Lei Fu

Department of Earth Science, Rice University
Cell: (713) 586-9211 E-mail: lei.fu.rice@gmail.com

EDUCATION

PhD, Rice University — 2011 - present
Seismic imaging, extended waveform inversion

MS, University of Utah — 2009 - 2011
Electromagnetic Modeling and Inversion

BS, University of Science and Technology of China — 2005 - 2009
Geophysics, global seismology

RESEARCH

Subsurface offset extended waveform inversion

Accelerate extended waveform inversion by adaptive multiscale methods

Fast extended waveform inversion using Morozov's discrepancy principle

Accelerate extended waveform inversion by adaptive multiscale methods

Lei Fu

The Rice Inversion Project (TRIP), Rice University

April 25, 2016

Overview

Extended model permits data fit throughout update process (avoid cycle skip)

Problem

BUT extended model $\Rightarrow$ computational cost $\uparrow$

Solution

Adaptive approach: reduce extension while maintain data fit with velocity updates

Multiscale method: refine grids (coarse $\rightarrow$ fine), frequencies (low $\rightarrow$ high)

Extended modeling concept

In order to fit the data, thus avoiding cycle-skipping.

Extended Born operator F

$$F[v]r = d$$

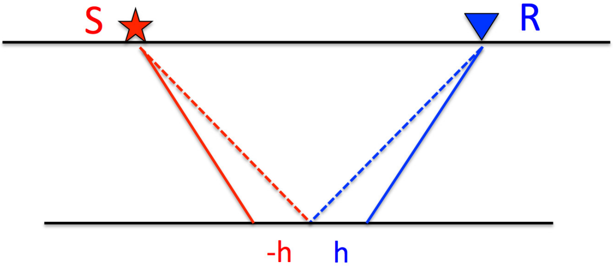
$v(\mathbf{x})$: background velocity

$r(\mathbf{x}, \mathbf{h})$: extended velocity perturbation, allows fit to d for any v

d: sampled pressure data at receivers

Subsurface offset extension

Originated in Claerbout's survey sinking concept [Symes, 2008; Biondi and Almomin, 2012; Shen, 2012; Shan and Wang, 2013; Weibull and Arntsen, 2013]



Subsurface offset: distance between subsurface scattering points

Physical meaning: action at a distance

Subsurface offset extended linearized acoustic modeling

δu - scattered (perturbation) pressure field

$$\left(\frac{\partial^2}{\partial t^2} - v^2(\mathbf{x})\nabla^2\right)\delta u(t,\mathbf{x};\mathbf{x_s}) = \int_{-H}^{H} d\mathbf{h} r(\mathbf{x},\mathbf{h})\nabla^2 u(t,\mathbf{x}+2\mathbf{h};\mathbf{x_s})$$

RHS involves an integration over $\mathbf{h}$, equivalent of a full matrix multiply at every time step - can easily overwhelm cost of ordinary time-stepping [Mulder, 2013]

v: P-wave velocity
$w(t)$: source function, t: time
$\mathbf{h}$: horizontal subsurface offset

$\mathbf{x}$: position
$\mathbf{x}_s$: source location
H: limit of $\mathbf{h}$

Extended full waveform inversion (EFWI)

Objective function:

$$J[v, r] = \frac{1}{2}\|F[v]r - d\|^2 + \frac{\alpha}{2}\|Ar\|^2$$

A: annihilator (A = h).

$\alpha > 0$, penalty for non-focus. Choose $\alpha \rightarrow$ talk 14:45

Solved with variable projection method, Nested optimization: [Golub and Pereyra, 1973, Golub and Pereyra, 2003]

- inner loop, given v, find r optimizes $J[v, r]$.
- outer loop, find v optimizes reduced $J[v, r[v]]$.

Inner loop, optimize J over r

Gradient of $J_\alpha[v, r]$ with respect to r

$$\nabla_r J_\alpha[v, r] = F[v]^T (F[v]r - d) + \alpha A^T A r$$

where T denotes transpose.

$\nabla_r J_\alpha[v, r] = 0 \Rightarrow$ normal equation:

$$(F^T F + \alpha A^T A) r = F^T d$$

Extended least-squares reverse time migration (ELSRTM) solved by linear iterative method, e.g., conjugate gradient (CG).

