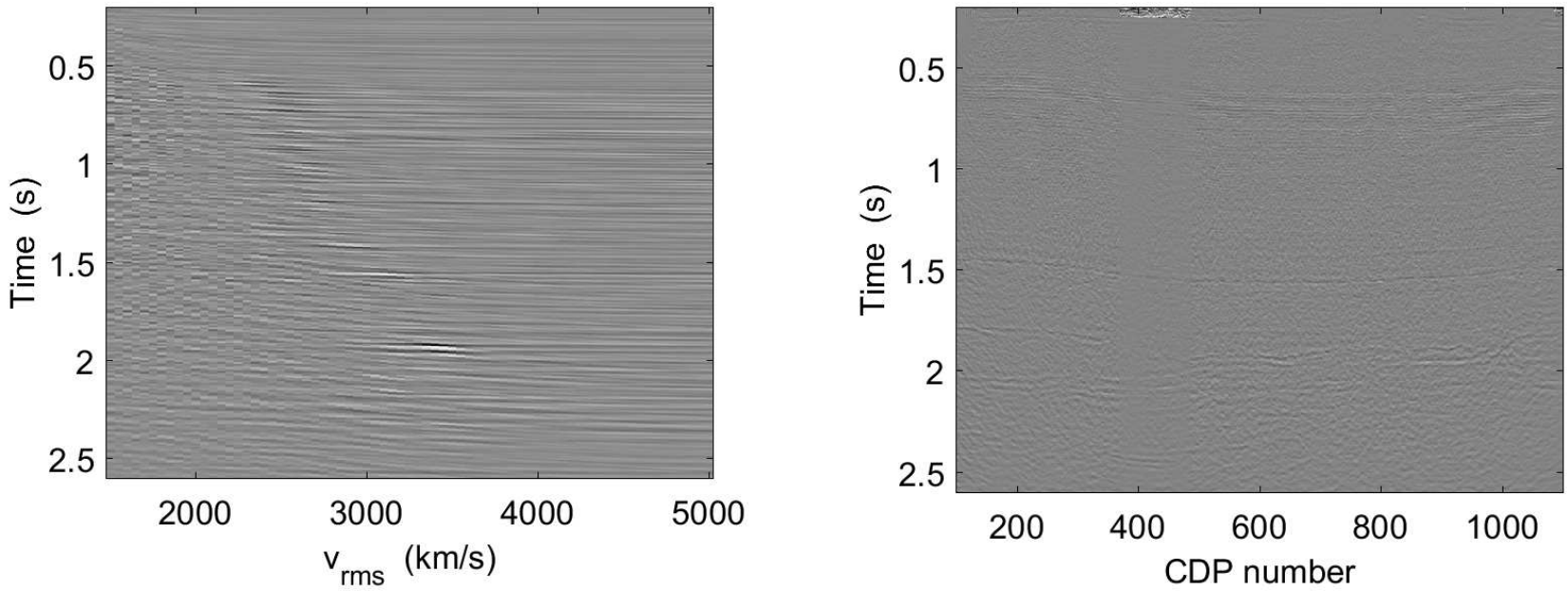
(c)

Percentage error: (a) predicted, $N_m = 1$ with 25% training; (b) predicted, $N_m = 1$, with 50% training; (c) predicted, $N_m = 3$, with 25% training.

QC


Const. velocity
midpoint
time
option B:
$L_S \times J_m \times J_v$
option A:
$S_{L_S \times J_m \times J_o}$
offset
midpoint
time
$V_{rms} \in \mathbb{R}^{L_S \times J_m}$
Analysis Volume
$A_{L_t \times N_m \times N_o}$
m_i
$f: x \to y$
RNN+FC
L_t
$z_{L_t \times 1}$
midpoint
time
Keep only Last predicted value
$H^{2D}_{Gaussian\ filter}$ *
$\tilde{V}_{RMS} \in \mathbb{R}^{L_S \times J_m}$
=
estimated
$\hat{V}_{RMS} \in \mathbb{R}^{L_S \times J_m}$

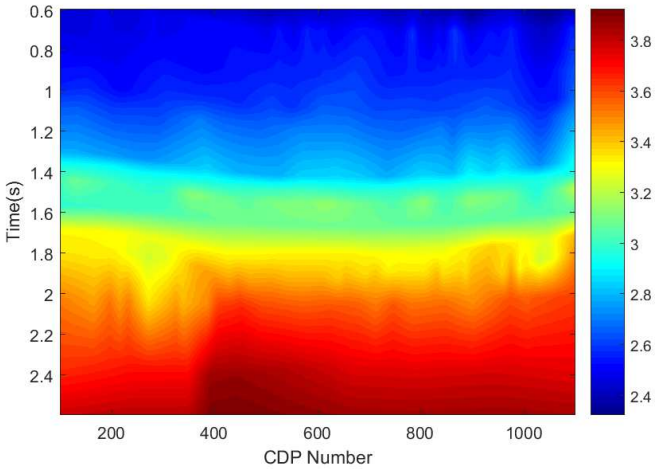
Real Data Results - CVM assisted



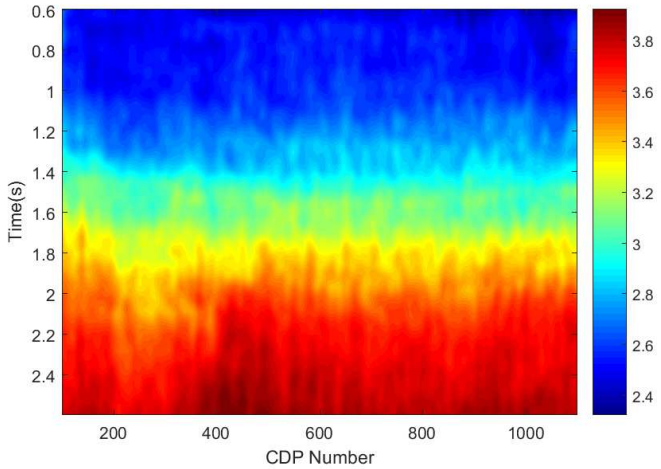
(a) (b)

Real seismic CVM panels: (a) a cross section at CDP index $m = 500$; (b) a cross section at a constant velocity of 4000 m/sec.

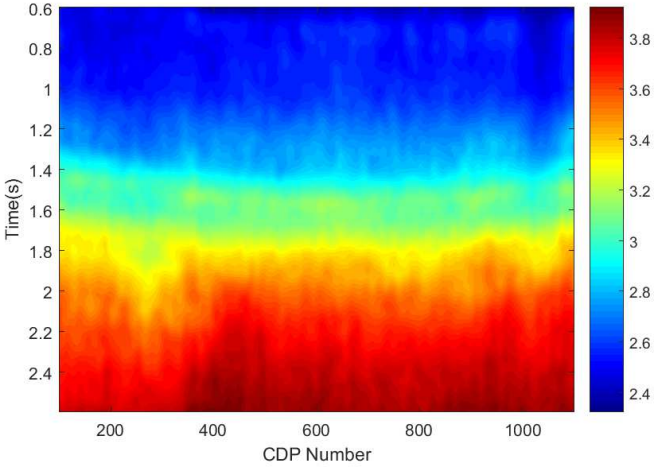
Real Data Results - CVM assisted



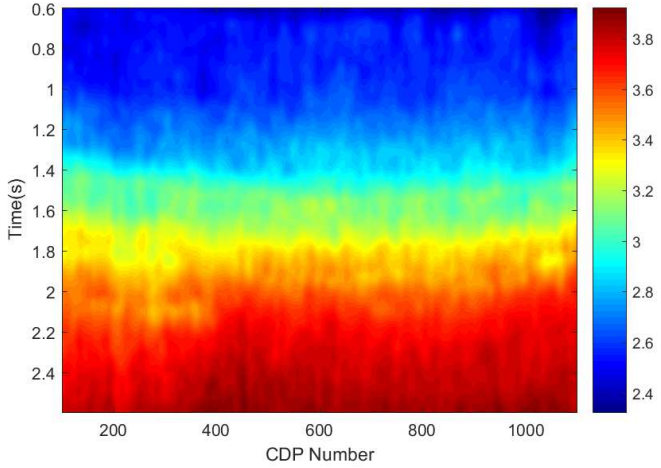
(a)



(b)



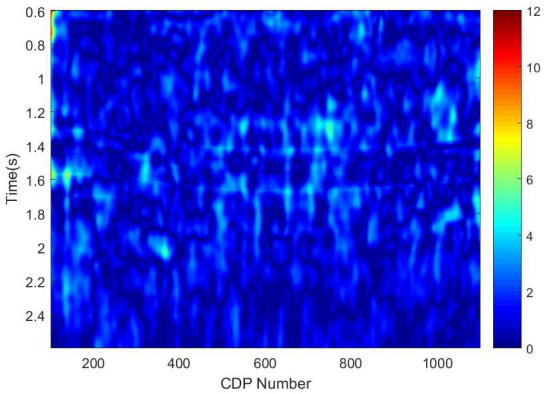
(c)



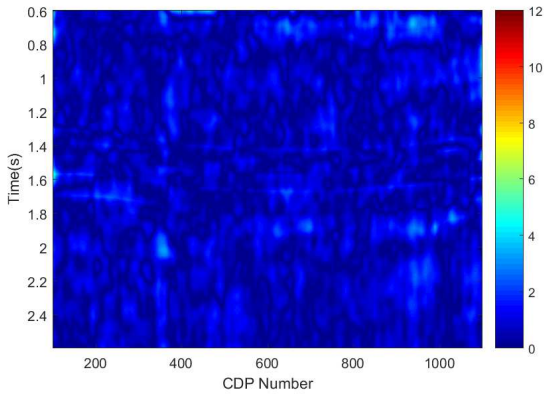
(d)

(a) manually picked velocities - ground truth; (b) predicted, $N_m = 1$, with 25% training; (c) predicted, $N_m = 1$ with 50% training; (d) predicted, $N_m = 3$, with 25% training.

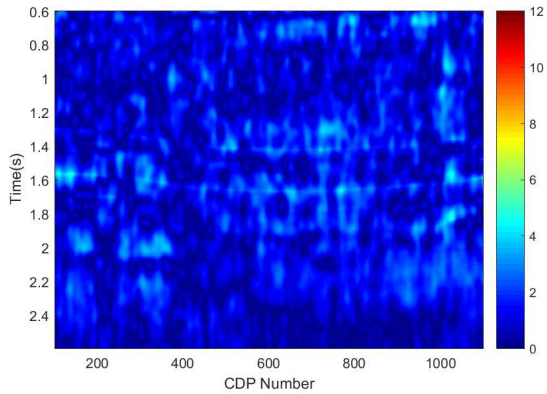
Real Data Results - CVM Assisted



(a)



(b)

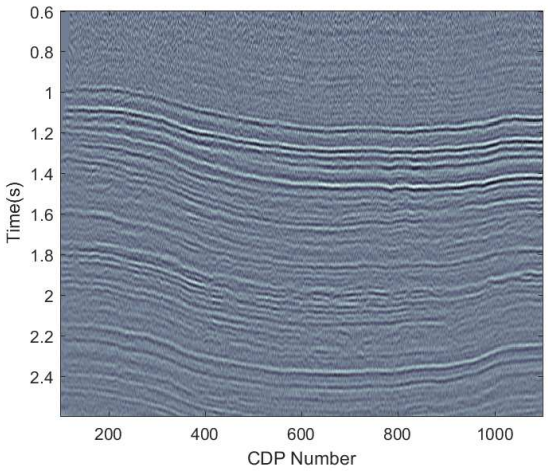


(c)

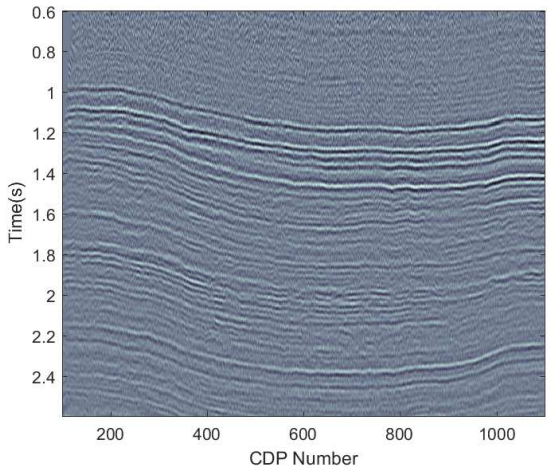
Percentage error: (a) predicted, $N_m = 1$ with 15% training; (b) predicted, $N_m = 1$, with 50% training; (c) predicted, $N_m = 3$, with 25% training.

- Training time on GPU (RTX-2080 Ti) $\sim$ **1 hour**
- Execution time $\sim 1 - 10$ secs

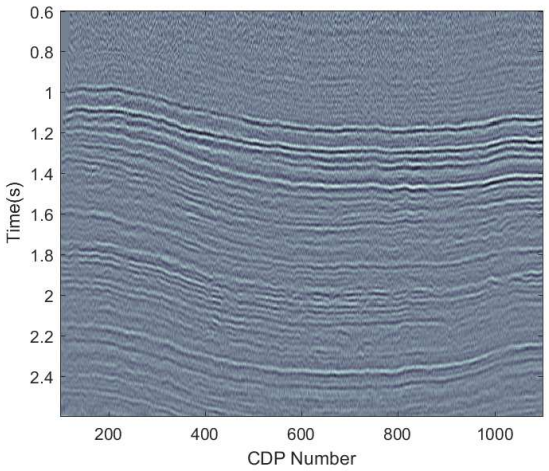
Real Data Results - CVM Assisted Migrated Images



(a)



(b)



(c)

PTSM based on CVMs MVA with: (a) picked velocities (courtesy of GeoEnergy); (b) predicted single channel velocities, $N_m = 1$, with 50% training; (c) predicted multichannel velocities, $N_m = 3$, with 25% training.

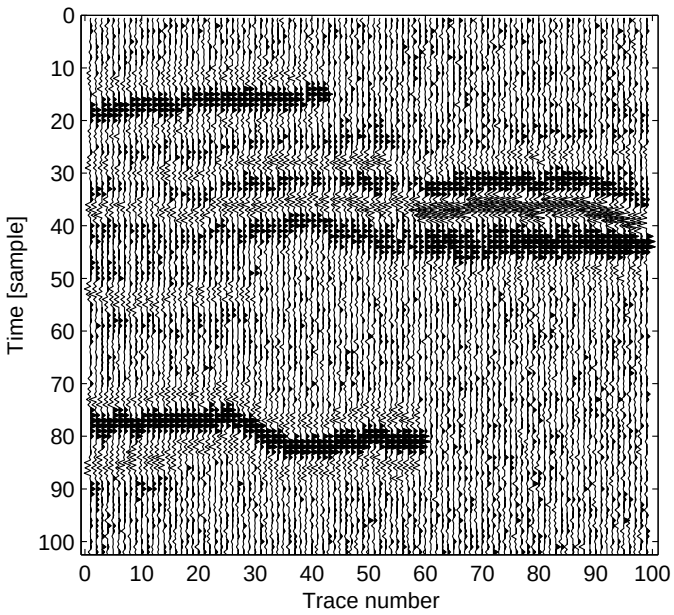
Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

Seismic Deconvolution

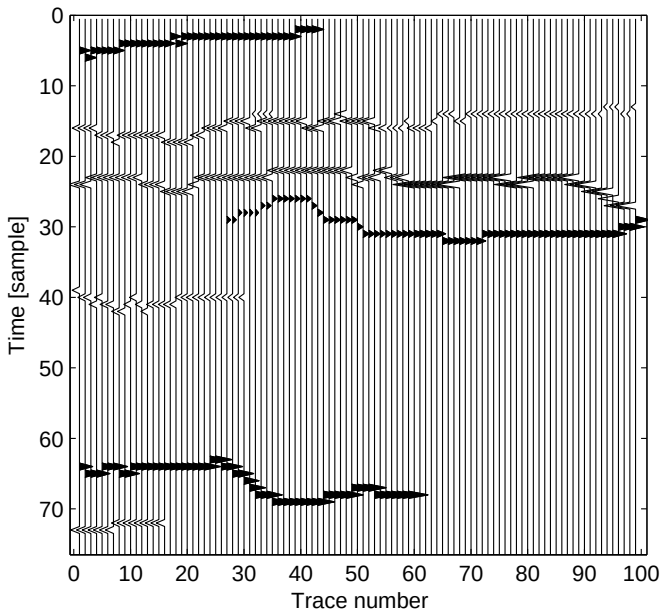
Final goal: Reveal ground layers structure.

Synthetic seismic data



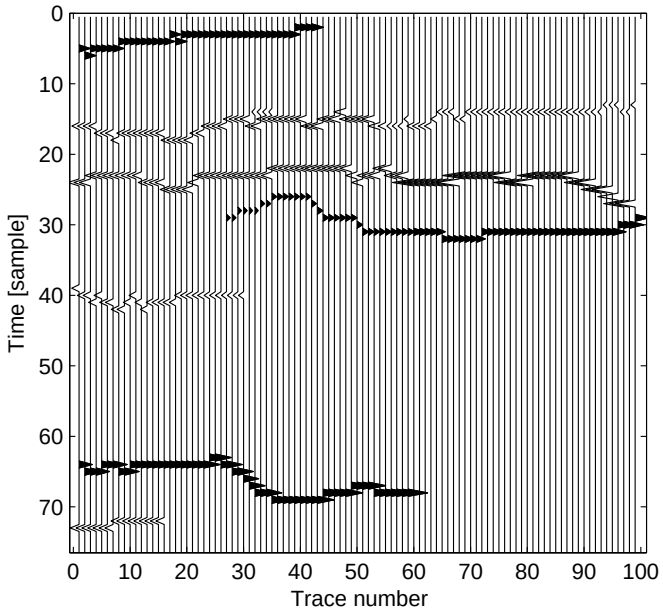
$\Longrightarrow$

Reflectivity



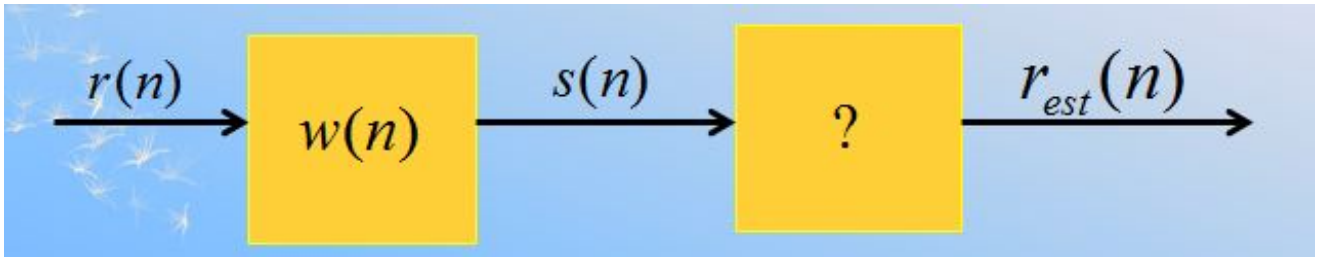
Problem Formulation

- Find an estimate of **the reflectivity**.
- Reflectivity is assumed to be **sparse**.
- Option 1: break the problem into independent **1D** deconvolution problems.

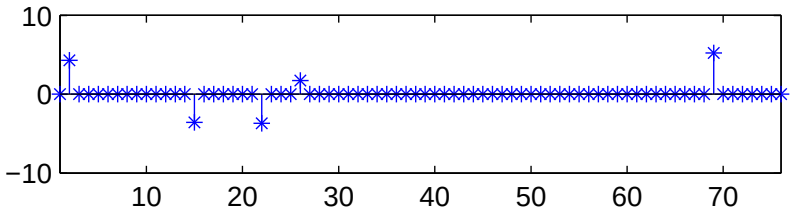


Deconvolution

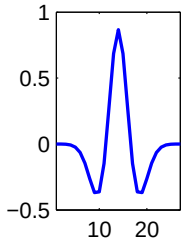
- Deconvolution - ideally reverses the effect of a convolution performed by LTI system operating on an input signal.
- Output and operating system are known. Input is unknown.
- Goal: reconstruct the input signal.



Time Invariant Model

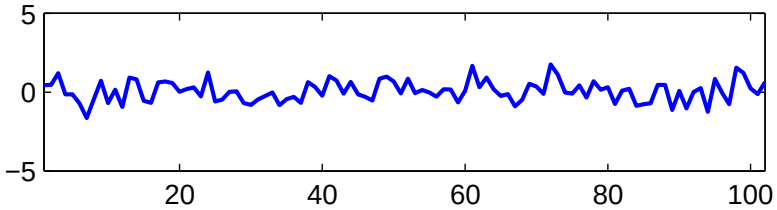


$*$



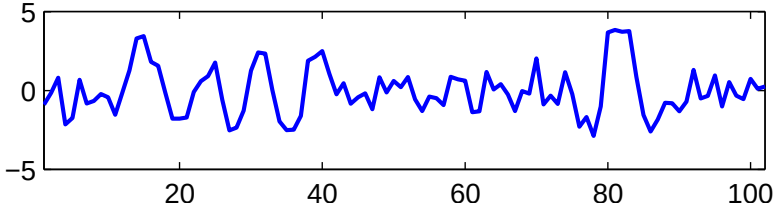
$r[k] * w[k]$

$+$



$n[k]$

$=$



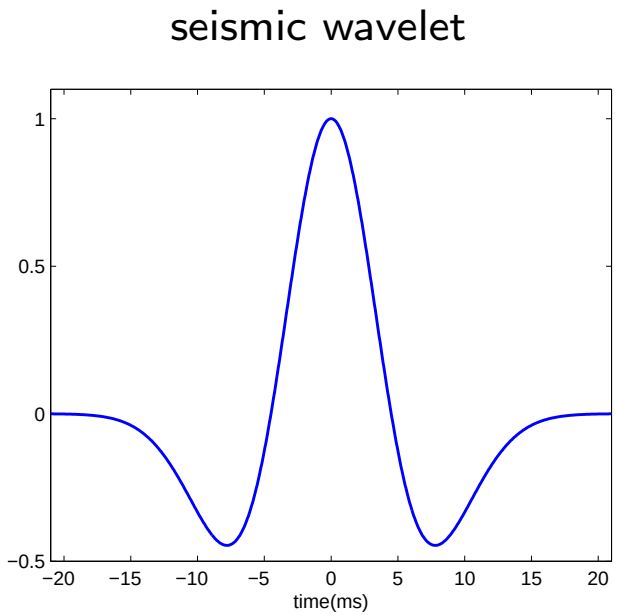
$s[k]$

Problem Formulation

$w(t)$ - the seismic wavelet (the blurring kernel)

Assumed to be **known**.

In practice it is estimated in pre-processing stages.



Atomic Decomposition

Assume a discrete dictionary of M waveforms $\{\phi_i\}_{i=1}^{M}$.

These waveforms are columns of an N by M matrix W.

The decomposition problem can be written as

$$\mathbf{s}_{N\times 1} = \sum_{i=1}^{M} \phi_i r_i$$

In matrix-vector form

$$\mathbf{s}_{N\times 1} = W_{N\times M}\mathbf{r}_{M\times 1} \quad => \quad \text{Goal : Find} \quad \mathbf{r}.$$

Problem Formulation

- After time discretization, our problem could be formulated as

$$\boxed{\mathbf{s} = W\mathbf{r} + \mathbf{n}}$$

$W \in \mathbb{R}^{L_s \times L_r}$, the dictionary, is a convolution matrix formed by shifts of the seismic wavelet.

- Since w is band-limited $W^{\dagger}$ is severely ill-conditioned, the solution obtained by using $W^{\dagger}$ is non-interpretable.

$$\mathbf{r} \neq W^{\dagger}\mathbf{s}$$

Previous Works

- **Single-Channel** Blind Deconvolution Using Statistical Methods
  - Linear Prediction

    Robinson and Treitel, 1964
  - Minimum Entropy Deconvolution

    Wiggins, 1978

Previous Works

- Seismic **Single-Channel** Deconvolution Using **Sparse Representations**
  - Sparse Spike Inversion (SSI)

    Taylor at al., 1979
  - Basis Pursuit Inversion (BPI)

    Zhang and Castagna, 2011

Sparse Seismic Inversion (SSI)

$$\mathbf{s} = \mathbf{W}\mathbf{r} + \mathbf{n}$$

Recover each column of the reflectivity by solving:

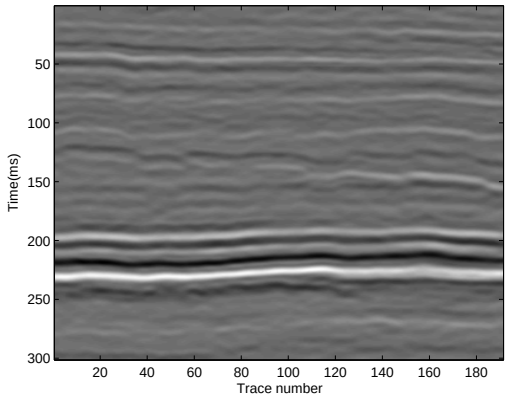
$$\min \|\mathbf{r}\|_0 \ \text{subject to} \ \|\mathbf{s} - \mathbf{W}\mathbf{r}\|_2^2 < \varepsilon$$

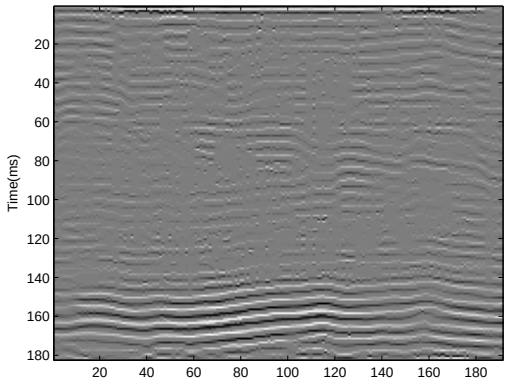
which is relaxed into:

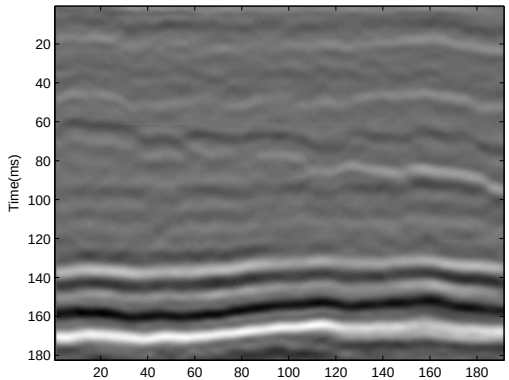
$$\min_{\mathbf{r}} \frac{1}{2} \|\mathbf{s} - \mathbf{W}\mathbf{r}\|_2^2 + \lambda \|\mathbf{r}\|_1$$

Sparse Seismic Inversion (SSI)

Seismic data



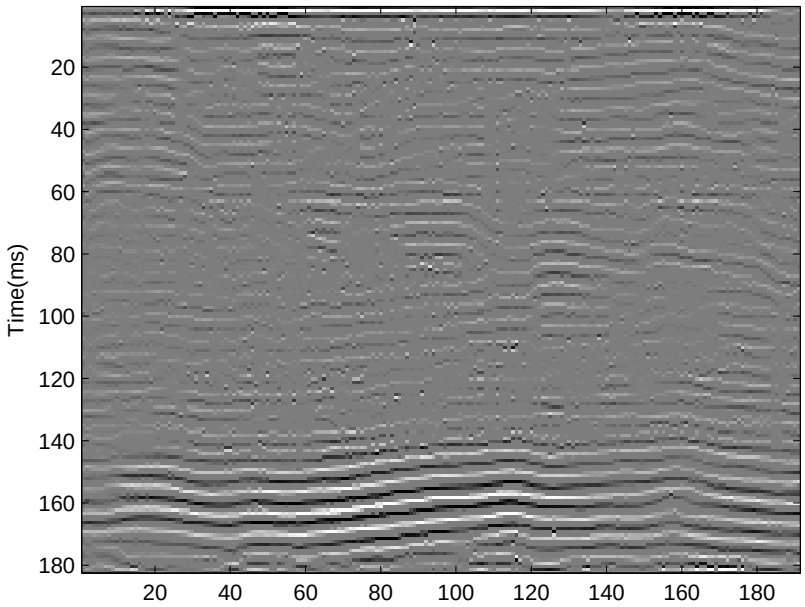




(a) Estimated reflectivity matrix; (b) Reconstructed seismic data. $\rho_{s,\hat{s}} = 0.95$ for $\lambda_{\text{opt}} = 9.4 \times 10^{-3}$.

Sparse Seismic Inversion (SSI)

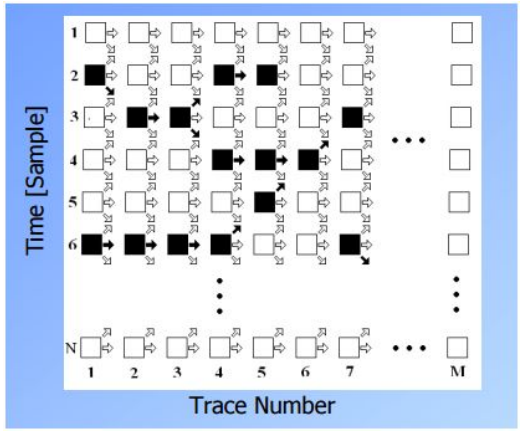
Estimated reflectivity matrix



- Sufficient performance under high signal to noise ratio (SNR).
- Ignores the correlation between neighboring traces

Multichannel Deconvolution Statistical Methods

- "Multichannel Seismic Deconvolution Using Markov-Bernoulli Random Field Modeling"
  Alon Heimer and Israel Cohen, 2009
- "Multichannel Deconvolution of Seismic Signals Using Statistical MCMC Methods"
  Idan Ram, Israel Cohen and Shalom Raz, 2010



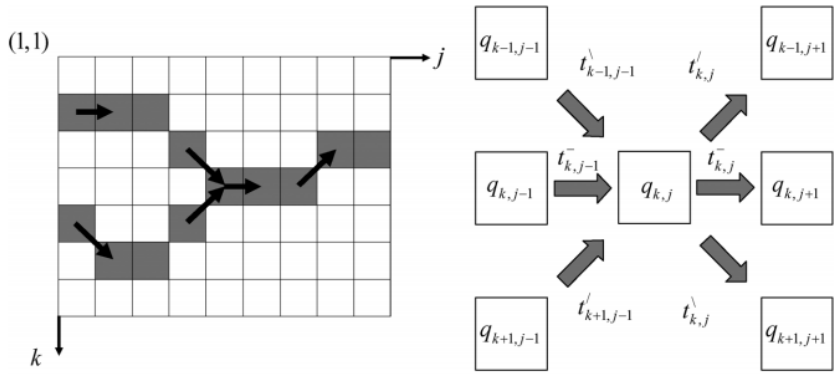


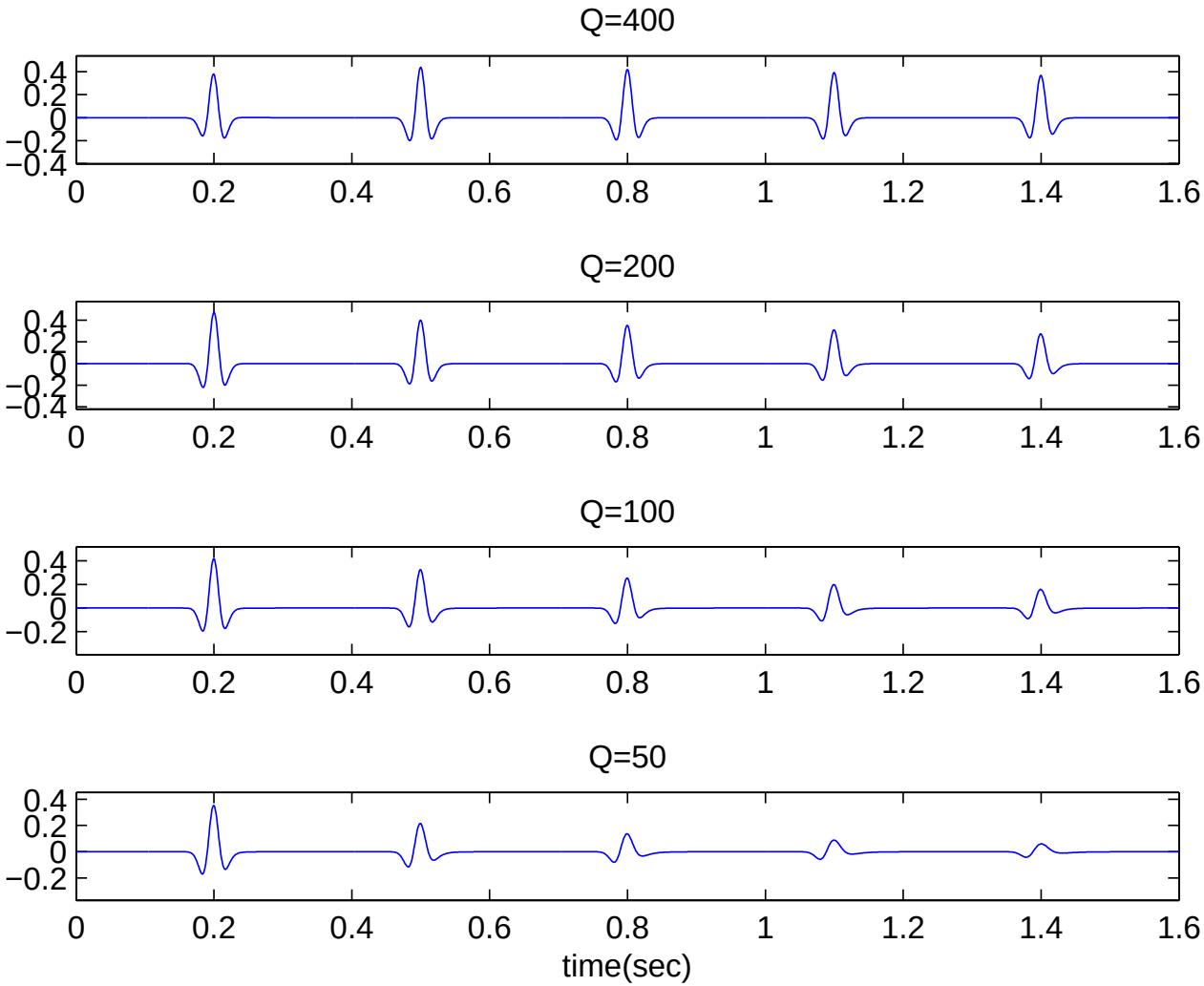
Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

Earth Q model

- When seismic waves propagate through the earth subsurface we observe two effects:
  1. Dissipation
     - $\Longrightarrow$ Wave energy is absorbed by the media.
     - $\Longrightarrow$ Wavelet's amplitude decreases with depth.
  2. Velocity dispersion
     - $\Longrightarrow$ Different frequencies travel with different speeds.
     - $\Longrightarrow$ Wavelet's width broadens with increased traveltime.

Earth Q model

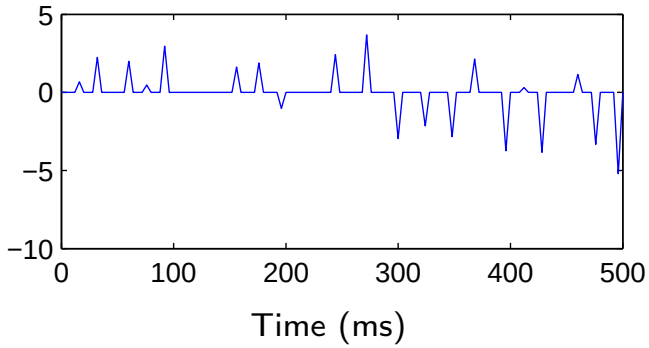


Problem Formulation

A 1D channel (column) in the unknown 3D reflectivity signal is a superposition of point sources. In the discrete setting

$$x[k] = \sum_m c_m \delta[k - k_m], \qquad k \in \mathbb{Z}, \qquad c_m \in \mathbb{R}$$

where $\delta[k]$ denotes the Kronecker delta function, and $\sum_m |c_m| < \infty$. $K = \{k_m\}$ is the set of discrete delays corresponding to the spikes locations.



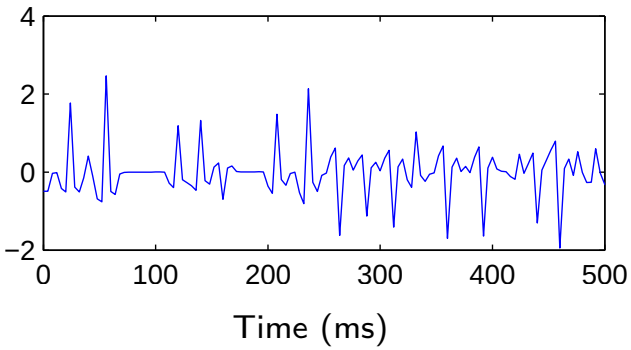
Problem Formulation

We consider a seismic discrete trace of the form

$$y[k] = \sum_n x[n] g_{\sigma,n}[k-n] + w[k], \qquad n \in \mathbb{Z} \quad (*)$$

$w[k]$ is additive noise.

- The set of kernels (pulses) $\{g_{\sigma,n}\}$ is known.
- The delays $T \triangleq \{t_m\}$ and the real amplitudes c_m are unknown.



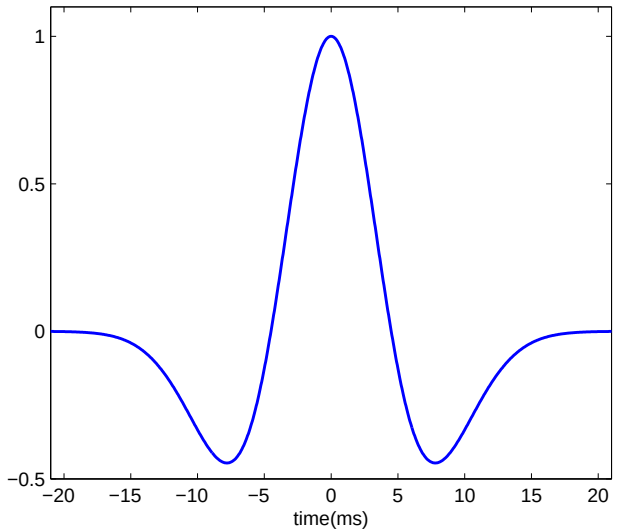
Earth Q model

Assume a source waveform $s(t)$ defined as the real-valued Ricker wavelet.

$$s(t) = \left(1 - \frac{1}{2}\omega_0^2 t^2\right) exp\left(-\frac{1}{4}\omega_0^2 t^2\right)$$

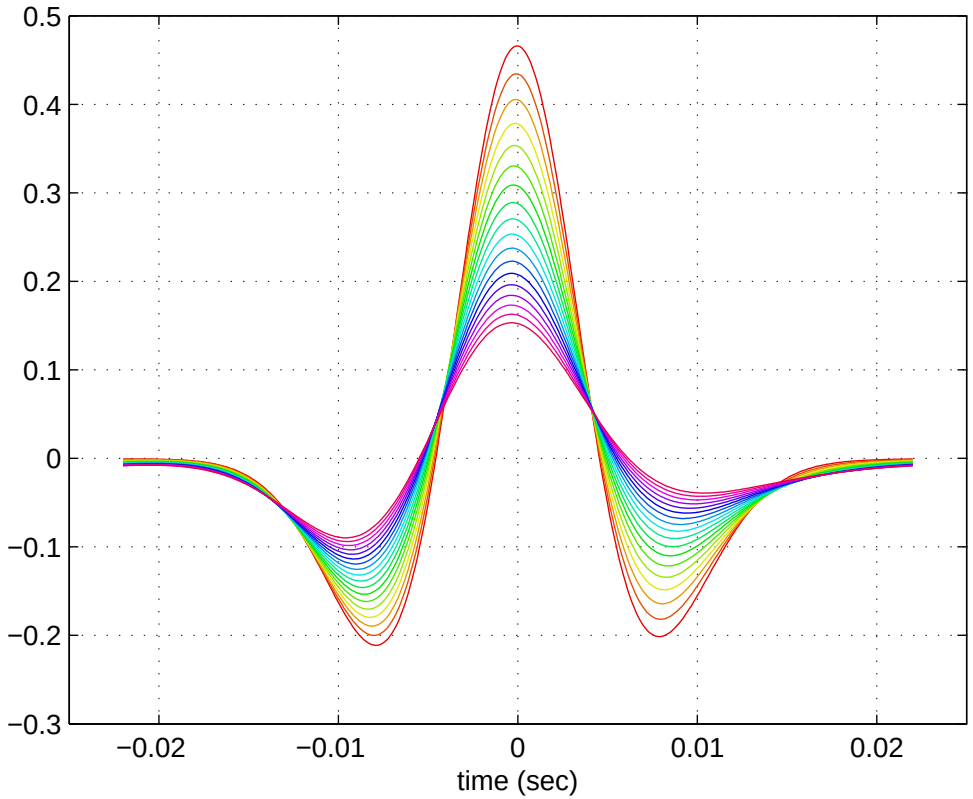
where ω_0 is the most energetic (dominant) radial frequency.

seismic wavelet



A Closer Look

Centered synthetic reflected wavelets $g_{\sigma,m}(t)$



$$Q = 25, \quad t_m = 10, 20, ..., 180ms, \quad \omega_0 = 100\pi \ (50)\text{Hz}$$

Our Objective

Given 3D seismic signals, we are to:

- Develop time-variant multichannel model
- Recover seismic reflectivity and find out:
  - To what **accuracy** can we recover each spike?
  - How does this accuracy depend on
    - noise level
    - spikes amplitude
    - medium Q constant
    - spikes separation
    - wavelets' shape
    - number of channels

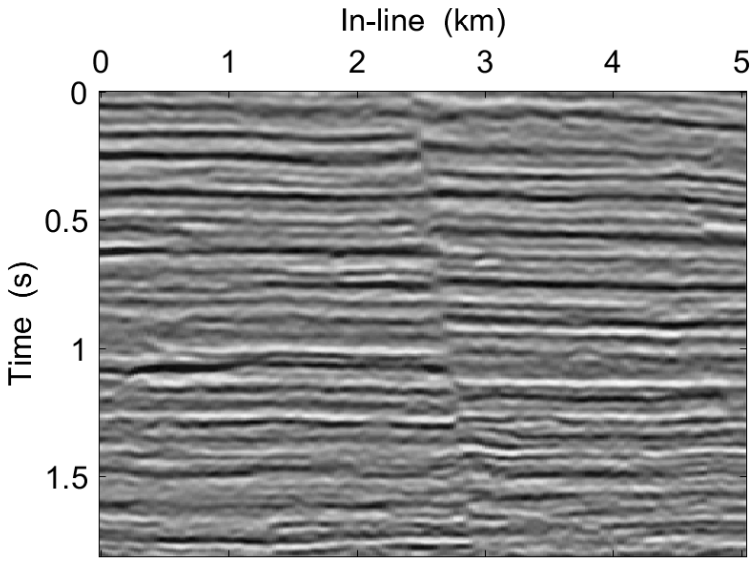
Multichannel Seismic Deconvolution

- $\mathbf{y}_{i,j}$ - an inline or a crossline seismic trace
- $\mathbf{y}_{i+u,j+v}$ - $N-1$ spatially neighboring traces.

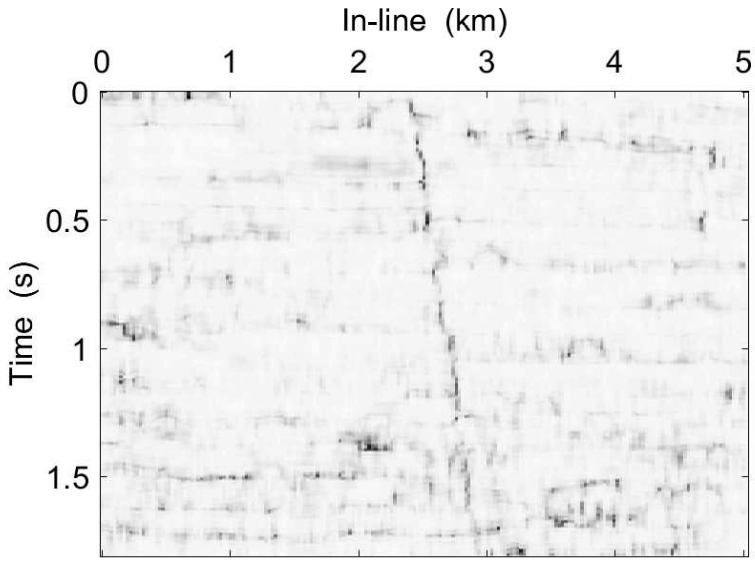
  $(u,v) \in \Gamma$ such that $\Gamma \subseteq \{(u,v) \in \mathbb{Z}^2 \ , \ (u,v) \neq (0,0)\}$ and

  $|\Gamma| = N-1$.
- Denote some local discontinuity measure as a column vector $\mathbf{a}_{i,j}$.
  - Each element $a_{i,j}[k]$ is associated with a distinguished point (i,j,k) in some analysis cube .
  - $0 \leq a_{i,j}[k] \leq 1$.
  - $a_{i,j}[k] = 0$ - maximum discontinuity.
  - $a_{i,j}[k] = 1$ - minimum discontinuity .

Local Discontinuity Measure - Example



(a) seismic data



(b) Local Structural Entropy, volume size [2 2 14].

LSE

Multichannel Seismic Deconvolution

- $0 \leq a_{i,j}[k] \leq 1$ - discontinuity measure.
- $\mathbf{y}_{i,j} = G\mathbf{x}_{i,j} + \mathbf{w}_{i,j}$
  - G is an operator matrix such that $(G)_{k,l} = g_l[k-l]$.
  - $\mathbf{x}_{i,j}$ is the corresponding reflectivity column and $\mathbf{w}_{i,j}$ is an additive noise.