Outer loop, update v

Gradient of reduced objective function $J_\alpha[v, r[v]]$ respect to v:

$$\nabla_v J_\alpha[v, r[v]] = DF^T\,(r, F[v]r - d)$$

DF^T tomographic or WEMVA operator.

DF, bilinear operator, linear in dm and r.

How to choose H?

Avoid cycle-skipping, fit data by using large enough H

$$\left(\frac{\partial^2}{\partial t^2} - v^2(\mathbf{x})\nabla^2\right)\delta u(t,\mathbf{x};\mathbf{x_s}) = \int_{-H}^{H} d\mathbf{h}\delta\bar{r}(\mathbf{x},\mathbf{h})\nabla^2 u(t,\mathbf{x}+2\mathbf{h};\mathbf{x_s})$$

Computational cost

Number of grid points in h, $N_h = \frac{2H}{dh}$

$N_h \downarrow \Leftarrow dh \uparrow$, coarse grid

$N_h \downarrow \Leftarrow H \downarrow$, how to choose H?

Computational cost

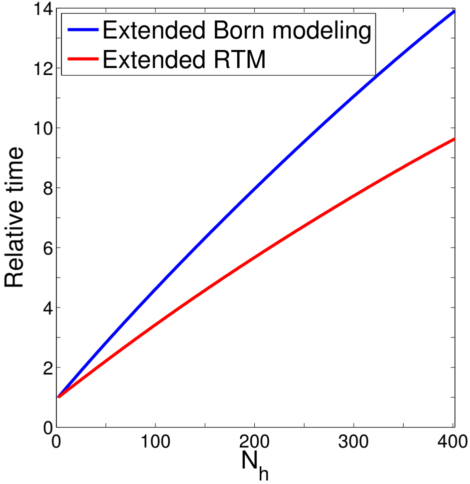


Figure : Relative computing time as a function of number of grid points N_h. The extended model r measures 1000×1000 cells with different N_h.

Single reflector example

2D constant density acoustic 2-8 order finite difference code

Parameter	Measurements
Source	Ricker wavelet $f_{peak} = 15\,Hz$
Source position $\mathbf{x_s}$	$x : 300 - 2700\,m$ every $40\,m$, $z = 0\,m$
Receiver position $\mathbf{x_r}$	$x : 0 - 3000\,m$ every $20\,m$, $z = 0\,m$
Subsurface offset h	$-1500\,m \leqslant h \leqslant 1500\,m$
Space and time	$x = 3000\,m$, $z = 2500\,m$, $t = 1.6\,s$
Grid size	$dx = dh = dz = 20\,m$, $dt = 2\,ms$
Background velocity	$v_{true} = 3.0\,km/s$
Max iter inner loop	20
α	0

Single reflector example

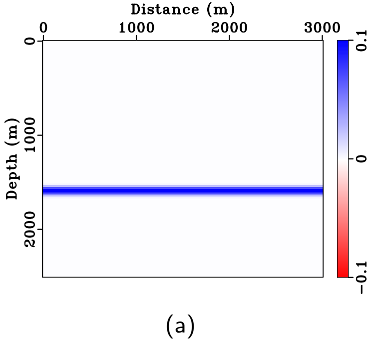


Figure : (a) Extended reflectivity r at $h = 0\,m$ (b) data at shot 31 at the center