Multichannel Seismic Deconvolution

The estimated reflectivity column $\hat{\mathbf{x}}_{i,j}$ is the solution of the optimization problem

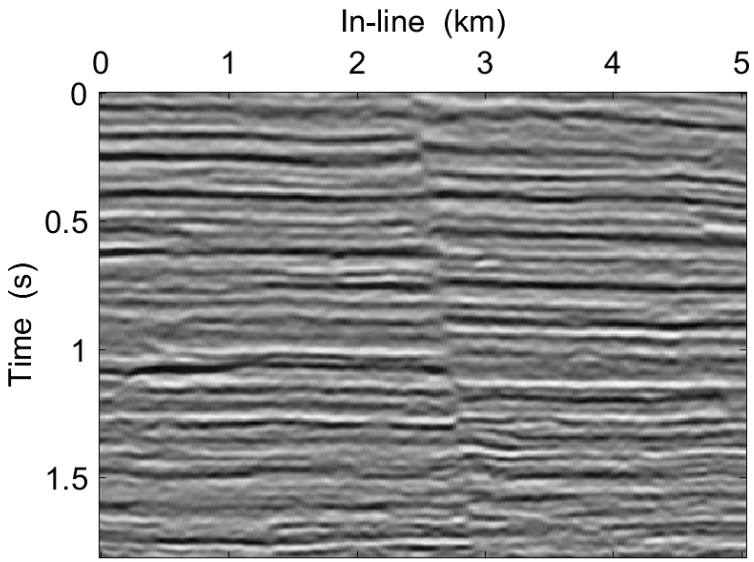
$$\begin{aligned} \min \quad & \|\hat{\mathbf{x}}_{i,j}\|_1 \\ \text{subject to} \quad & f(\hat{\mathbf{x}}_{i,j}) \leq \Delta, \end{aligned}$$

where

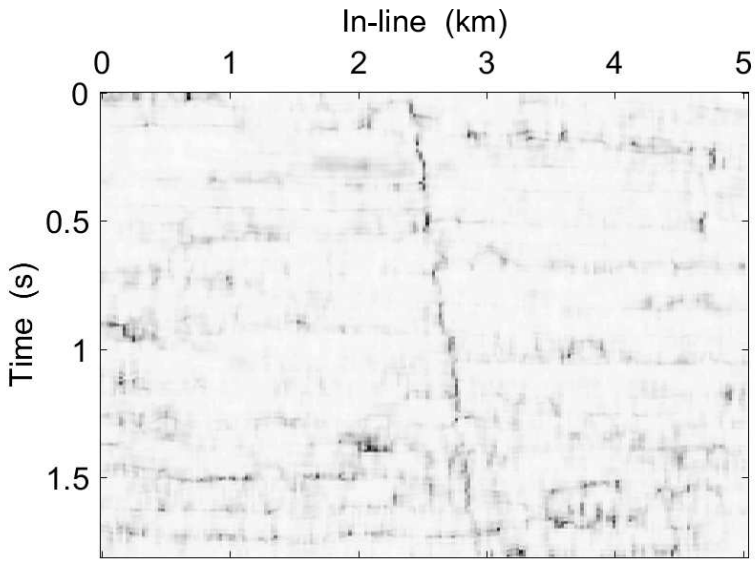
$$f(\hat{\mathbf{x}}_{i,j}) = \|\mathbf{y}_{i,j} - G\hat{\mathbf{x}}_{i,j}\|_2 + \sum_{(u,v)\in\Gamma} \|A_{i,j}A_{i+u,j+v}(\mathbf{y}_{i+u,j+v} - G\hat{\mathbf{x}}_{i,j})\|_2$$

$A_{i,j} = \mathrm{diag}(\mathbf{a}_{i,j})$, and $A_{i+u,j+v} = \mathrm{diag}(\mathbf{a}_{i+u,j+v})$.

Real Data Results



(a)



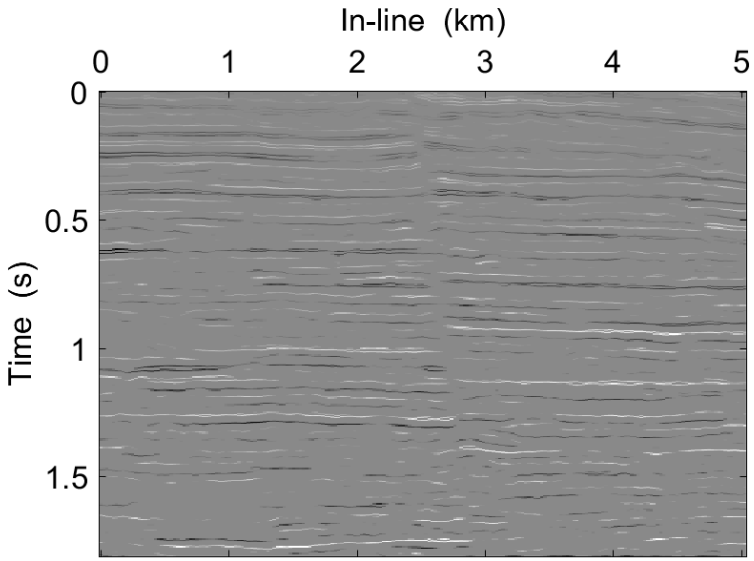
(b)

Vertical cross-section at $x = 2.5$km through: (a) seismic data (b) LSE (Local Structural Entropy) volume [2 2 14]. Estimated $Q = 233$*.

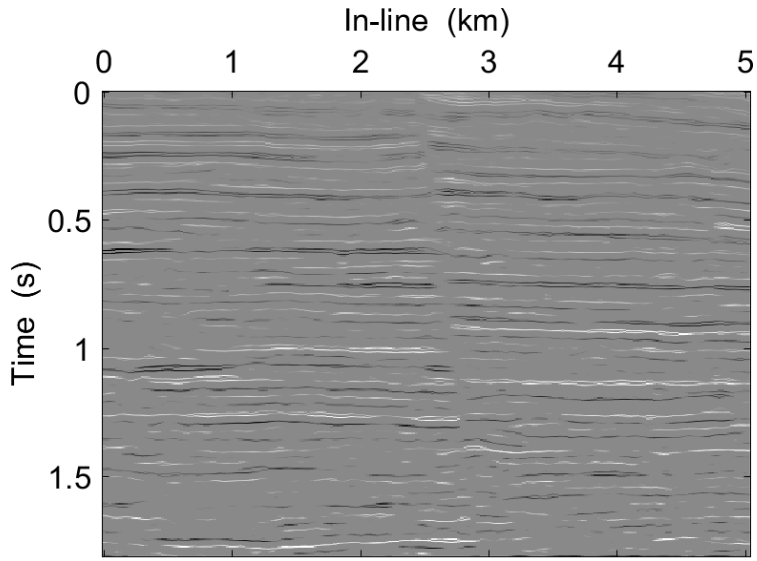
*Zhang & Ulrych, 2002

Real Data Results

Real data inversion results: Estimated reflectivity.

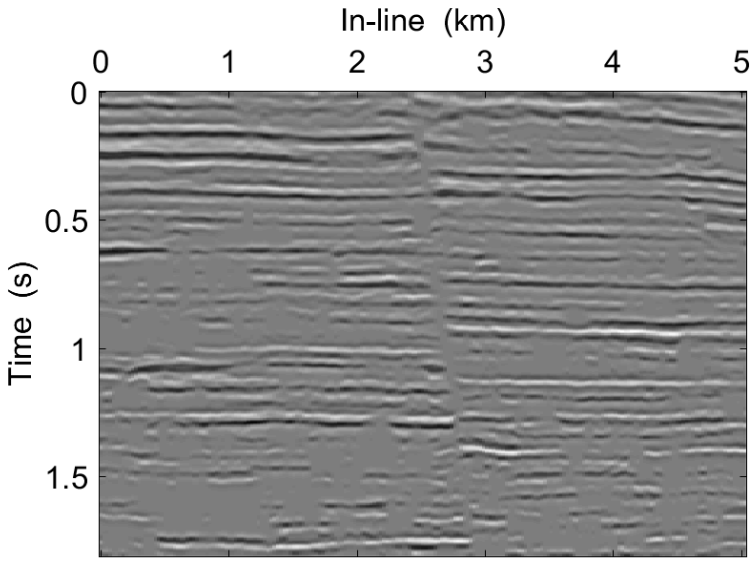


(a) single-channel

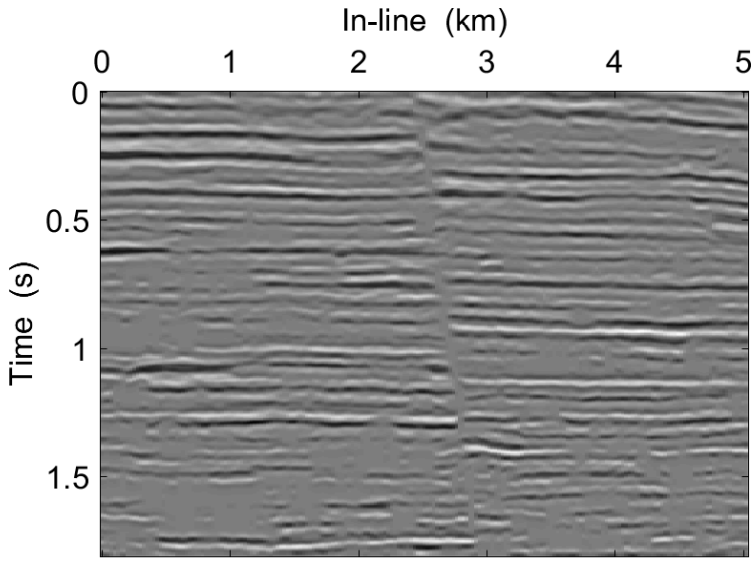


(b) multichannel

Real Data Results - Reconstructed seismic data

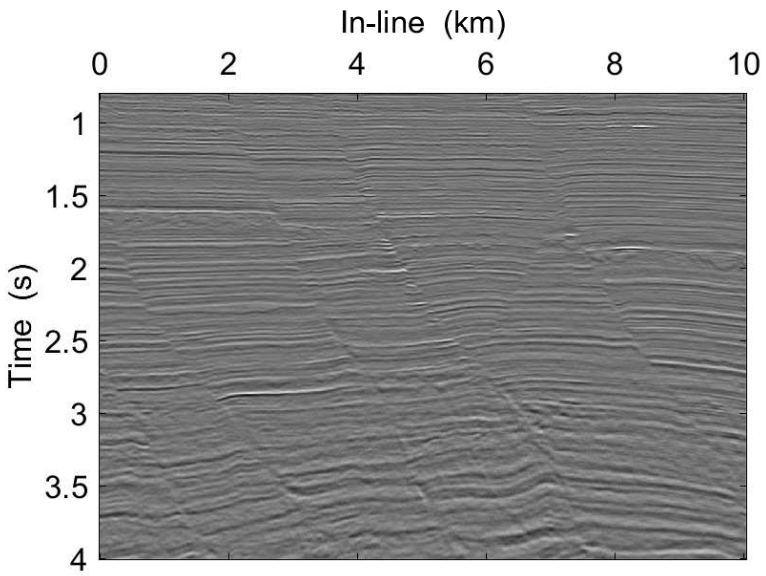


(a) single-channel, $\rho_{y,\hat{y}} = 0.78$

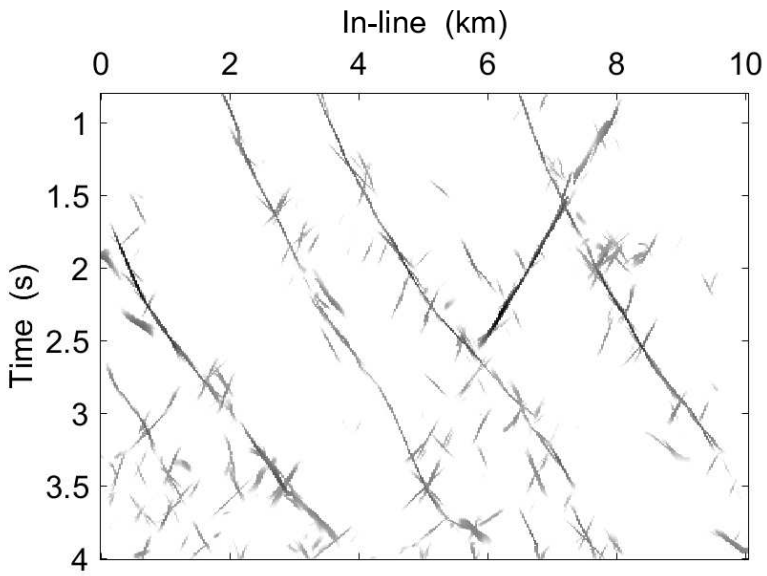


(b) multichannel, $\rho_{y,\hat{y}} = 0.82$

Real Data Results



(a)



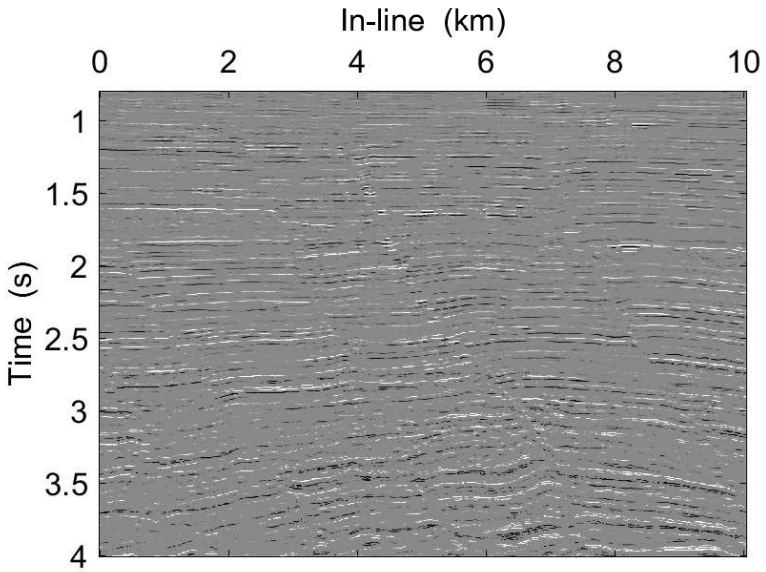
(b)

Vertical cross-section at $y = 1.3$km through:

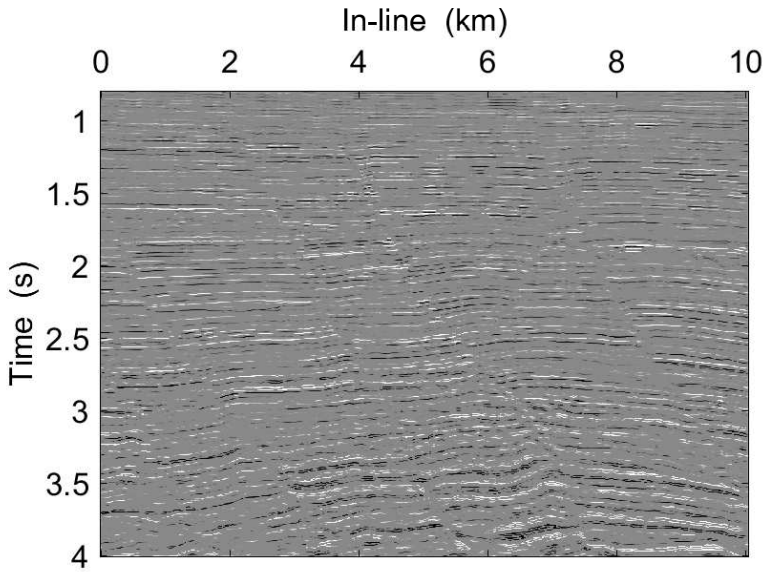
(a) seismic data (b) skeletonized LFE (Local Fault Extraction) volume.

Estimated $Q = 200$.

Real Data Results - Estimated reflectivity.

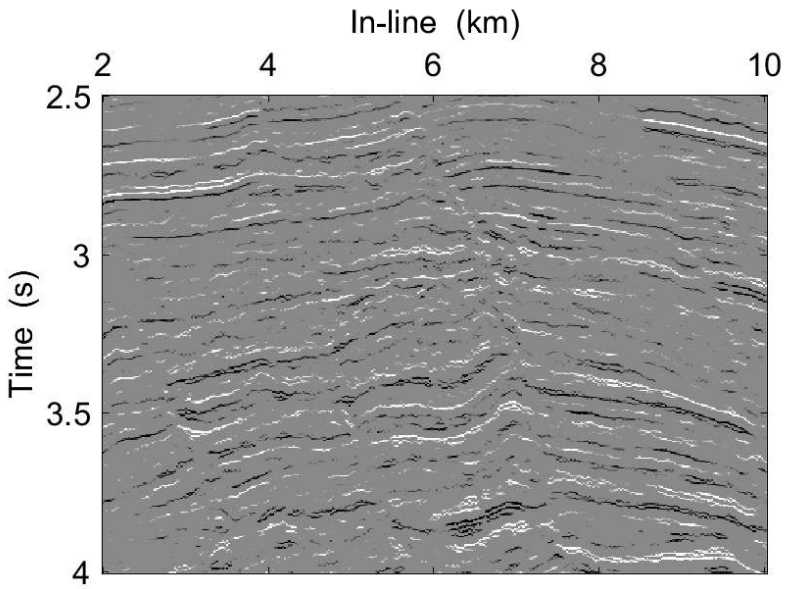


(a) single-channel

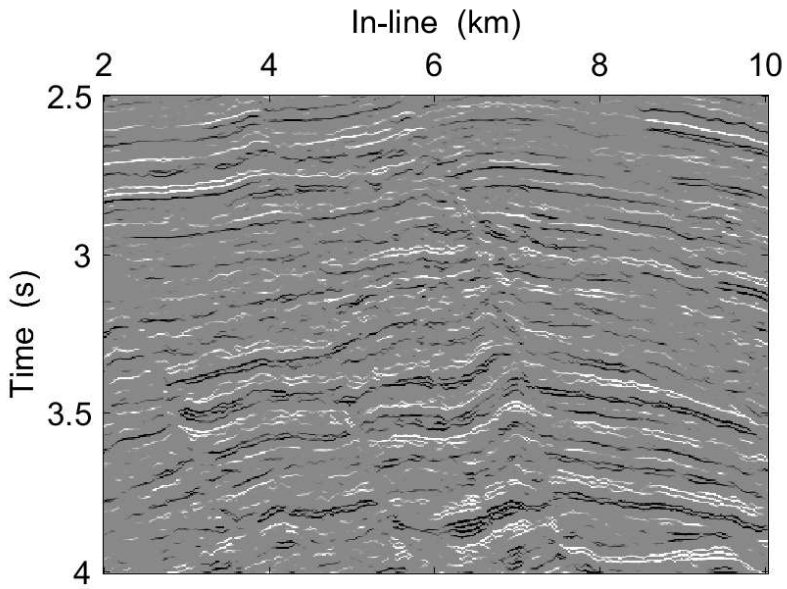


(b) multichannel

Real Data Results - zoom in

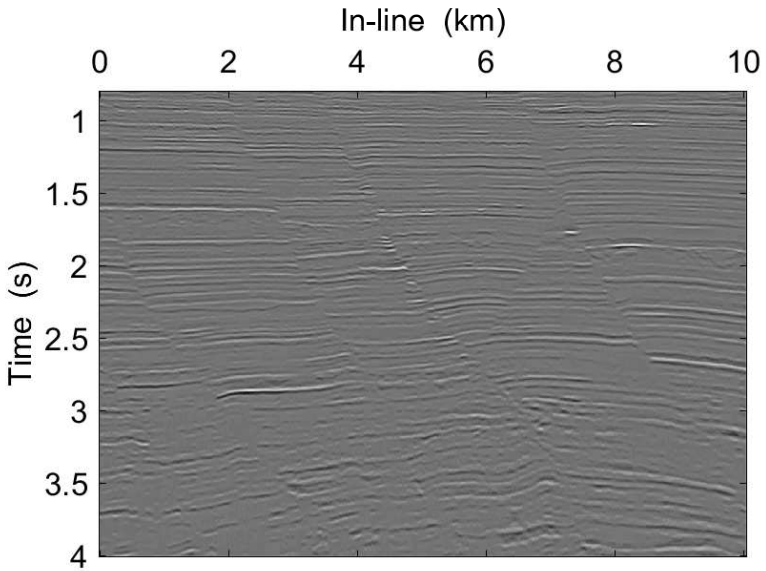


(a) single-channel

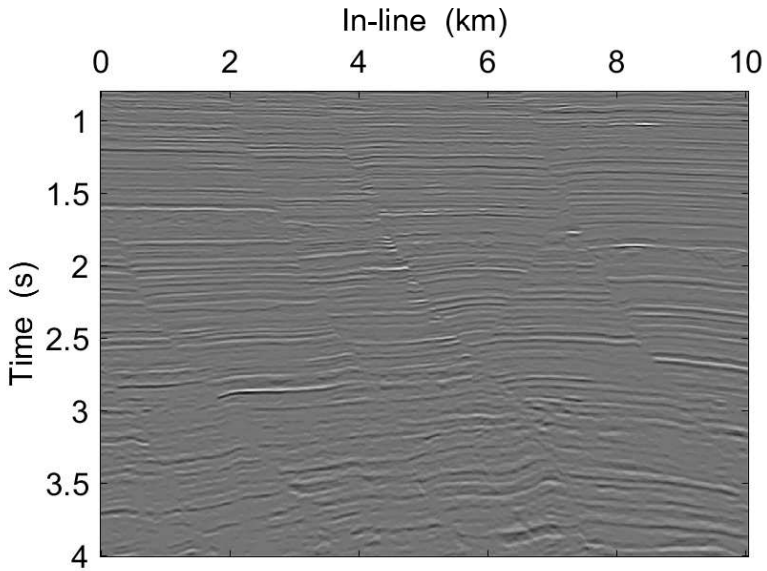


(b) multichannel

Real Data Results - Reconstructed Seismic Data



(a) single-channel, $\rho_{y,\hat{y}} = 0.88$

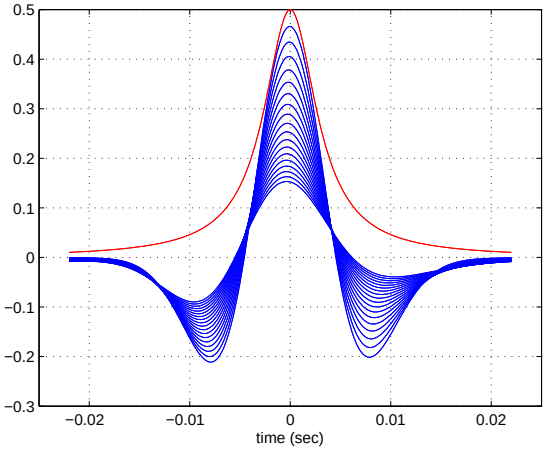


(b) multichannel, $\rho_{y,\hat{y}} = 0.90$

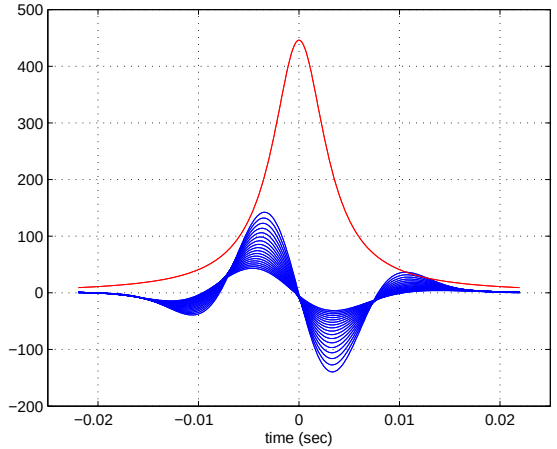
Admissible Kernel

A kernel g is **admissible** if it has the following properties:

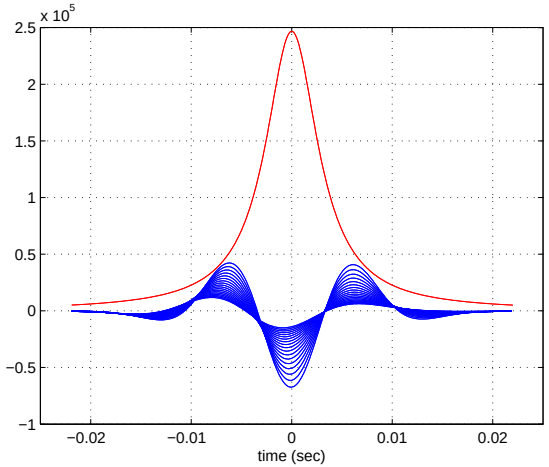
- $g \in \mathbb{R}$ is real and even.
- Global Property: There exist constants $C_l > 0$, $l = 0, 1, 2, 3$, such that $\left|g^{(l)}(t)\right| \leq \frac{C_l}{1+t^2}$, where $g^{(l)}(t)$ denotes the l^{th} derivative of g.



(a) $g_m(t - t_m)$;



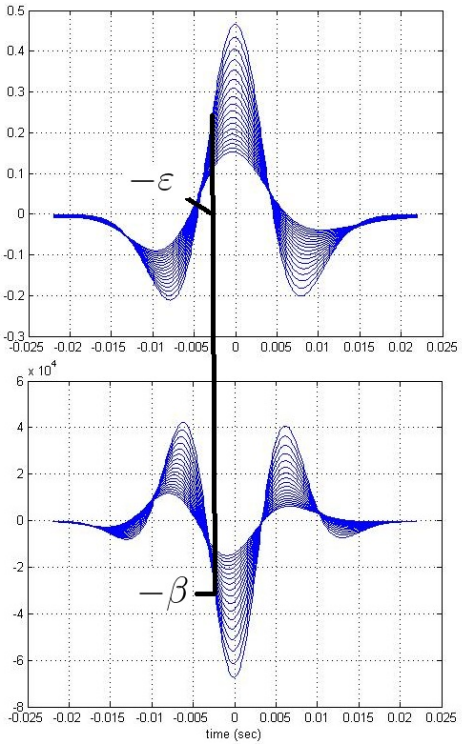
(b) $g_m^{(1)}(t - t_m)$;



(c) $g_m^{(2)}(t - t_m)$

Admissible Kernel

- Local Property: There exist constants $\varepsilon, \beta > 0$ such that
  1. $g(t) > 0$ for all $|t| \leq \varepsilon$ and $g(\varepsilon) > g(t)$ for all $|t| \geq \varepsilon$.
  2. $g^{(2)}(t) < -\beta$ for all $|t| \leq \varepsilon$.



(ε, β) determine the concavity of the pulse near the origin.

The minimal separation condition

A set of points $K \subset \mathbb{Z}$ is said to satisfy

the minimal separation condition

for a kernel dependent $\nu > 0$ and a given scaling $\sigma > 0$ and a sampling spacing $T_s = 1/F_s > 0$ if

$$\min_{k_i, k_j \in K, i \neq j} |k_i - k_j| \geq \frac{\nu\sigma}{T_s}.$$

ν is the "separation constant".

$\nu\sigma$ is the smallest time interval between two reflectors for which we can still recover two separated spikes.

Recovery Error Bound

Assume N seismic traces corresponding to N identical reflectivity channels

$$y_i[k] = \sum_n x[n] g_{\sigma,n}[k-n] + w_i[k], \quad E\|w_i\|_2^2 \leq S_{\mathrm{w}}^2, \quad i = 1, 2 \ldots N,$$

If K satisfies the separation condition then the solution $\hat{x}$ of

$$\begin{array}{ll} \min & \|\hat{\mathbf{x}}\|_1 \\ \text{subject to} & \sum_{i=1}^{N} \|\mathbf{y}_i - G\hat{\mathbf{x}}\|_2, \leq N S_{\mathrm{w}}, \end{array}$$

where satisfies

$$E\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq 64 \frac{\rho}{\beta^2} \left(\frac{\alpha_0}{\gamma_0}\right)^2 \frac{S_{\mathrm{w}}^2}{N}.$$

$$\rho \triangleq \max\left\{\frac{\gamma_0}{\varepsilon^2}, (F_s\sigma)^2\alpha_0\right\}$$

Recovery Error Bound

In the time-variant setting most cases comply with

$\frac{\gamma_0}{\varepsilon^2} < (F_s\sigma)^2\alpha_0$.

Then, the recovery error is bounded by

$$E||\hat{x} - x||_2^2 \leq 64\frac{L_r(F_s\sigma)^2}{\beta^2}\left(\frac{\alpha_0}{\gamma_0}\right)^2\frac{S_w^2}{N}.$$

*

Recovery Error Bound

$$\log(\|x - \hat{x}\|_2^2)$$

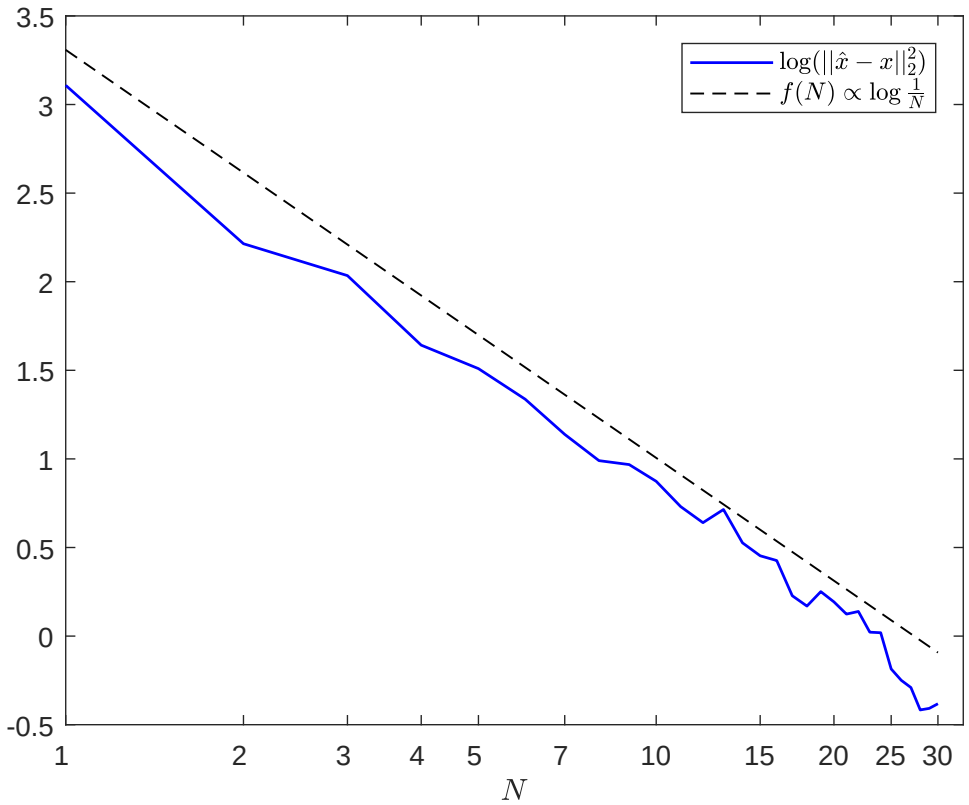


Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

Atomic Decomposition

Suppose we have a discrete dictionary of M waveforms $\{\varphi_i\}_{i=1}^{M}$ (atoms), and we collect all these waveforms as columns of a dictionary matrix $\mathbf{D}$.

The decomposition problem can be written as

$$\mathbf{y} = \sum_{i=1}^{M} \varphi_i x_i$$

In matrix-vector form

$$\mathbf{y} = \mathbf{D}\mathbf{x}$$

Finding $\mathbf{x}$ is decomposing the signal $\mathbf{y}$ into an optimal superposition of dictionary elements.

Sparse Model

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{n}$$

Recover sparsest $\mathbf{x}$ attempts at solving:

$$\min \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - D\mathbf{x}\|_2^2 < \varepsilon$$

which is relaxed into:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

The Basis Pursuit Problem

Statistical point of view

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

Solve MAP problem where

$$p(\mathbf{x}) = \exp^{-\lambda \|\mathbf{x}\|_1}$$

Atomic Decomposition - Why?

Our Data is Structured

"We are surrounded by sources of information. Each of these sources have an internal structure, which can be exploited. This structure, when identified, is the engine behind the ability to process data."

(Prof. Michael Elad, Sparse Modeling in Image Processing and Deep Learning presentation, 2018)

- We meet this problem in many fields such as
  - Compressed Sensing
  - Image Processing
  - Radar
  - Ultrasound
  - Seismology
  - And more...

Iterative-Soft Thresholding Algorithm (ISTA)

Consider the cost function

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

We build a *surrogate* function

$$Q(\mathbf{x}, \mathbf{x}_k) = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 + \frac{c}{2}\|\mathbf{x} - \mathbf{x}_k\|_2^2 - \frac{1}{2}\|\mathbf{D}\mathbf{x} - \mathbf{D}\mathbf{x}_k\|_2^2$$

... the parameter c is chosen such that the added expression $d(\mathbf{x}, \mathbf{x}_k) = Q(\mathbf{x}, \mathbf{x}_k) - f(\mathbf{x})$ is strictly convex, requiring its Hessian to be positive definite, $c\mathbf{I} - \mathbf{D}^T\mathbf{D} \prec \mathbf{0}$. Therefore $c > \|\mathbf{D}^T\mathbf{D}\|_2 = \lambda_{max}(\mathbf{D}^T\mathbf{D})$.

*Daubechies et al. 2004.

Iterative-Soft Thresholding Algorithm (ISTA)

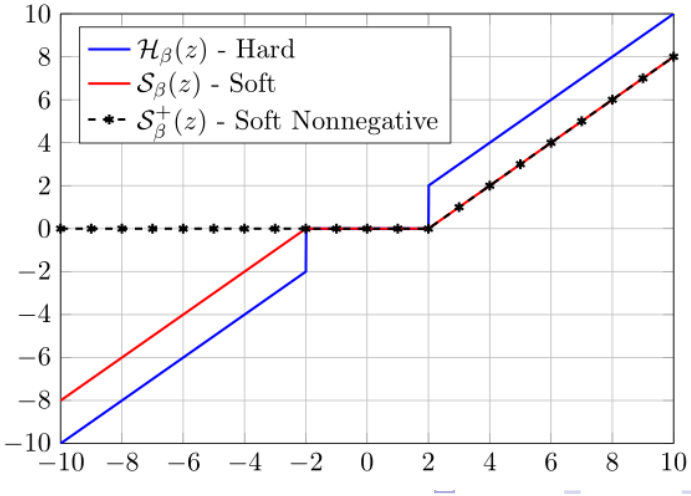
Following Majorization Minimization strategy, i.e., minimizing $Q(\mathbf{x}, \mathbf{x}_k)$ instead of $f(\mathbf{x})$ The sequence of solution generated by the recurrent formula

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} Q(\mathbf{x}, \mathbf{x}_k)$$

We can manage to find closed-form solution for its global minimizer

$$\mathbf{x}_{k+1} = \mathcal{S}_{\frac{\lambda}{c}}\left(\frac{1}{c}\mathbf{D}^T(\mathbf{y} - \mathbf{D}\mathbf{x}_k) + \mathbf{x}_k\right)$$

$$\mathcal{S}_\beta(z) = \begin{cases} z + \beta, & z < -\beta \\ 0, & -\beta \leq z \leq \beta \\ z - \beta, & z > \beta \end{cases}.$$



Iterative-Soft Thresholding Algorithm (ISTA)

Let's look at the solution again

$$\mathbf{x}_{k+1} = \mathcal{S}_{\frac{\lambda}{c}}\left(\frac{1}{c}\overbrace{\mathbf{D}^T}^{\text{project on dictionary}}\underbrace{(\mathbf{y}-\mathbf{D}\mathbf{x}_k)}_{\text{residual term}} + \underbrace{\mathbf{x}_k}_{\text{add to current solution}}\right)$$

Iterative-Soft Thresholding Algorithm (ISTA)

Note that the rectified linear units (ReLU) activation function satisfies

$$\mathrm{ReLU}(z-\beta) = \mathcal{S}_{\beta}^{+}(z) = \begin{cases} 0, & z \leq \beta \\ z-\beta, & z > \beta \end{cases}.$$

A soft-threshold solution can be written as

$$\begin{aligned} \mathbf{X} &= \mathcal{S}_{\beta}^{+}(\mathbf{D}^T\mathbf{Y}) - \mathcal{S}_{\beta}^{+}(-\mathbf{D}^T\mathbf{Y}) \\ &= \mathrm{ReLU}(\mathbf{D}^T\mathbf{Y}-\beta) - \mathrm{ReLU}(-\mathbf{D}^T\mathbf{Y}-\beta) \end{aligned}$$

$\Rightarrow$ ISTA can be viewed as an *Recurrent Neural Net* (RNN)

unfolded through time.

Extensions: Fast-ISTA (FISTA), Learned-ISTA (LISTA) - Lecun et al. 2010.

Current Challenges

- High dictionary mutual coherence

$$\mu(\mathbf{D}) = \max_{i \neq j} \frac{\left| \mathbf{d}_i^T \mathbf{d}_j \right|}{\|\mathbf{d}_i\|_2 \cdot \|\mathbf{d}_j\|_2},$$

For example, in the seismic scenario usually $0.5 < \mu(\mathbf{G}) < 1$.

Current Challenges

- High dictionary mutual coherence

$$\mu(\mathbf{D}) = \max_{i \neq j} \frac{\left| \mathbf{d}_i^T \mathbf{d}_j \right|}{\|\mathbf{d}_i\|_2 \cdot \|\mathbf{d}_j\|_2},$$

  For example, in the seismic scenario usually $0.5 < \mu(\mathbf{G}) < 1$.

- The signal (or even a signal stripe (or a patch)) is not sufficiently sparse for a successful recovery (see Papyan et al. 2017).

Current Challenges

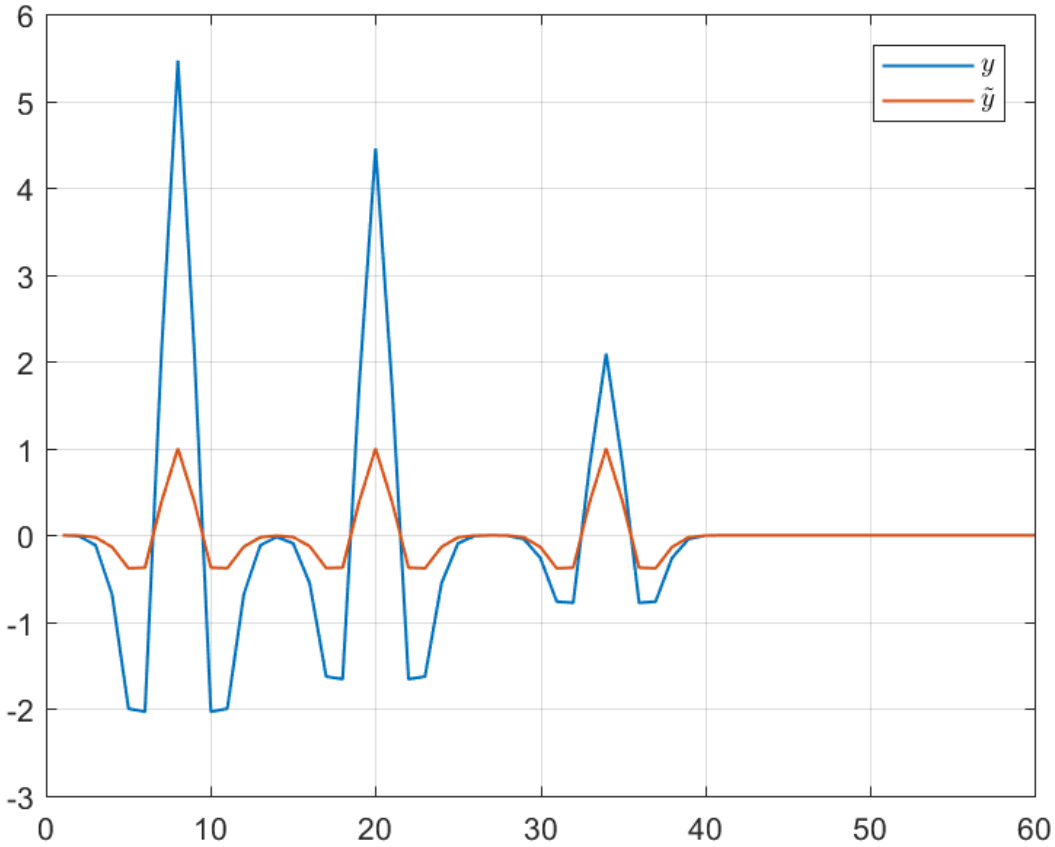
- The ratio between the global maximal and minimal $\mathbf{x}$ values in absolute value $\frac{|\mathbf{x}|_{\max}}{|\mathbf{x}|_{\min}}$ is high.

Current Challenges

- The ratio between the global maximal and minimal $\mathbf{x}$ values in absolute value $\frac{|\mathbf{x}|_{\max}}{|\mathbf{x}|_{\min}}$ is high.

- Recovering the sparse code takes too much time and cannot be applied to real-time applications.

Receptive Field Normalization



Receptive Field Normalization

$$\sigma_{\mathrm{y}}[k] \triangleq \left(\sum_{n=-\frac{L_{\mathrm{h}}-1}{2}}^{\frac{L_{\mathrm{h}}-1}{2}} h[n] y^2[k-n] \right)^{\frac{1}{2}} = \sqrt{h[k] * y^2[k]}$$

$h[k]$ is a receptive field normalization kernel such that:

1. The kernel is positive: $h[k] \geq 0 \quad \forall k$.
2. The kernel is symmetric: $h[k] = h[-k] \quad \forall k$.
3. The kernel's global maximum is at its center: $h[0] = 1 \geq h[k] \quad \forall k \neq 0$.
4. The kernel's energy is finite: $\sum_k h[k] < \infty$.

Step 1 - Support Estimation

The detected support at the first iteration is

$$\mathbf{q}_1 = \mathcal{I}_{\beta_1}(\mathbf{W}_{\mathrm{D}}\mathbf{D}^T\mathbf{W}_0\mathbf{y})$$

Step 1 - Support Estimation

The detected support at the first iteration is

$$\mathbf{q}_1 = \mathcal{I}_{\beta_1}(\mathbf{W}_{\mathrm{D}}\mathbf{D}^T\mathbf{W}_0\mathbf{y})$$

$\mathbf{W}_{\mathrm{D}} = \mathrm{diag}\left(\sigma_{\mathrm{d}_i}^{-1}\right), i = 1, ..., mL_{\mathrm{x}}$. where $\sigma_{\mathrm{d}_i} = \|\mathbf{d}_i\|_2$

Step 1 - Support Estimation

The detected support at the first iteration is

$$\mathbf{q}_1 = \mathcal{I}_{\beta_1}(\mathbf{W}_{\mathrm{D}}\mathbf{D}^T\mathbf{W}_0\mathbf{y})$$

$\mathbf{W}_{\mathrm{D}} = \mathrm{diag}\left(\sigma_{\mathrm{d}_i}^{-1}\right), i = 1, ..., mL_{\mathrm{x}}$. where $\sigma_{\mathrm{d}_i} = \|\mathbf{d}_i\|_2$

$$\tilde{\sigma}_{\mathrm{y}}[k] = \begin{cases} \sigma_{\mathrm{y}}[k] & \left|\sigma_{\mathrm{y}}[k]\right| \geq \tau_1 \\ 1 & \left|\sigma_{\mathrm{y}}[k]\right| < \tau_1 \end{cases},$$

$\mathbf{W}_0 = \mathrm{diag}\left(\tilde{\sigma}_{\mathrm{y}}^{-1}[k]\right), k = 1, ..., L_{\mathrm{y}}$

Step 1 - Support Estimation

The detected support at the first iteration is

$$\mathbf{q}_1 = \mathcal{I}_{\beta_1}(\mathbf{W}_{\mathrm{D}}\mathbf{D}^T\mathbf{W}_0\mathbf{y})$$

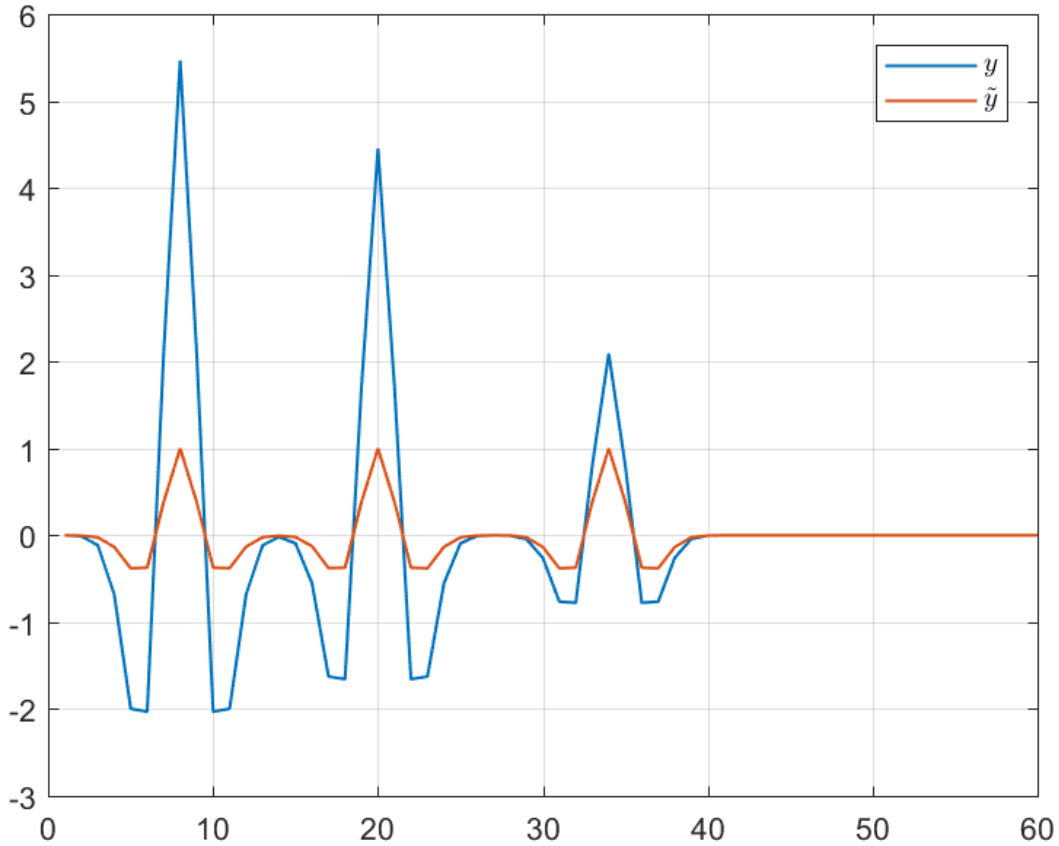
$\mathbf{W}_{\mathrm{D}} = \mathrm{diag}\left(\sigma_{\mathrm{d}_i}^{-1}\right), i = 1, ..., mL_{\mathrm{x}}$. where $\sigma_{\mathrm{d}_i} = \|\mathbf{d}_i\|_2$

$$\tilde{\sigma}_{\mathrm{y}}[k] = \begin{cases} \sigma_{\mathrm{y}}[k] & \left|\sigma_{\mathrm{y}}[k]\right| \geq \tau_1 \\ 1 & \left|\sigma_{\mathrm{y}}[k]\right| < \tau_1 \end{cases},$$

$$\mathbf{W}_0 = \mathrm{diag}\left(\tilde{\sigma}_{\mathrm{y}}^{-1}[k]\right), k = 1, ..., L_{\mathrm{y}}$$