Single reflector example

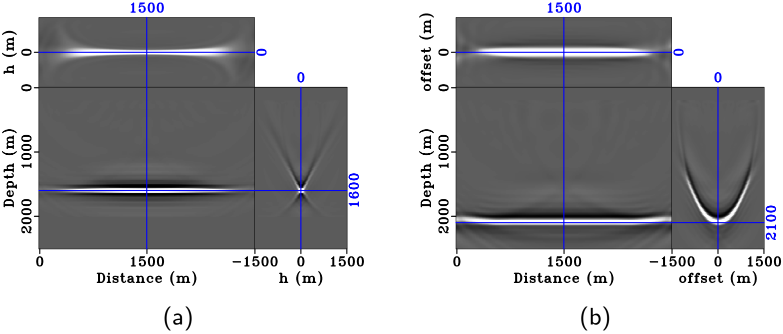


Figure : Inverted r (a) $v = v_{true}$ (b) $v = 1.3v_{true}$

Inverted r at $x = 1.5\,km$ for different v

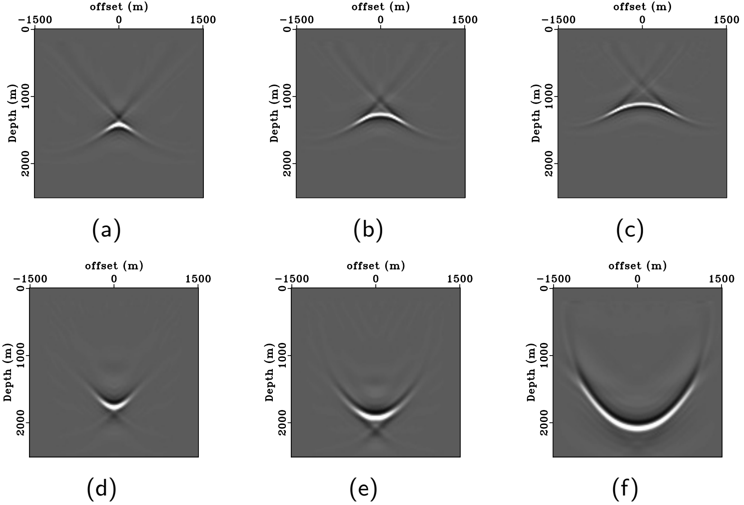


Figure : (a) $v = 0.9v_{true}$, (b) $0.8v_{true}$, (c) $0.7v_{true}$, (d) $1.1v_{true}$, (e) $1.2v_{true}$, (f) $1.3v_{true}$

Relation between H and data misfit

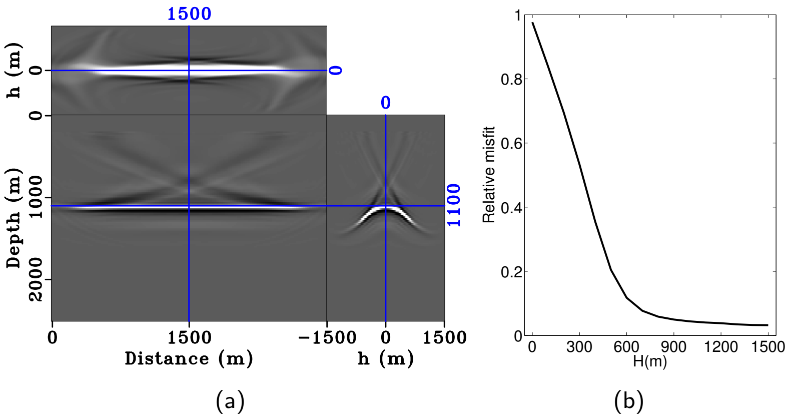
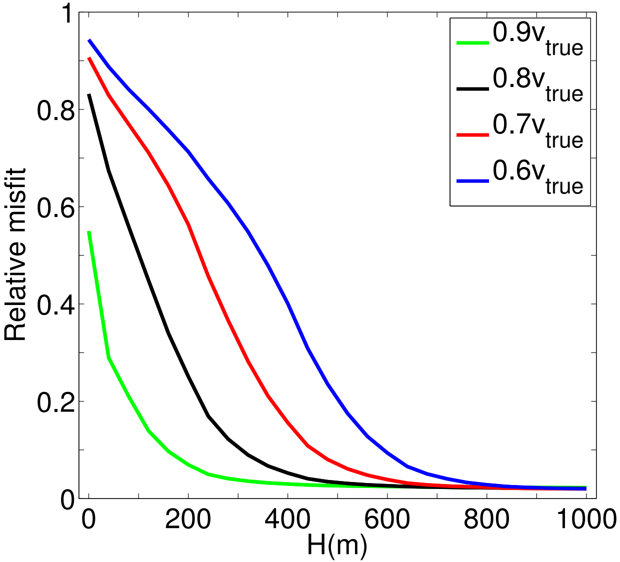


Figure : $v = 0.7v_{true}$ (a) inverted $\bar{r}$ (b) H vs relative Δd_H

Relationship between H and Δd_H at different v



Adaptive multiscale method

Idea: set H adaptively using data fitting criterion

For any “reasonable” v, can fit data if H adequate;