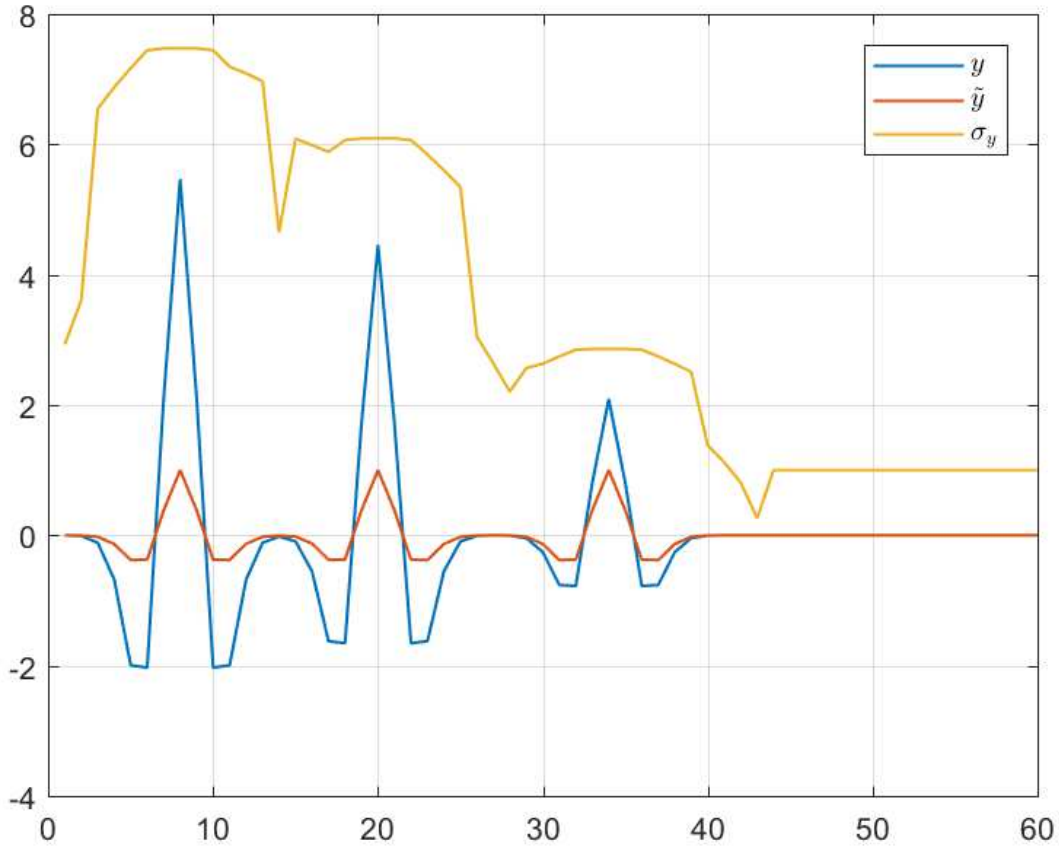
$$\mathcal{I}_{\beta_1}(x_k) = \begin{cases} 1 & |x_k| \geq \beta_1 \\ 0 & |x_k| < \beta_1 \end{cases}$$

Receptive Field Normalization



$$\hat{\mathbf{y}} = \tilde{\mathbf{W}}_{\mathrm{D}} \mathbf{W}_0 \mathbf{y}$$

Receptive Field Normalization



$$\hat{\mathbf{y}} = \tilde{\mathbf{W}}_{\mathrm{D}}\mathbf{W}_0\mathbf{y}$$

Step 2 - Amplitude Estimation

- Solve the subsystem $\mathbf{y} = \mathbf{D}_{K_1}\mathbf{x}_{K_1}$,

  $\mathbf{D}_{K_1}$ - the partial dictionary matrix having $|K_1|$ columns from the columns of $\mathbf{D}$ with indices in K_1 - the estimated support in the first iteration.

- Or approximately: $\mathbf{x}_1 = \mathbf{q}_1 \odot \left(\mathbf{W}_{\mathrm{D}}^2 \mathbf{D}^T \mathbf{y}\right)$

input : signal $\mathbf{y}$, dictionary $\mathbf{D}$

Init: $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{\Delta q}_0 = \mathbf{0}$, $\theta = 0$

compute: $\mathbf{W}_{\mathrm{D}} = \mathrm{diag}\left(\sigma_{\mathrm{d}_i}^{-1}\right), i = 1, ..., mL_{\mathrm{x}}.$

while $\|\mathbf{x}_{\theta+1} - \mathbf{x}_{\theta}\|_2 < \delta$ *or* $\theta \leq N_{\mathrm{it}}$ **do**

$\mathbf{\Delta r}_{\theta+1} = \mathbf{y} - \mathbf{D}\mathbf{x}_{\theta}$

compute:

$$\tilde{\sigma}_{\Delta r}^{\theta+1}[k] = \begin{cases} \sigma_{\Delta r}^{\theta+1}[k] & \left|\sigma_{\Delta r}^{\theta+1}[k]\right| \geq \tau_{\theta} \\ 1 & \left|\sigma_{\Delta r}^{\theta+1}[k]\right| < \tau_{\theta} \end{cases},$$

$\mathbf{W}_{\theta} = \mathrm{diag}\left(\tilde{\sigma}_{\Delta r}^{\theta+1}[k]\right)^{-1}, k = 1, ..., L_{\mathrm{y}}$

$\mathbf{\Delta q}_{\theta+1} = \mathcal{I}_{\beta_{\theta+1}}(\mathbf{W}_{\mathrm{D}}\mathbf{D}^{T}\mathbf{W}_{\theta}\mathbf{\Delta r}_{\theta+1})$

solve the subsystem $\mathbf{\Delta r}_{\theta+1} = \mathbf{D}_{K_{\theta+1}}\mathbf{\Delta x}_{\theta+1}$,

$\mathbf{D}_{K_{\theta+1}}$ - partial dictionary matrix corresponding to $\mathbf{\Delta q}_{\theta+1}$.

$\mathbf{x}_{\theta+1} = \alpha_{\mathrm{r}}\mathbf{\Delta x}_{\theta+1} + \mathbf{x}_{\theta}$

$\theta \leftarrow \theta + 1$

end

Theoretical Analysis

Under what conditions can we

guarantee perfect support detection

within the first step

???

Support Recovery in the Presence of Noise

Define local sparsity measure

$$s = \|\mathbf{x}\|_{0,\infty}^{S_h^i} = \max_i \|\mathbf{x}_i\| = \max_i \sum_{l \in S_h^i} \mathcal{I}_0(\mathbf{x}[l]),$$

Let $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e}$, where $\mathbf{D}$ is a convolutional dictionary, with normalized atom's variance $\|\mathbf{d}_i\|_2 = \sigma_{\mathrm{d}} = 1 \; \forall i \in [1, mL_{\mathrm{x}}]$.

Assume a rectangular receptive field normalization kernel, that is $\mathbf{h}[k] = 1$, $\mathbf{h} \in \mathbb{R}^{L_{\mathrm{h}} \times 1}$, $L_{\mathrm{h}} = L_{\mathrm{d}}$. In other words $\mathbf{H}$ simply extracts a local data stripe.

Assuming that

1. $$\frac{|\mathbf{x}_i|_{\min}}{|\mathbf{x}_i|_{\max} + \frac{\varepsilon_{\mathrm{d}}}{s}} > \frac{s\mu}{1+\mu}\left(1 + \frac{\sqrt{s}}{\sqrt{1-(s-1)\mu} - \tilde{\varepsilon}_{\mathrm{d},\infty}}\right) + \frac{2s\varepsilon_{\mathrm{s}}}{1+\mu} \quad \forall i \in K.$$

   where $|\mathbf{x}_i|_{\max}$ and $|\mathbf{x}_i|_{\min}$ are the highest and the lowest entries in a stripe $\mathbf{x}_i$ respectively, $\varepsilon_{\mathrm{d}} \triangleq \max_i \|\mathbf{e}_i\|_2$, $\tilde{\varepsilon}_{\mathrm{d},\infty} \triangleq \frac{\varepsilon_{\mathrm{d}}}{|\mathbf{x}|_{\min}}$ and $\varepsilon_{\mathrm{s}} \triangleq \frac{\varepsilon_{\mathrm{d}}}{\tau}$.

Assuming that

1. $$\frac{|\mathbf{x}_i|_{\min}}{|\mathbf{x}_i|_{\max}+\frac{\varepsilon_{\mathrm{d}}}{s}} > \frac{s\mu}{1+\mu}\left(1+\frac{\sqrt{s}}{\sqrt{1-(s-1)\mu}-\tilde{\varepsilon}_{\mathrm{d},\infty}}\right)+\frac{2s\varepsilon_{\mathrm{s}}}{1+\mu} \quad \forall i \in K.$$

   where $|\mathbf{x}_i|_{\max}$ and $|\mathbf{x}_i|_{\min}$ are the highest and the lowest entries in a stripe $\mathbf{x}_i$ respectively, $\varepsilon_{\mathrm{d}} \triangleq \max_i \|\mathbf{e}_i\|_2$, $\tilde{\varepsilon}_{\mathrm{d},\infty} \triangleq \frac{\varepsilon_{\mathrm{d}}}{|\mathbf{x}|_{\min}}$ and $\varepsilon_{\mathrm{s}} \triangleq \frac{\varepsilon_{\mathrm{d}}}{\tau}$.

2. The threshold β_1 is chosen according to

$$\frac{(1+\mu)|\mathbf{x}_i|_{\min}}{s|\mathbf{x}_i|_{\max}+\varepsilon_{\mathrm{d}}}-\mu-\varepsilon_{\mathrm{s}} > \beta_1 > \frac{\sqrt{s}\mu}{\sqrt{1-(s-1)\mu}}+\varepsilon_{\mathrm{s}}.$$

Assuming that

1. $$\frac{|\mathbf{x}_i|_{\min}}{|\mathbf{x}_i|_{\max}+\frac{\varepsilon_{\mathrm{d}}}{s}} > \frac{s\mu}{1+\mu}\left(1+\frac{\sqrt{s}}{\sqrt{1-(s-1)\mu}-\tilde{\varepsilon}_{\mathrm{d},\infty}}\right)+\frac{2s\varepsilon_{\mathrm{s}}}{1+\mu} \quad \forall i \in K.$$

   where $|\mathbf{x}_i|_{\max}$ and $|\mathbf{x}_i|_{\min}$ are the highest and the lowest entries in a stripe $\mathbf{x}_i$ respectively, $\varepsilon_{\mathrm{d}} \triangleq \max_i \|\mathbf{e}_i\|_2$, $\tilde{\varepsilon}_{\mathrm{d},\infty} \triangleq \frac{\varepsilon_{\mathrm{d}}}{|\mathbf{x}|_{\min}}$ and $\varepsilon_{\mathrm{s}} \triangleq \frac{\varepsilon_{\mathrm{d}}}{\tau}$.

2. The threshold β_1 is chosen according to

$$\frac{(1+\mu)|\mathbf{x}_i|_{\min}}{s|\mathbf{x}_i|_{\max}+\varepsilon_{\mathrm{d}}}-\mu-\varepsilon_{\mathrm{s}} > \beta_1 > \frac{\sqrt{s}\mu}{\sqrt{1-(s-1)\mu}}+\varepsilon_{\mathrm{s}}.$$

then the support of $\mathbf{x}$ is perfectly recovered with Receptive-Field Normalization thresholding,

$$\mathrm{Supp}(\mathbf{x}_1) = \mathrm{Supp}(\mathbf{x}),$$

where $\mathrm{Supp}(\cdot)$ is the support of a vector, and $\mathbf{x}_1$ is the recovered sparse vector within one step of RFN thresholding (or one iteration of Algorithms 1-4).

Support Recovery under Sufficient Separation

Stable recovery of *Hard Thresholding* in the noiseless case is acquired as long as

$$\min_{i \in K} |\mathbf{d}_i^T \mathbf{y}| > \max_{j \notin K} |\mathbf{d}_j^T \mathbf{y}|.$$

Leading to

$$\frac{|\mathbf{x}|_{\min}}{|\mathbf{x}|_{\max}} > (2s-1)\mu.$$

Suppose $s = 1$, meaning that there can only be one atom in each stripe, than we have

$$\frac{|\mathbf{x}|_{\min}}{|\mathbf{x}|_{\max}} > \mu.$$

Practically, in many convolutional models, if the stride is not large enough, this condition is not be met.

Support Recovery under Sufficient Separation

Perfect recovery in the noiseless case where $s = 1$ *with RFN* is guaranteed for any ratio

$$0 < \frac{|\mathbf{x}|_{\min}}{|\mathbf{x}|_{\max}} \leq 1.$$

Support Recovery with Attenuating RFN kernel in the Noiseless case

Assume a symmetric strictly decreasing receptive field normalization kernel, that is $h[|k|] > h[|k+1|]$ (e.g., a truncated Gaussian window), of length $L_{\mathrm{h}} = L_{\mathrm{d}}$, such that $\mu(\mathbf{HD}_i) << \mu(\mathbf{D}),\ \forall i \in [1, L_{\mathrm{y}}]$.

Support Recovery with Attenuating RFN kernel in the Noiseless case

Assume a symmetric strictly decreasing receptive field normalization kernel, that is $h[|k|] > h[|k+1|]$ (e.g., a truncated Gaussian window), of length $L_{\mathrm{h}} = L_{\mathrm{d}}$, such that $\mu(\mathbf{HD}_i) << \mu(\mathbf{D}),\ \forall i \in [1, L_{\mathrm{y}}]$. Assuming that

1. K obeys the minimal separation condition with a separation constant ν.

Support Recovery with Attenuating RFN kernel in the Noiseless case

Assume a symmetric strictly decreasing receptive field normalization kernel, that is $h[|k|] > h[|k+1|]$ (e.g., a truncated Gaussian window), of length $L_{\mathrm{h}} = L_{\mathrm{d}}$, such that $\mu(\mathbf{HD}_i) << \mu(\mathbf{D}),\ \forall i \in [1, L_{\mathrm{y}}]$. Assuming that

1. K obeys the minimal separation condition with a separation constant ν.
2. $$\frac{\sqrt{s}\mu}{h_{\mathrm{d,min}}} < \beta_1 < 1 - \left(h_{\mathrm{d}}(\nu) + \mu\right) \frac{\|\mathbf{x}_{-i}\|_1}{|\mathbf{x}[i]| + h_{\mathrm{d}}(\nu)\|\mathbf{x}_{-i}\|_1} \quad \forall i \in K,$$

   where $h_{\mathrm{d}}(\nu) \triangleq \max_{p\in[1,m]} \mathbf{H}_{\mathrm{d}}[\Delta_k + (p-1)L_{\mathrm{s}}, \Delta_k + (p-1)L_{\mathrm{s}}]$, such that $\mathbf{H}_{\mathrm{d}}[k,k] = \|\mathbf{H}\mathbf{d}_k\|_2^2,\ k \in S_j^h$.
   $h_{\mathrm{d,min}} \triangleq \min_k \mathbf{H}_{\mathrm{d}}[k,k]$, and $\|\mathbf{x}_{-i}\|_1 \triangleq \sum_{t\in S_h^i, t\neq i} |\mathbf{x}[t]| = \|\mathbf{x}_i\|_1 - |\mathbf{x}[i]|$ is the sum of weights affecting the ith stripe, besides the spike at location i.

Support Recovery with Attenuating RFN kernel in the Noiseless case

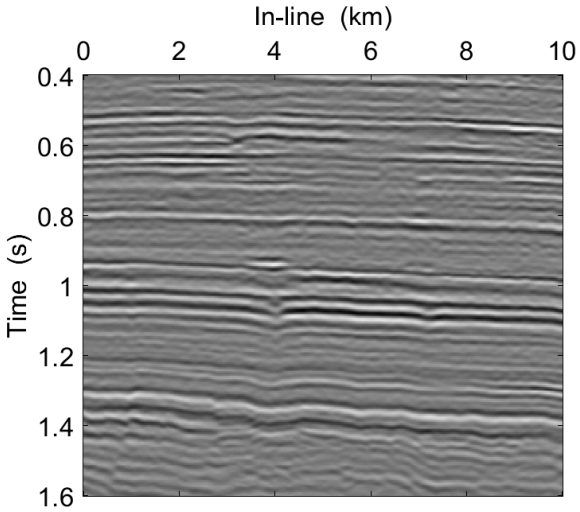
Assume a symmetric strictly decreasing receptive field normalization kernel, that is $h[|k|] > h[|k+1|]$ (e.g., a truncated Gaussian window), of length $L_{\mathrm{h}} = L_{\mathrm{d}}$, such that $\mu(\mathbf{HD}_i) << \mu(\mathbf{D}),\ \forall i \in [1, L_{\mathrm{y}}]$. Assuming that

1. K obeys the minimal separation condition with a separation constant ν.
2. $$\frac{\sqrt{s}\mu}{h_{\mathrm{d,min}}} < \beta_1 < 1 - \left(h_{\mathrm{d}}(\nu) + \mu\right)\frac{\|\mathbf{x}_{-i}\|_1}{|\mathbf{x}[i]| + h_{\mathrm{d}}(\nu)\|\mathbf{x}_{-i}\|_1} \quad \forall i \in K,$$

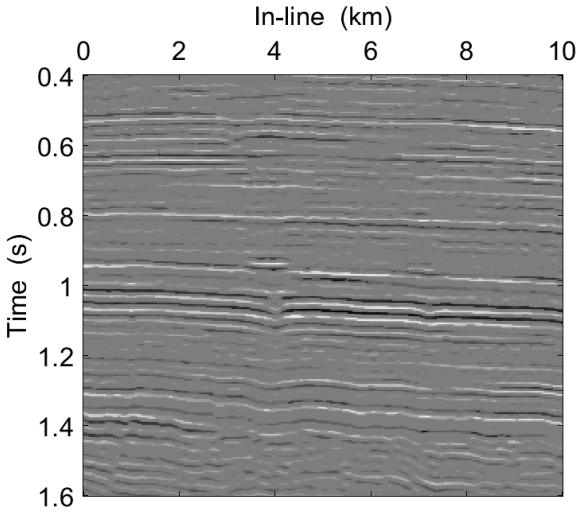
   where $h_{\mathrm{d}}(\nu) \triangleq \max_{p\in[1,m]} \mathbf{H}_{\mathrm{d}}[\Delta_k + (p-1)L_{\mathrm{s}}, \Delta_k + (p-1)L_{\mathrm{s}}]$, such that $\mathbf{H}_{\mathrm{d}}[k,k] = \|\mathbf{H}\mathbf{d}_k\|_2^2,\ k \in S_j^h$.
   $h_{\mathrm{d,min}} \triangleq \min_k \mathbf{H}_{\mathrm{d}}[k,k]$, and $\|\mathbf{x}_{-i}\|_1 \triangleq \sum_{t\in S_h^i, t\neq i} |\mathbf{x}[t]| = \|\mathbf{x}_i\|_1 - |\mathbf{x}[i]|$ is the sum of weights affecting the ith stripe, besides the spike at location i.

then the support of $\mathbf{x}$ is perfectly recovered with Receptive-Field Normalization thresholding, $\mathrm{Supp}(\mathbf{x}_1) = \mathrm{Supp}(\mathbf{x})$.

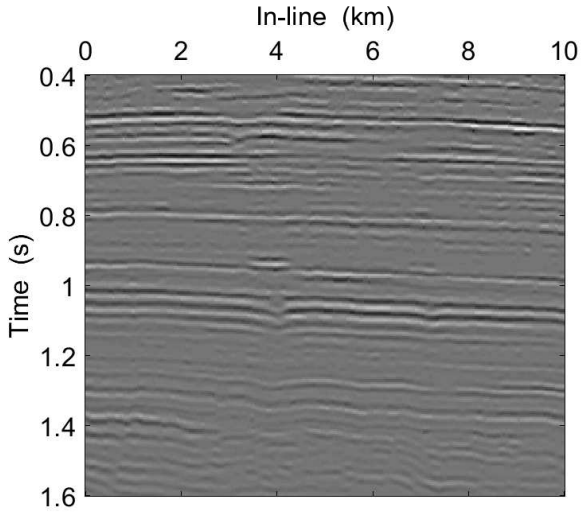
Real Data Results



(a)



(b)

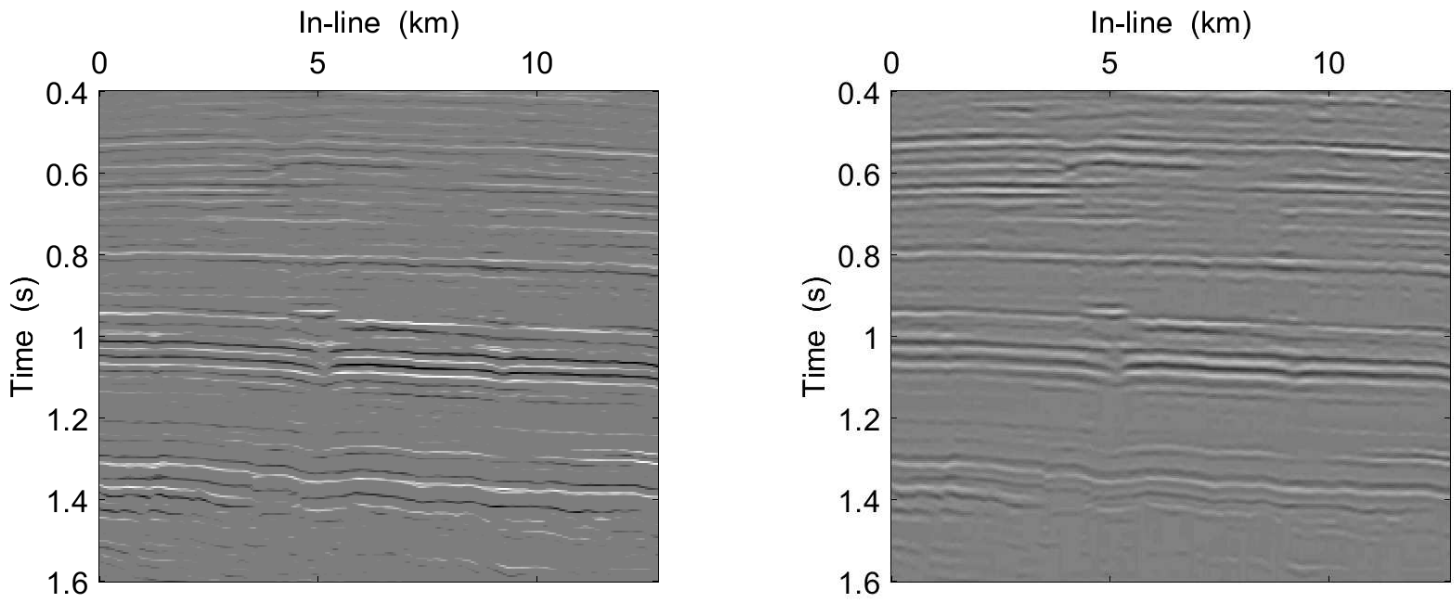


(c)

Real data inversion results:

(a) Seismic data; (b) Estimated reflectivity with $N_{\mathrm{it}} = 2$ maximum number of iterations per trace; (c) Reconstructed seismic data, $\rho_{\mathbf{S},\hat{\mathbf{S}}} = 0.89$.

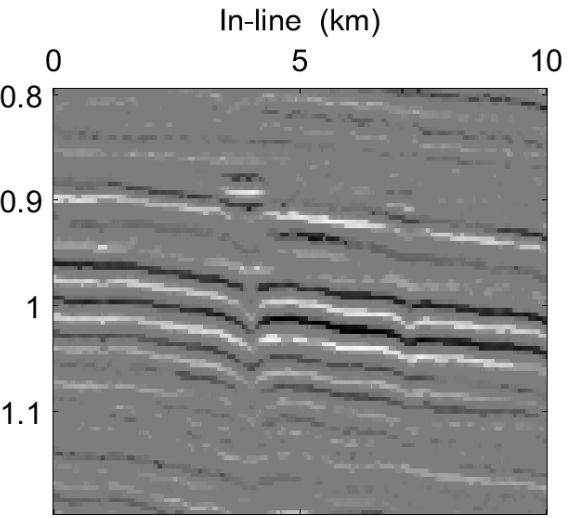
Real Data Results



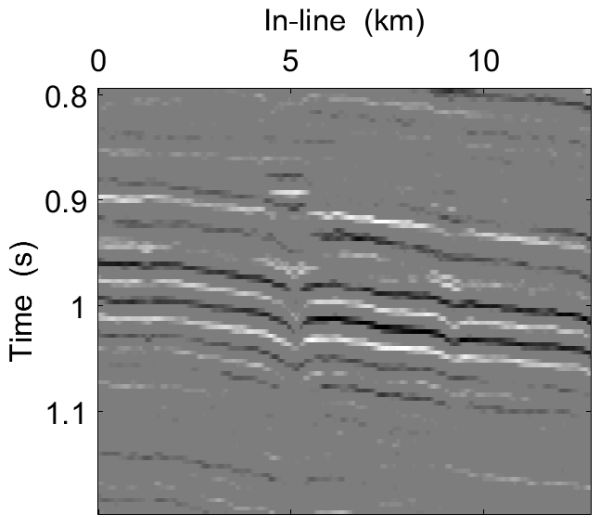
(a) (b)

Real data inversion results using ISTA: (a) Estimated reflectivity with $M_{\text{it}} = 1087$ average number of iterations per trace; (b) Reconstructed seismic data, $\rho_{\mathbf{S},\hat{\mathbf{S}}} = 0.91$.

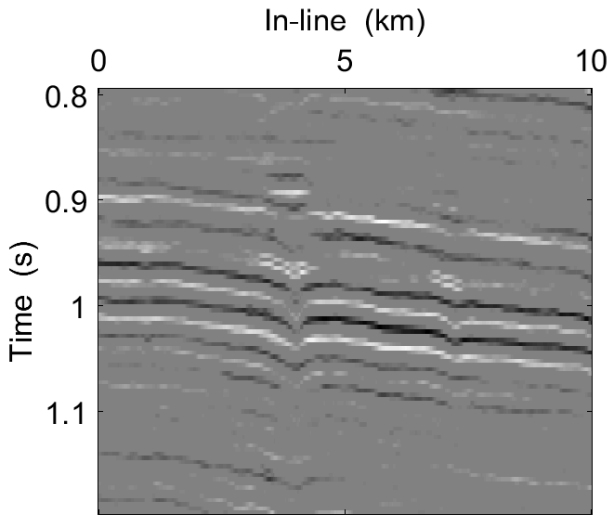
Real Data Results



(a)



(b)

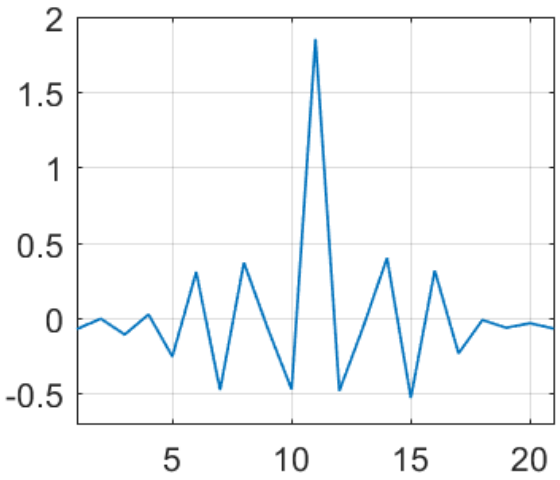


(c)

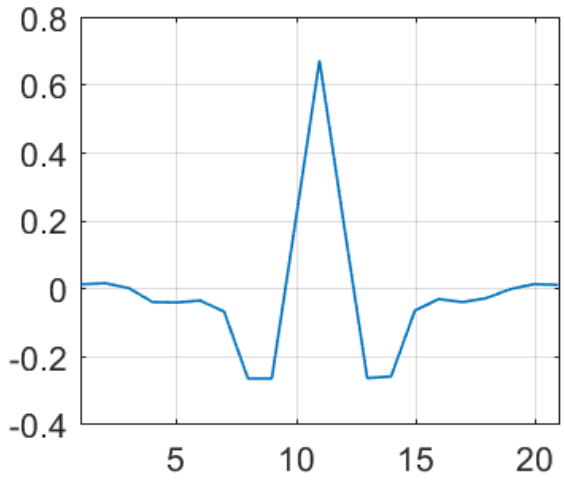
Zoom into the predicted reflectivities:

(a) RFN-ITA; (b) ISTA; (c) Multichannel time-variant deconvolution.

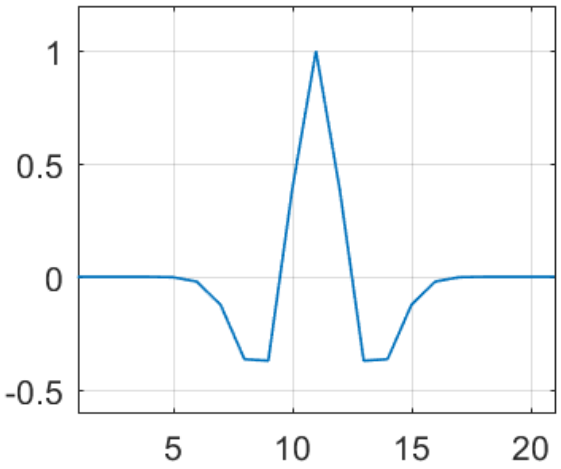
Learned RFN-ITA



(a)



(b)



(c)

Learned RFN-ITA parameters:

(a) $\mathbf{D}_1$ kernel; (b) $\mathbf{D}_2$ kernel; (c) 80π Ricker wavelet ($F_\mathrm{s} = 250\mathrm{Hz}$).

Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

Reminder

RNN learns a mapping from the known input-output patterns $f : \mathbf{x} \rightarrow \mathbf{y}$. Suppose a given observation sequence $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t)$, and an expected output vector $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t)$. At time t, every neuron receives both the input vector $\mathbf{x}_t$ and the previous output $\mathbf{y}_{t-1}$, which can formulated as

$$\mathbf{y}_t = \phi(\mathbf{W}_{xy}^T \mathbf{x}_t + \mathbf{W}_{yy}^T \mathbf{y}_{t-1} + \mathbf{b}),$$

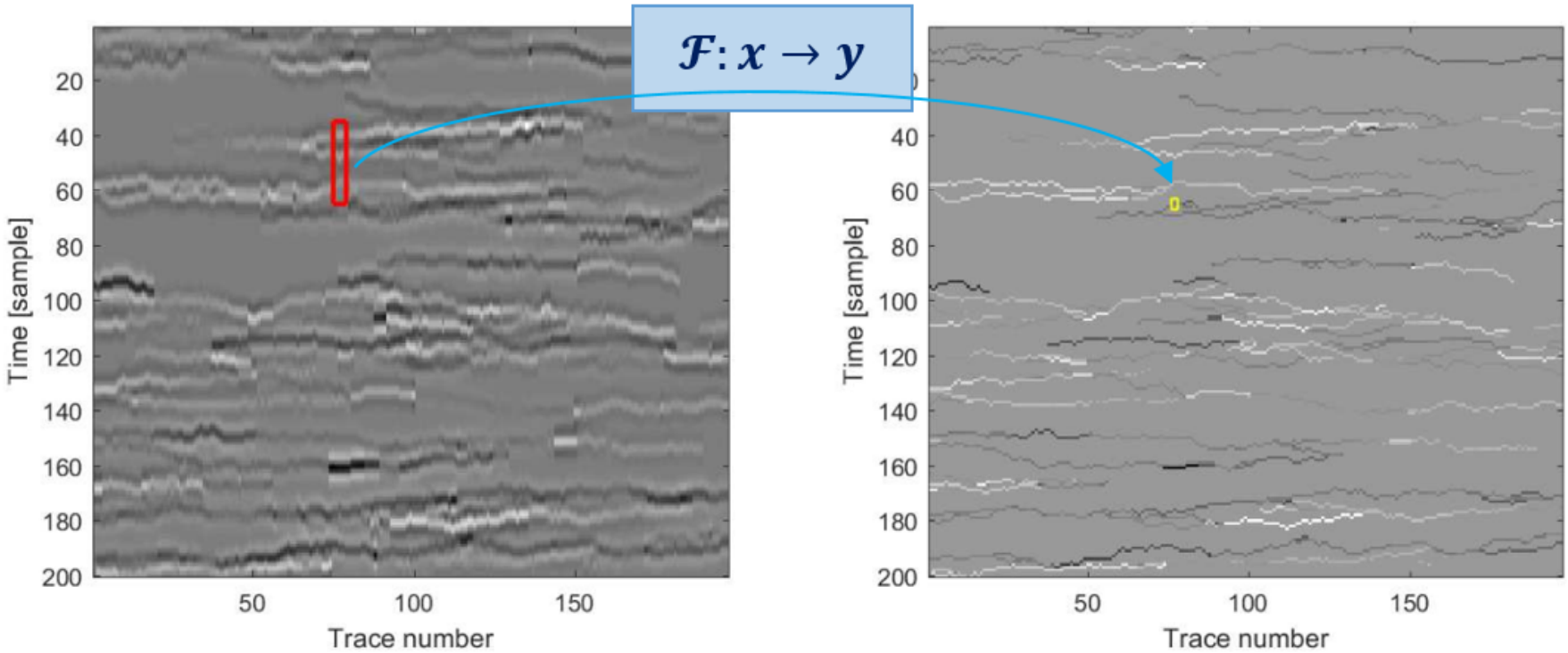
where ϕ is the activation function, $\mathbf{W}_{xy}$ and $\mathbf{W}_{yy}$ are the weight matrices, and $\mathbf{b}$ is the bias vector.

Sparse seismic deconvolution using RNN

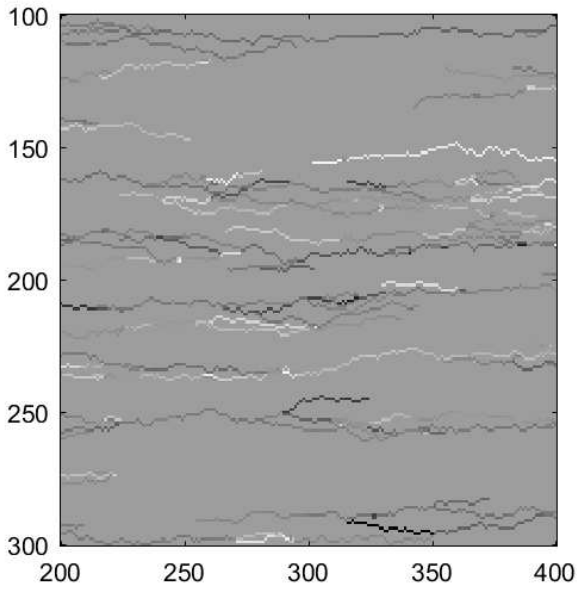
Why RNN?

- RNN remembers both in space and time:
  - Signal travels in both directions.
  - Output of current state depends on output of previous states.
  - Estimation of velocity at each point considers neighboring points.
- Hypothesis: Data in same feature space and have the same distribution.

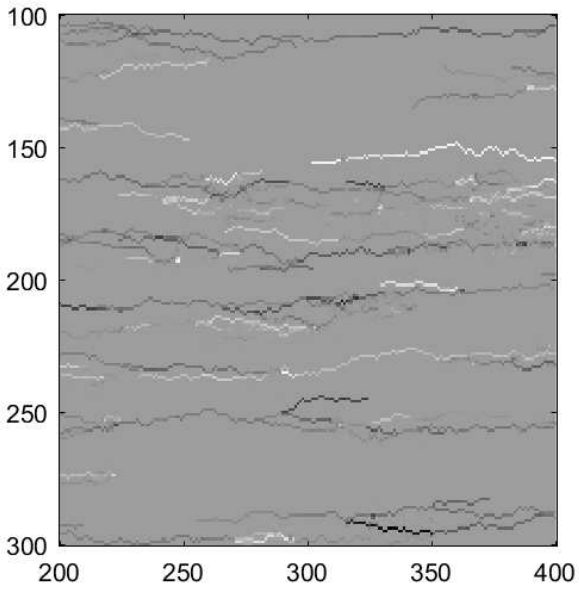
Deconvolution using RNN



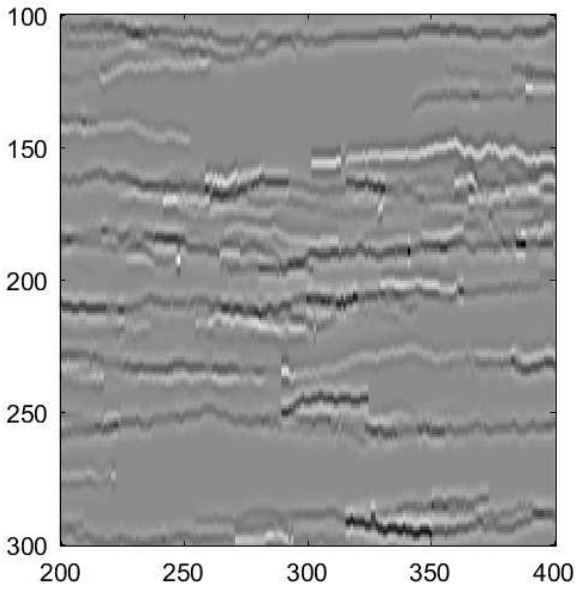
Synthetic Experiments



(a)



(b)



(c)

(a) Synthetic 2D reflectivity; (b) Estimated 2D reflectivity;

(c) 2D noise free seismic data.

Real Data Seismic Deconvolution

Real data is of inherent uncertainty

How does one train a neural network

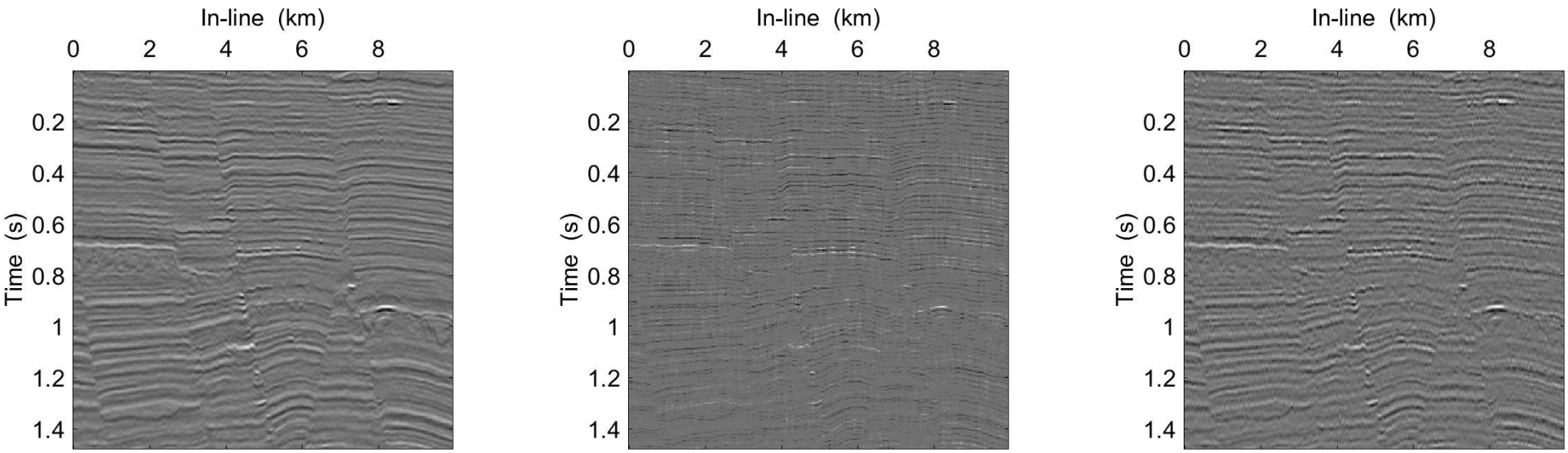
to deconvolve real seismic data

???

Real Data Seismic Deconvolution

- Train with real data deconvolved by other algorithm
- Train with synthetic data of complex structure
- Train with synthetic data of simple structure

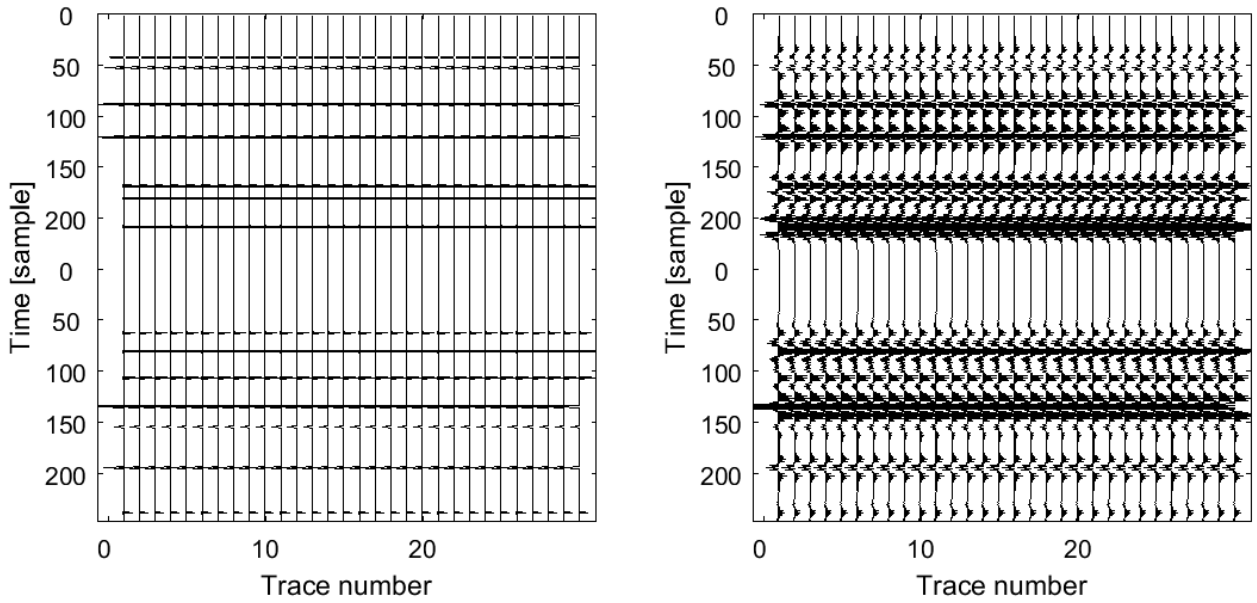
Real Data Results



(a) (b) (c)

Real data inversion results: (a) seismic data; (b) Estimated reflectivity; (c) Reconstructed seismic data, $\rho_{\mathbf{s},\hat{\mathbf{s}}} = 0.87$.

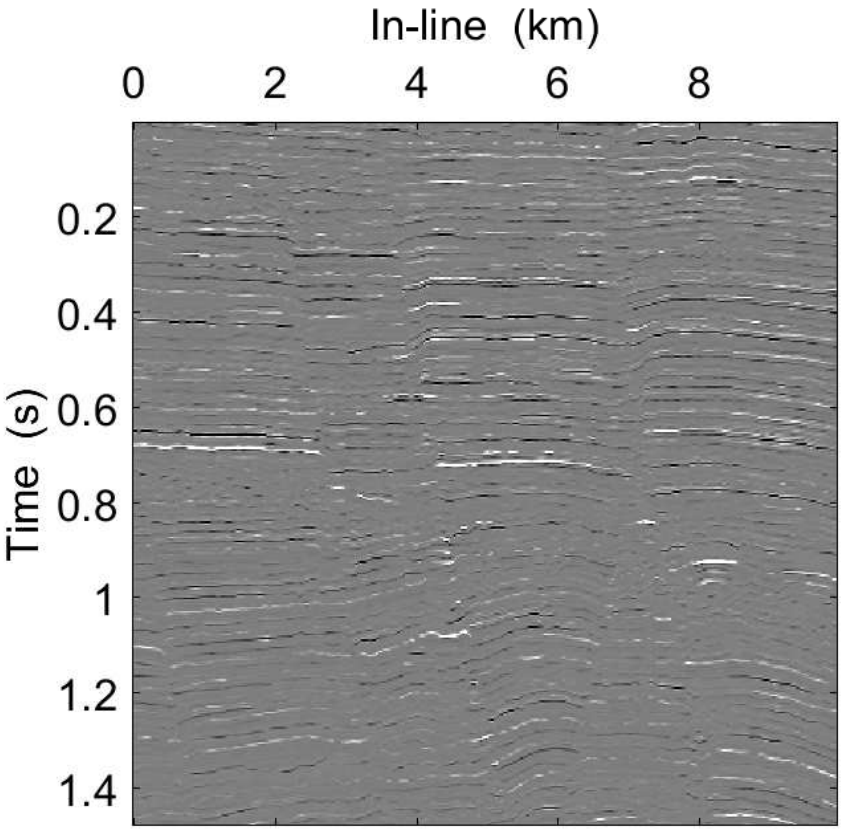
Real Data Experiments

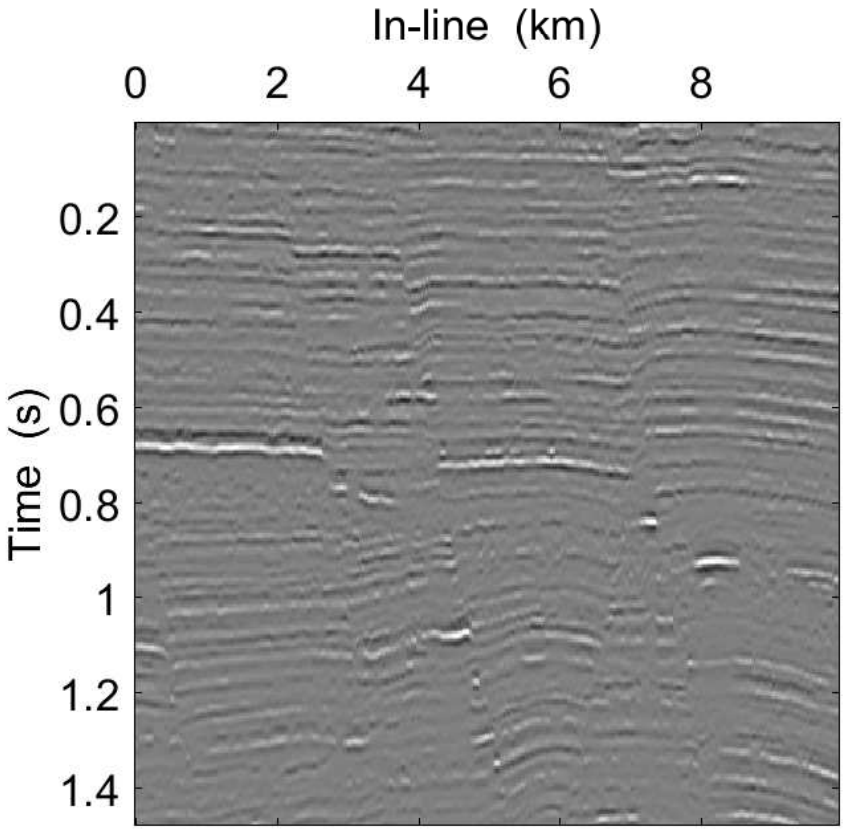


(a) (b)

Synthetic reflectivity and seismic data for RNN training: (a) Synthetic 2D reflectivity section of horizontal layers only; (b) 2D seismic data.

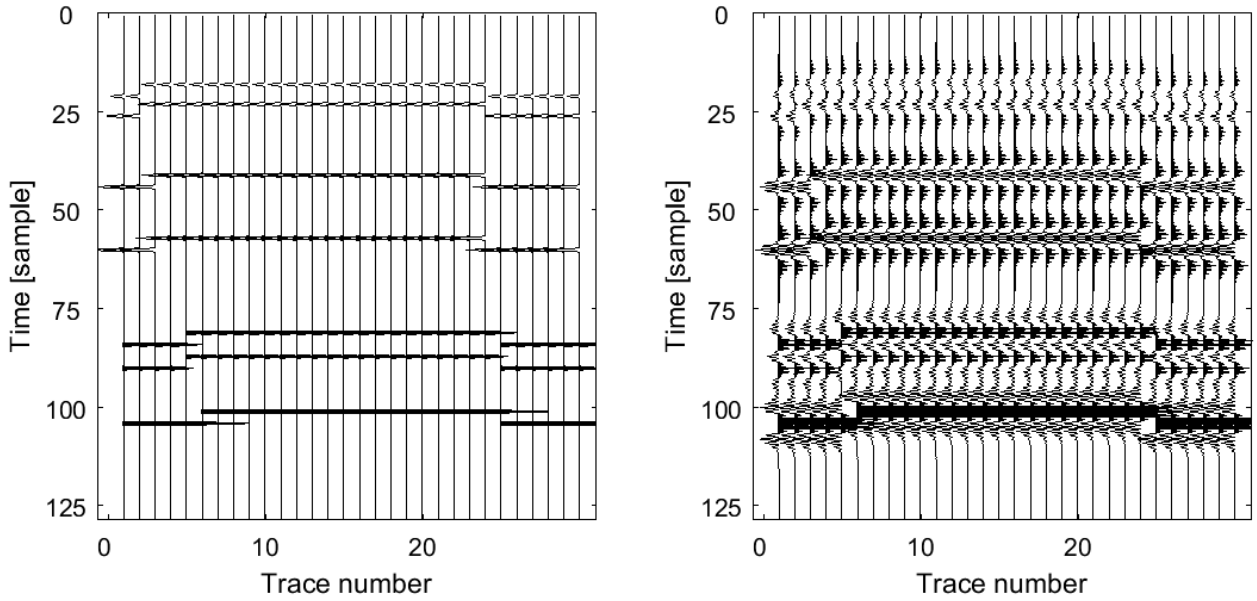
Real Data Experiments





Estimated reflectivity when RNN is trained with the horizontal layers only, and corresponding reconstructed seismic data, $\rho_{\mathbf{S},\hat{\mathbf{S}}} = 0.77$.

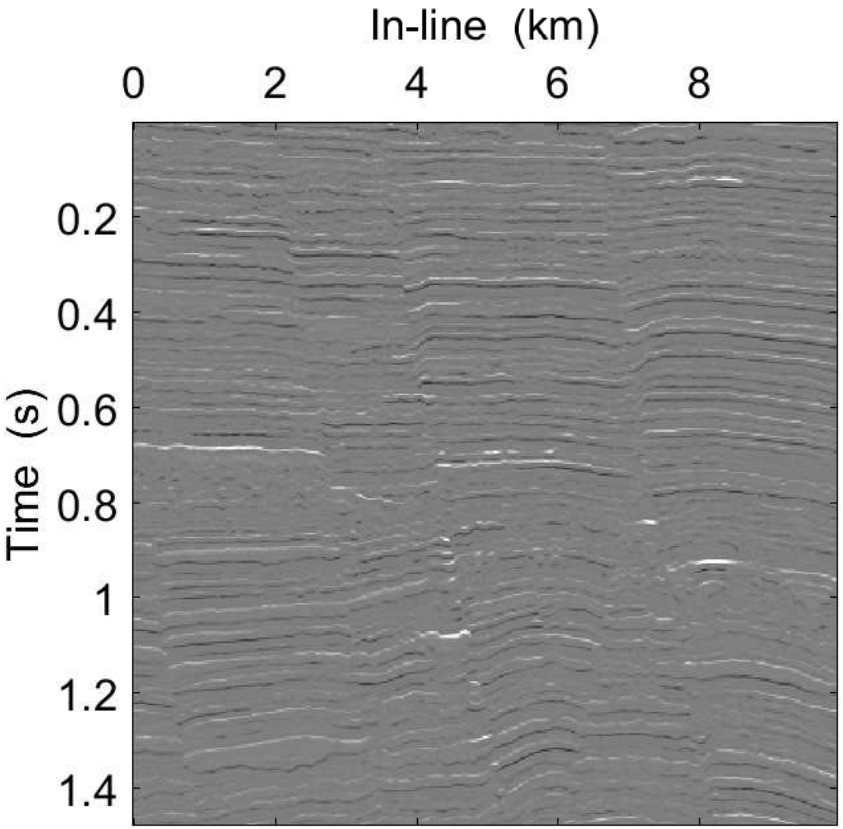
Real Data Experiments

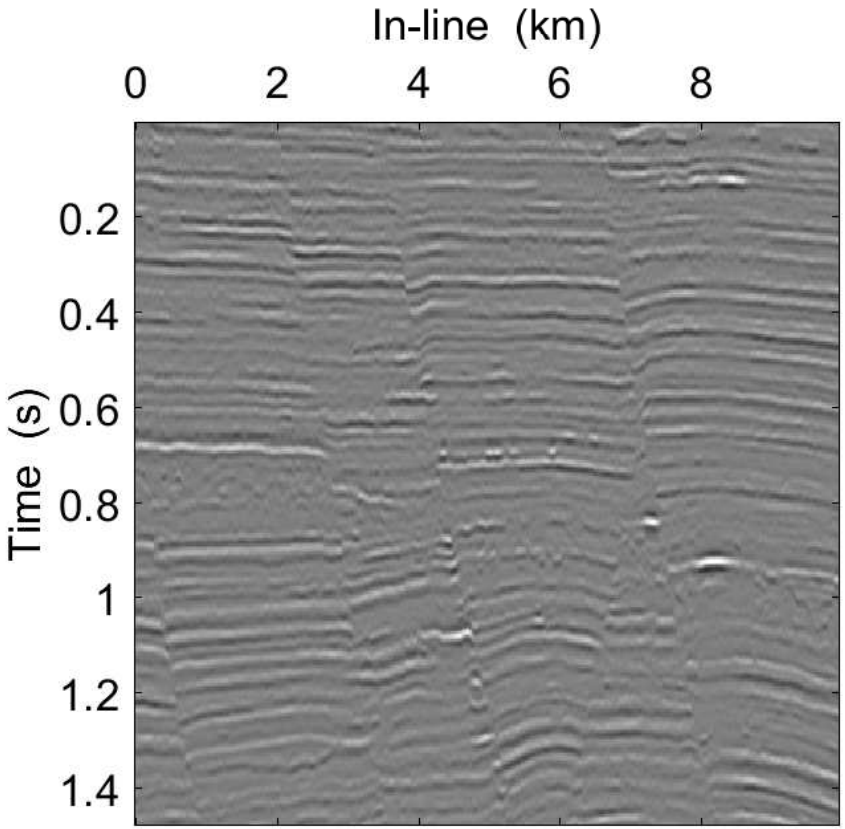


(a) (b)

Synthetic reflectivity and seismic data for RNN training: (a) Synthetic 2D reflectivity section of horizontal layers and two apparent faults; (b) 2D seismic data.

Real Data Experiments

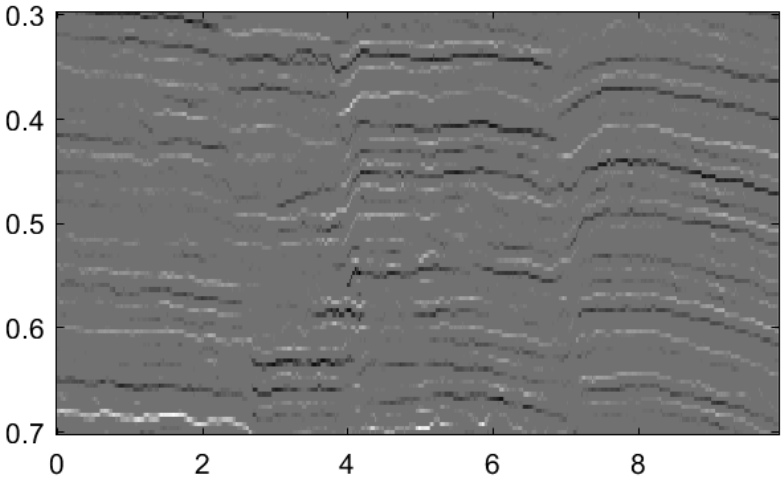


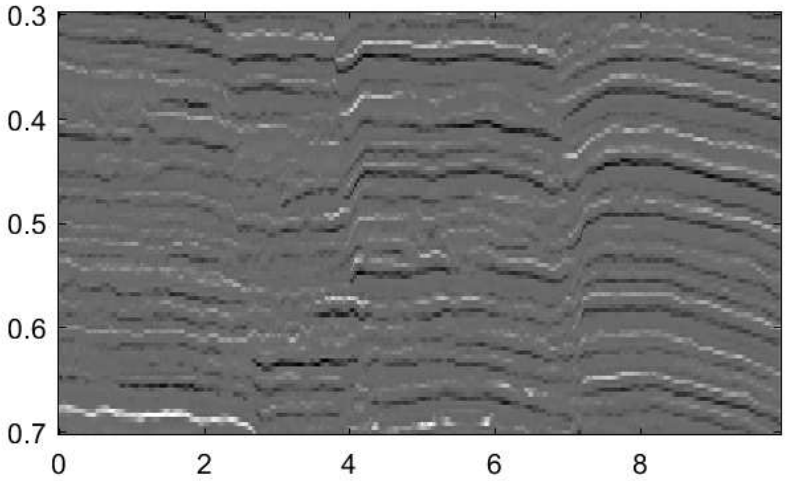


Estimated reflectivity and corresponding reconstructed seismic data,

$$\rho_{\mathbf{S},\hat{\mathbf{S}}} = 0.88.$$

Real Data Experiments





Zoom in: Left - Estimated reflectivity by 3D time-variant deconvolution (Pereg et al.,2019); Right - Estimated reflectivity by RNN trained with horizontal layers and faults.

- Training time on GPU (RTX-2080 Ti) $\sim$ **10 min**
- Execution time $\sim 1 - 10$ secs

Table of Contents

1. Research Goals
2. Background and Previous Works
3. Prestack Time-Migration Velocity Analysis
4. Sparse Inversion
5. Multichannel Time-variant Recovery
6. Fast Sparse Coding
7. Deconvolution via RNN
8. Conclusion & Future Work

Conclusions - Sparse Seismic Inversion

- We presented sparse 3D seismic inversion algorithm based on discontinuity measures.

Conclusions - Sparse Seismic Inversion

- We presented sparse 3D seismic inversion algorithm based on discontinuity measures.
- Solving a convex optimization problem has shown to locate the elements of the support with high resolution as long as they are separated by $\nu\sigma$.

Conclusions - Sparse Seismic Inversion

- We presented sparse 3D seismic inversion algorithm based on discontinuity measures.
- Solving a convex optimization problem has shown to locate the elements of the support with high resolution as long as they are separated by $\nu\sigma$.
- The method both promotes sparsity of the solution and also takes into account the spatial dependency between neighboring traces.

Conclusions - Sparse Seismic Inversion

- We presented sparse 3D seismic inversion algorithm based on discontinuity measures.
- Solving a convex optimization problem has shown to locate the elements of the support with high resolution as long as they are separated by $\nu\sigma$.
- The method both promotes sparsity of the solution and also takes into account the spatial dependency between neighboring traces.
- Results are improved

Conclusions - Sparse Seismic Inversion

- We presented fast sparse coding approximation algorithm based on receptive field normalization.

Conclusions - Sparse Seismic Inversion

- We presented fast sparse coding approximation algorithm based on receptive field normalization.
- We proposed seismic deconvolution based on RNN.

Conclusions - Sparse Seismic Inversion

- We presented fast sparse coding approximation algorithm based on receptive field normalization.
- We proposed seismic deconvolution based on RNN.
- We presented two automatic methods for seismic migration velocity analysis based on RNN.

Future Work

- Extend to blind-deconvolution
- Combine 3D algorithm and fast sparse coding tailored for local analysis
- Non stationary data by transfer learning
- Upgrade architecture. e.g.: skipping connections, and dropout
- Explore other applications, such as: ultrasound imaging

thank you

Local Structural Entropy (LSE)

An analysis cube of $2L_2$ inline trace, $2L_2$ crossline traces and N samples. The analysis cube is split into 4 $L_1 \times L_2 \times N$ quadrants, rearranged into columns vectors $\{\mathbf{a_i}|\mathbf{i} = \mathbf{1}, ..., \mathbf{4}\}$.

S - the correlation matrix of the analysis cube

$$\mathbf{S} = \frac{1}{NL_1L_2} \begin{pmatrix} \mathbf{a_1}^T\mathbf{a_1} & \dots & \mathbf{a_1}^T\mathbf{a_4} \\ \vdots & \ddots & \vdots \\ \mathbf{a_4}^T\mathbf{a_1} & \dots & \mathbf{a_4}^T\mathbf{a_4} \end{pmatrix}$$

The LSE measure associated with a point within the analysis cube

$$\mathcal{L}(i,j,t) = \frac{tr\mathbf{S}}{||\mathbf{S}||} - 1 = \frac{\sum_{i=1}^{4} \mathbf{a_i}^T\mathbf{a_i}}{\sqrt{\sum_{i,j=1}^{4} (\mathbf{a_i}^T\mathbf{a_j})^2}} - 1$$

back

Earth Q model

Wang* showed that given a travel time t_n, the reflected wave can be modeled as

$$u_n(t - t_n) = Re\left\{\frac{1}{\pi}\int_0^\infty S(\omega)\exp[j(\omega t - \kappa r(\omega))]d\omega\right\},$$

where $S(\omega)$ is the Fourier transform of the source waveform $s(t)$,

$$\kappa r(\omega) \triangleq \left(1 - \frac{j}{2Q}\right)\left|\frac{\omega}{\omega_0}\right|^{-\gamma}\omega t_n,$$

$$\gamma \triangleq \frac{2}{\pi}\tan^{-1}\left(\frac{1}{2Q}\right) \approx \frac{1}{\pi Q},$$

and Q is the medium quality factor, assumed to be frequency independent and known.

*Y. Wang, Seismic inverse Q-Filtering, Blackwell Pub., 2008.

Earth Q model

The expression of the earth Q filter consists of two exponential operators that express the phase effect (caused by velocity dispersion) and the amplitude effect (caused by energy absorption).

$$U_n(\omega)\exp^{-j\omega t_n} = S(\omega)exp\Big(\ -j\left|\frac{\omega}{\omega_0}\right|^{-\gamma}\omega t_n\Big)exp\Big(-\left|\frac{\omega}{\omega_0}\right|^{-\gamma}\frac{\omega t_n}{2Q}\Big)$$

Then, the time-domain seismic signal is

$$u_n(t-t_n) = \frac{1}{2\pi}\int U_n(\omega)\exp^{-j\omega t_n}\exp^{j\omega t}\,d\omega.$$

Earth Q model

The known set of kernels (pulses) $\{g_{\sigma,n}\}$ is defined as

$$g_{\sigma,n}(t - t_n) = u(t - t_n)|_{\sigma=\omega_0^{-1}}.$$

- $\sigma > 0$ is the scaling parameter: $g_\sigma(t) := g(\frac{t}{\sigma})$.
- For the Ricker wavelet $\sigma = \omega_0^{-1}$.

back

Q Estimation

- Q represents the effects of amplitude attenuation and velocity dispersion on the propagation of a plane wave through a homogeneous viscoelastic medium.

Q Estimation

- Q represents the effects of amplitude attenuation and velocity dispersion on the propagation of a plane wave through a homogeneous viscoelastic medium.
- By definition, Q is the energy loss during each cycle or wavelength.

$$Q = 2\pi \frac{E_0}{\Delta E}$$

E_0 is the maximum strain energy, and ΔE is the energy dissipation per cycle.

Q Estimation

Assuming an initial Ricker wavelet with ω_0, we compute Q according to the relation between Q and the peak frequency at time t,

$$Q = \frac{\pi t f_p f_m^2}{2(f_m^2 - f_p^2)},$$

f_m is the initial peak frequency. f_p is the peak frequency at time t.

Q Estimation

Assuming an initial Ricker wavelet with ω_0, we compute Q according to the relation between Q and the peak frequency at time t,

$$Q = \frac{\pi t f_p f_m^2}{2(f_m^2 - f_p^2)},$$

f_m is the initial peak frequency. f_p is the peak frequency at time t. We estimate f_m from two points in time t_1 and t_2, and two corresponding peak frequencies f_{p1} and f_{p2}

$$f_m = \sqrt{\frac{f_{p1} f_{p2}(t_2 f_{p1} - t_1 f_{p2})}{(t_2 f_{p2} - t_1 f_{p1})}}.$$

back

Proof outline

Build a *dual certificate* of the form:

$$q(t) = \sum_m a_m g_m(t - t_m) + b_m g_m^{(1)}(t - t_m).$$

$q(t)$ satisfies

$$|q(t_m)| = 1 \quad \forall t_m \in T,$$
$$q^{(1)}(t_m) = 0 \quad \forall t_m \in T.$$
$$|q(t)| < 1 \quad \forall t \notin T.$$

$q(t)$ enables us to decouple the estimation error at t_m from the amplitude of the rest of the spikes. $< q, x > = \sum_m |c_m|$.

Proof outline

- We prove $q(t)$ exists and the coefficients are bounded

$$||\mathbf{a}||_\infty \leq \frac{3\nu^2}{3\gamma_0\nu^2 - 2\pi^2\tilde{C}_0},$$

$$||\mathbf{b}||_\infty \leq \frac{3\pi^2\tilde{C}_1\nu^2}{(3\gamma_2\nu^2 - \pi^2\tilde{C}_2)(3\gamma_0\nu^2 - 2\pi^2\tilde{C}_0)},$$

where $\mathbf{a} \triangleq \{a_m\}$, $\mathbf{b} \triangleq \{b_m\}$ are coefficient vectors and

$$\tilde{C}_l = \max_m C_{l,m}, \ \ l = 0, 1, 2, 3.$$

Proof outline

- And also

$$a_m \geq \frac{1}{\alpha_0 + 2\tilde{C}_0 E(\nu) + \frac{(2\tilde{C}_1 E(\nu))^2}{\gamma_2 - 2\tilde{C}_2 E(\nu)}}.$$

$$E(\nu) \triangleq \frac{\pi^2}{6\nu^2}.$$

Proof outline - Multichannel

Assume N seismic traces that correspond to N identical reflectivity channels

$$y_i[k] = \sum_n x[n] g_{\sigma,n}[k-n] + w_i[k], \quad E\|w_i\|_2^2 \leq S_{\mathrm{w}}^2, \quad i = 1, 2 \ldots N,$$

If K satisfies the separation condition then the solution $\hat{x}$ of

$$\begin{aligned} &\min && \|\hat{\mathbf{x}}\|_1 \\ &\text{subject to} && \textstyle\sum_{i=1}^{N} \|\mathbf{y}_i - G\hat{\mathbf{x}}\|_2, \leq N S_{\mathrm{w}}, \end{aligned}$$

where satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \frac{64 L_r \rho^2}{\beta^2 \gamma_0^2} \frac{S_{\mathrm{w}}^2}{N}.$$

$$\rho \triangleq \max\left\{\frac{\gamma_0}{\varepsilon^2}, (F_s \sigma)^2 \alpha_0\right\}$$

Proof outline - Multichannel

$$\left|\left|q_\sigma[k]h[k]\right|\right|_2 = \left|\left|\left(\sum_{k_m\in K} a_m g_{m,\sigma}[k-k_m] + b_m g^{(1)}_{m,\sigma}[k-k_m]\right)h[k]\right|\right|_2$$

$$\leq \|a\|_\infty \left|\left|\sum_m g_{m,\sigma}[k-k_m]h[k]\right|\right|_2 + \|b\|_\infty \left|\left|\sum_m g^{(1)}_{m,\sigma}[k-k_m]h[k]\right|\right|_2 .$$

Proof outline - Multichannel

$$
\begin{aligned}
\left\| \sum_m g_{m,\sigma}[k-k_m]h[k] \right\|_2 &= \left\| \sum_m g_{m,\sigma}[k-k_m]\hat{x}[k] - \sum_m g_{m,\sigma}[k-k_m]x[k] \right\|_2 \\
&= \left|\left| G\hat{\mathbf{x}} - G\mathbf{x} \right|\right|_2 \\
&= \left\| \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i - G\mathbf{x} - \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i - G\hat{\mathbf{x}} \right) \right\|_2 \\
&\leq \left\| \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i - G\mathbf{x} \right\|_2 + \left\| \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i - G\hat{\mathbf{x}} \right\|_2 \\
&\leq \frac{2S_{\mathrm{w}}}{\sqrt{N}},
\end{aligned}
$$

Proof outline - Multichannel

$$\left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i - G\mathbf{x}\right|\right|_2 \leq \frac{S_{\mathrm{w}}}{\sqrt{N}}.$$

Since we know

$$\left|\left|\mathbf{y}_i - G\mathbf{x}\right|\right|_2 \leq S_{\mathrm{w}},$$

and the noise signals are uncorrelated,

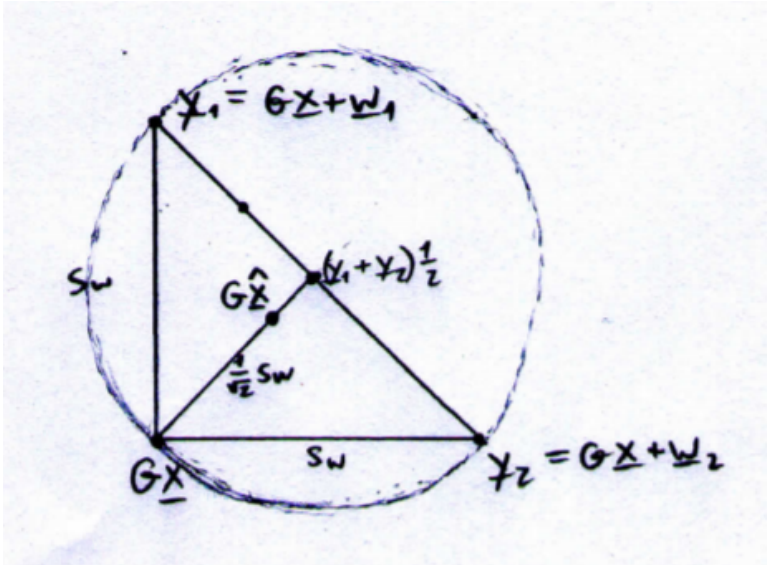
$$\begin{aligned}\left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i - G\mathbf{x}\right|\right|_2 &= \left|\left|\frac{1}{N}\sum_{i=1}^{N}G\mathbf{x} + \mathbf{w}_i - G\mathbf{x}\right|\right|_2 \\ &= \left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i\right|\right|_2 = \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{w}_i\mathbf{w}_j} \leq \frac{S_{\mathrm{w}}}{\sqrt{N}}.\end{aligned}$$

Proof outline - Multichannel

$$\left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i - G\hat{\mathbf{x}}\right|\right|_2 \leq \frac{S_{\mathrm{w}}}{\sqrt{N}},$$

An estimated reflectivity channel obeys

$$\sum_{i=1}^{N}\left|\left|\mathbf{y}_i - G\hat{\mathbf{x}}\right|\right|_2^2 \leq NS_{\mathrm{w}}^2.$$



$$\sum_{i=1}^{N}\left|\left|Gx - G\hat{x} + \mathbf{w}_i\right|\right|_2^2 \leq N\left|\left|\mathbf{w}_i\right|\right|_2^2$$

$$N\left|\left|Gx - G\hat{x}\right|\right|_2^2 + 2\sum_{i=1}^{N} < \mathbf{w}_i, Gx - G\hat{x} > + \sum_{i=1}^{N}\left|\left|\mathbf{w}_i\right|\right|_2^2 \leq N\left|\left|\mathbf{w}_i\right|\right|_2^2$$

$$\left|\left|Gx - G\hat{x}\right|\right|_2^2 + 2 < \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i, Gx - G\hat{x} > + \left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i\right|\right|_2^2 \leq + \left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i\right|\right|_2^2$$

$$\left|\left|Gx - G\hat{x} + \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i\right|\right|_2^2 \leq \frac{S_{\mathrm{w}}^2}{N}$$

Proof outline - Multichannel

Which means that,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i - G\hat{\mathbf{x}} \right\|_2^2 \leq \frac{S_{\mathrm{w}}^2}{N},$$

The estimated trace $G\hat{\mathbf{x}}$, reconstructed from the solution $\hat{\mathbf{x}}$, is as close to the averaged trace as the true noise-free trace $G\mathbf{x}$.

Proof outline

As mentioned above $\{g_m\}$ is a set of admissible kernels. Therefore,

$$|g^{(1)}_{m,\sigma}[k-k_m]| = \left|g^{(1)}_m\left(\frac{k-k_m}{N\sigma}\right)\right| \leq \frac{C_{1,m}}{1+\left(\frac{k-k_m}{N\sigma}\right)^2}.$$

Under the separation condition we have $|k_i - k_j| \geq N\nu\sigma \quad \forall k_i, k_j \in K$.

Hence, for any k we have

$$\sum_{k_m \in K} \frac{1}{1+\left(\frac{k-k_m}{N\sigma}\right)^2} < 2(1+P(\nu)).$$

Since

$$\sum_{n=1}^{\infty} \frac{1}{1+(n\nu)^2} \leq P(\nu) \triangleq \frac{\pi^2}{6\nu^2}.$$

Proof outline

Then,

$$E\Big\| \sum_{k_m \in K} \Big(g^{(1)}_{m,\sigma}[k-k_m]h[k] \Big) \Big\|_2^2 < 4\tilde{C}_1^2 (1+P(\nu))^2 E\|h\|_2.$$

$$E\Big\| q_\sigma[k]h[k] \Big\|_2 \leq \frac{2S_w}{\sqrt{N}} \delta ||a||_\infty + 2\tilde{C}_1 (1+E(\nu)) ||b||_\infty E||h||_2.$$

On the other hand,

$$\begin{aligned} E\Big\| q_\sigma[k]h[k] \Big\|_2 &= E\Big\| q_\sigma[k](h_K[k]+h_{K^C}[k]) \Big\|_2 \\ &\geq E||h_K||_2 - \max_{k \in \mathbb{Z}\setminus K} |q_\sigma[k]| E\|h_{K^C}[k]\|_2. \end{aligned}$$

Proof outline - Multichannel

We get

$$E\left\|q_\sigma[k]h[k]\right\|_2 \leq \frac{2S_w}{\sqrt{N}}\|a\|_\infty + 2\tilde{C}_1(1+E(\nu))\|b\|_\infty\|h\|_2.$$

And consequently we can derive

$$E\|h\|_2 \leq \frac{72\sqrt{L_r}\rho\gamma_2}{9\beta\gamma_0\gamma_2 - D_1\nu^{-2} + D_2\nu^{-4}}\frac{S_w}{\sqrt{N}}.$$

$$D_1 = 3\pi^2(\beta\gamma_0\tilde{C}_2 + 2\beta\gamma_2\tilde{C}_0 + 8\sqrt{L_r}\rho\tilde{C}_1^2), \quad D_2 = 2\beta\pi^4\tilde{C}_2\tilde{C}_0 - 4\pi^4\sqrt{L_r}\rho\tilde{C}_1^2.$$

For sufficiently large ν we have

$$E\|h\|_2^2 \leq \frac{64L_r\rho^2}{\beta^2\gamma_0^2}\frac{S_w^2}{N}.$$

This completes the proof.

Principles of velocity Analysis

The average velocity V_{rms} relates to a specific raypath

$$V_{rms}^2 = \frac{\int_0^t V^2(t)dt}{\int_0^t dt}$$

The interval velocity $v(\tau)$ - the average velocity over some travel path.

Dix formula: in a layered medium,

$$v^2(\tau) = \frac{\tau_i V_{rms}^2(\tau_i) - \tau_{i-1} V_{rms}^2(\tau_{i-1})}{\Delta\tau_i},$$

where $\Delta\tau_i$ is the time thickness of the ith layer.

$$V_{rms}^2(\tau_N) = \frac{\sum_{i=1}^N v^2(\tau_i)\Delta\tau_i}{\sum_{i=1}^N \Delta\tau_i},$$

$\tau_N = \sum_i \tau_i$ - total two-way traveltime to the bottom of the N'th layer. back

Non constant velocity

When velocity is not constant, the summation surfaces have more complex shapes, computed numerically through a velocity function.

All methods solve the Eikonal equation (an approximation of the wave equation (Bleinstein, 1984)), using

- Ray-tracing methods (Cerveny and Psencik, 1983)
- Finite difference methods (Vidale, 1990; Sava and Fomel, 2001; Popovici and Sethian, 2002)

The Eikonal equation

Consider a plane wave function $P(x, y, z; t)$ with a constant amplitude P_0 and spatially varying traveltime $\tau(x, y, z)$.

$$P(x, y, z; t) = P_0 \exp\{-j\omega[t - \tau(x, y, z)]\}. \tag{1}$$

The plane-wave solution of the form (1) satisfies the scalar wave equation

$$\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2}\right) = \frac{1}{v^2(x, y, z)} \frac{\partial^2 P}{\partial t^2} \tag{2}$$

Substitution of equation (2) into equation (3) yields

$$\left(\frac{\partial^2 \tau}{\partial x^2} + \frac{\partial^2 \tau}{\partial y^2} + \frac{\partial^2 \tau}{\partial z^2}\right) = \frac{1}{v^2(x, y, z)} \tag{3}$$

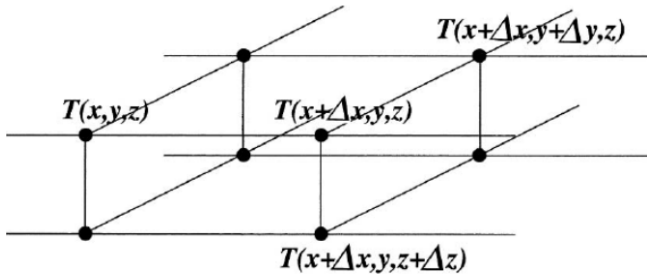
$\tau(x, y, z)$ is the traveltime value for a ray passing through a point

Ray tracing methods

- "Ray tracing system" - A system of 6 differential equations solved by standard numerical solvers.
- Based on asymptotic ray theory.
- Applicable both in isotropic and anisotropic models.
- The model parameters may vary both in vertical and lateral directions.
- Applicable to arbitrary source-receiver configurations.

Finite difference methods

we extrapolate the travel time from one point to another down to the source, by computing the traveltime τ of the eikonal equation at a grid point $(x+\Delta x, y+\Delta y, z+\Delta z)$ using the known traveltimes at grid points $(x,y,z),(x+\Delta x,y,z),(x+\Delta x,y+\Delta y,z)$.



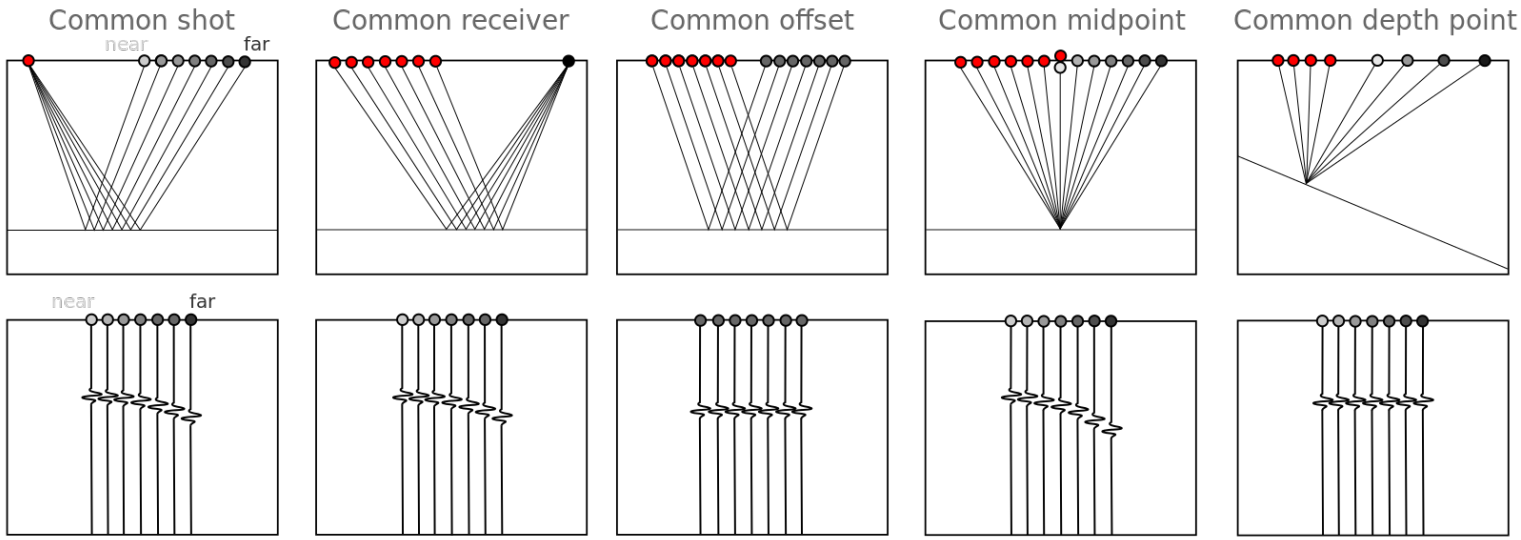
$$T_{i+1,j}^{k+1} = T_{i+1,j}^{k} + \frac{\Delta z}{v_{i+1,j}^{k+1}} \left[1 - \left(\frac{v_{i+1,j}^{k} T_{i+1,j}^{k} - v_{i,j}^{k} T_{i,j}^{k}}{\Delta x} \right)^2 - \left(\frac{v_{i+1,j+1}^{k} T_{i+1,j+1}^{k} - v_{i+1,j}^{k} T_{i+1,j}^{k}}{\Delta y} \right)^2 \right]^{1/2}$$

$T_{i,j}^{k}$ is the discrete form of $\tau(x,y,z)$.

back

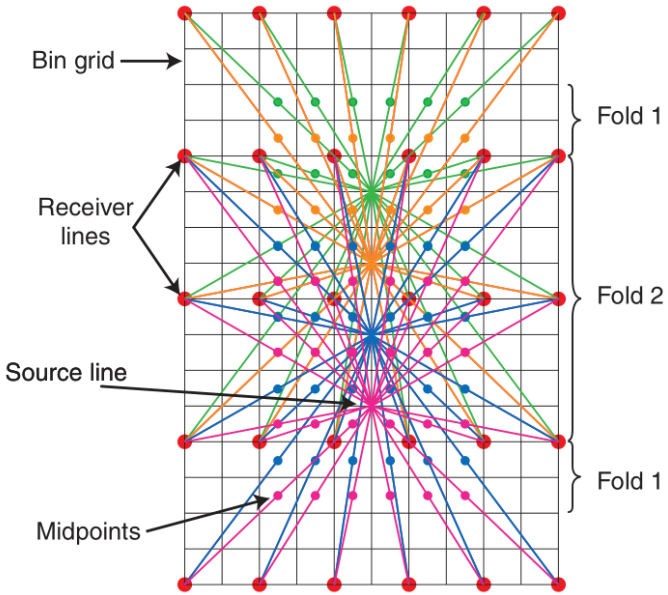
Stacking

The term stacking refers to the operation of adding together records of a particular gather to yield a single stacked trace. Stacking increases SNR, attenuating ground roll, air waves, and other noises, and attenuates multiples profoundly.

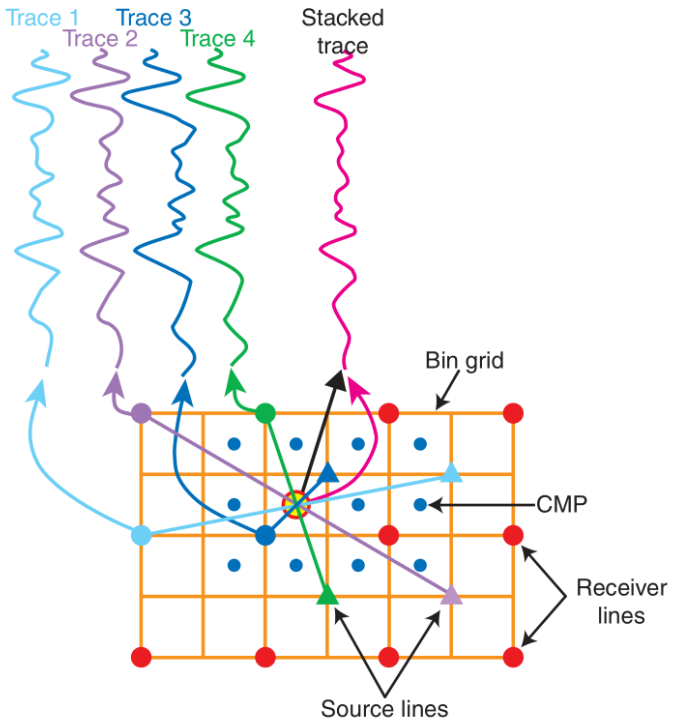


Seismic gathers.

Stacking



(a) 3D orthogonal design



(b) Illustration of CMP Stacking.

back

Principles of Velocity Analysis

- The simplest method for measuring relative delays in the data is to measure the moveout along the offset axis.
- When reflections originate from a flat interface and the medium above the interface is isotropic and homogeneous, the moveout is hyperbolic with offset.
- Stacking-velocity spectra (Taner and Koehler, 1969) is computed from a CMP gather by applying an NMO correction with several constant velocities, within a given velocity range.

Principles of Velocity Analysis

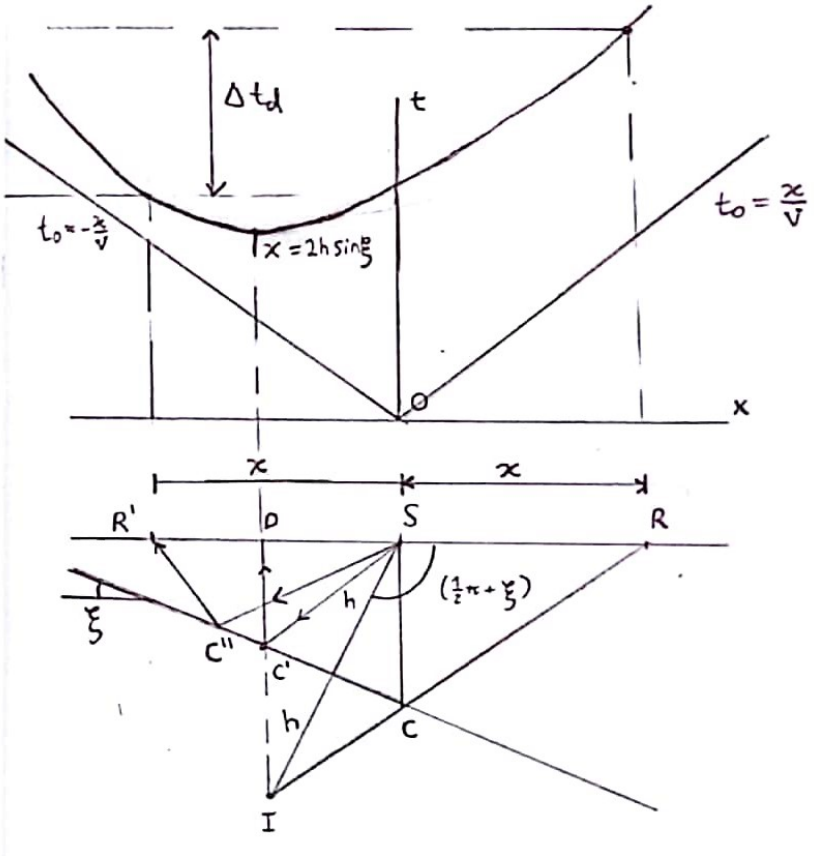
- The flatness of the NMO-corrected gathers is measured by using a coherency measure (such as semblance or stacking power).
- The stacking velocities corresponding to the maxima of the spectra are then picked, either by hand or automatically.
- Note:

$$V_{stacking} = \frac{V_{rms}}{\cos \xi}$$

for dipping parallel reflectors.

back

Dip Moveout (DMO)



Traveltime curve for a dipping reflector.

Assume ξ is the dip of a bed and h is the distance normal to the bed.

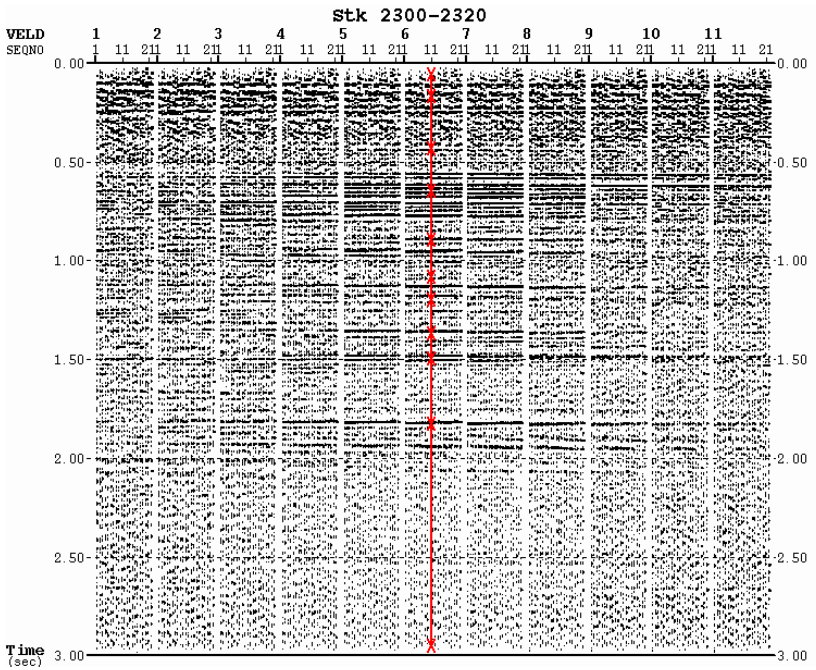
$$V^2 t^2 = x^2 + 4h^2 + 4hx \sin\xi$$

$$\frac{V^2 t^2}{(2h\cos\xi)^2} - \frac{(x + 2h\sin\xi)^2}{(2h\cos\xi)^2} = 1.$$

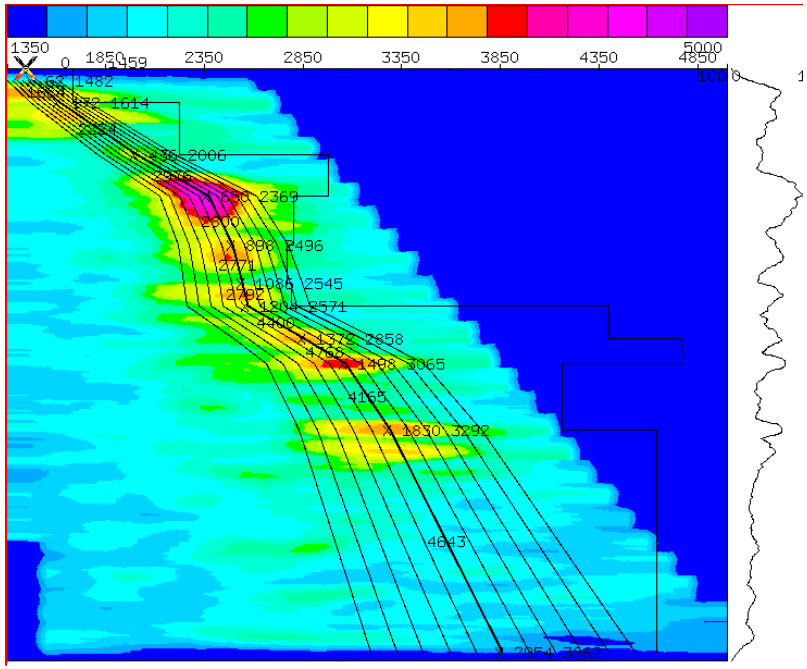
Traveltime curve is a hyperbola, but $x = -2h\sin\xi$ is the axis of symmetry.

back

Real Data - Quality Control Plots



(a)



(b)

Quality Control (QC) plots: (a) Migration stack of the supergather for the CDP range 2300-2320 for different velocities. (b) Semblance values for rms velocities. The migration velocities used for (a) are marked with a black solid line.

back

Iterative-Soft Thresholding Algorithm (ISTA)

Consider the cost function

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

We build a *surrogate* function

$$Q(\mathbf{x}, \mathbf{x}_k) = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 + \frac{c}{2}\|\mathbf{x} - \mathbf{x}_k\|_2^2 - \frac{1}{2}\|\mathbf{D}\mathbf{x} - \mathbf{D}\mathbf{x}_k\|_2^2$$

The related function $Q(\mathbf{x}, \mathbf{x}_k)$ obeys

1. Equality at $\mathbf{x}_k$: $Q(\mathbf{x}_k, \mathbf{x}_k) = f(\mathbf{x}_k)$.
2. Upper-bounding the original function: $Q(\mathbf{x}, \mathbf{x}_k) \geq f(\mathbf{x}) \quad \forall \mathbf{x}$.
3. Tangent at $\mathbf{x}_k$: $\nabla Q(\mathbf{x}, \mathbf{x}_k)|_{\mathbf{x}=\mathbf{x}_k} = \nabla f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k}$.

Iterative-Soft Thresholding Algorithm (ISTA)

Consider the cost function

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

The term

$$d(\mathbf{x}, \mathbf{x}_k) = \frac{c}{2}\|\mathbf{x} - \mathbf{x}_k\|_2^2 - \frac{1}{2}\|\mathbf{D}\mathbf{x} - \mathbf{D}\mathbf{x}_k\|_2^2,$$

is a measure of the distance to the previous solution.

- For directions $\mathbf{x} - \mathbf{x}_k$ that are close to the null space of $\mathbf{D}$, the distance is nearly Euclidean.
- For directions $\mathbf{x} - \mathbf{x}_k$ that are spanned by $\mathbf{D}$, the distance drops to nearly zero. (We usually choose $c = \|\mathbf{D}\|_2^2$.)

back