If v is "good", H can be "small" (physical model: $H = 0$)

Goal: control cost for large H

Decrease frequency, increase sample rates

v more accurate, r more focused, so smaller H is needed

1. Initialization

Initial H, constant velocity case [Shen, 2004, Mulder, 2014]

$$H = L|1 - \frac{v^2}{v_{true}^2}|$$

L - maximum surface offset

Number of refinement stages n

$$\frac{1}{2^{n-1}} f_{\max} > f_{\min}$$

Relative data misfit $\Delta d_H = \frac{\|F[v]r - d\|}{\|d\|}$

Half offset misfit $\Delta d_{H/2}$, restrict h range $[-H/2, H/2]$: only recompute $F[v]r$, NOT re-solve ELSRTM.

Data residual tolerance X, determined by accuracy of solving ELSRTM, target value $\Delta d_H \leq X \leq \Delta d_{H/2}$

Multiscale adaptive method

1. Initialization H, n, X

Multiscale adaptive method

1. Initialization H, n, X

2. Filter d, w with $(f_{\min} - \frac{1}{2^{n-k}} f_{\max})$, k - current refinement stage, $(dx, dz, dh, dt) \leftarrow 2^{n-k}(dx, dz, dh, dt)$.

Multiscale adaptive method

1. Initialization H, n, X

2. Filter d, w with $(f_{\min} - \frac{1}{2^{n-k}} f_{\max})$, k - current refinement stage, $(dx, dz, dh, dt) \leftarrow 2^{n-k}(dx, dz, dh, dt)$.

3. Solve ELSRTM to estimate r. Compute Δd_H and $\Delta d_{H/2}$.

Multiscale adaptive method

1. Initialization H, n, X
2. Filter d, w with $(f_{\min} - \frac{1}{2^{n-k}} f_{\max})$, k - current refinement stage, $(dx, dz, dh, dt) \leftarrow 2^{n-k}(dx, dz, dh, dt)$.
3. Solve ELSRTM to estimate r. Compute Δd_H and $\Delta d_{H/2}$.
4. If $\Delta d_H \leq X \leq \Delta d_{H/2}$, go to 7.

Multiscale adaptive method

1. Initialization H, n, X

2. Filter d, w with $(f_{\min} - \frac{1}{2^{n-k}} f_{\max})$, k - current refinement stage, $(dx, dz, dh, dt) \leftarrow 2^{n-k}(dx, dz, dh, dt)$.

3. Solve ELSRTM to estimate r. Compute Δd_H and $\Delta d_{H/2}$.

4. If $\Delta d_H \leq X \leq \Delta d_{H/2}$, go to 7.
5. If $\Delta d_{H/2} < X$, $H \leftarrow H/2$, go to 3.

Multiscale adaptive method

1. Initialization H, n, X

2. Filter d, w with $(f_{\min} - \frac{1}{2^{n-k}} f_{\max})$, k - current refinement stage, $(dx, dz, dh, dt) \leftarrow 2^{n-k}(dx, dz, dh, dt)$.

3. Solve ELSRTM to estimate r. Compute Δd_H and $\Delta d_{H/2}$.

4. If $\Delta d_H \leq X \leq \Delta d_{H/2}$, go to 7.
5. If $\Delta d_{H/2} < X$, $H \leftarrow H/2$, go to 3.
6. If $\Delta d_H > X$, $H \leftarrow 2H$, go to 3.

Multiscale adaptive method

1. Initialization H, n, X

2. Filter d, w with $(f_{\min} - \frac{1}{2^{n-k}} f_{\max})$, k - current refinement stage, $(dx, dz, dh, dt) \leftarrow 2^{n-k}(dx, dz, dh, dt)$.

3. Solve ELSRTM to estimate r. Compute Δd_H and $\Delta d_{H/2}$.

4. If $\Delta d_H \leq X \leq \Delta d_{H/2}$, go to 7.
5. If $\Delta d_{H/2} < X$, $H \leftarrow H/2$, go to 3.
6. If $\Delta d_H > X$, $H \leftarrow 2H$, go to 3.

7. Update v_+: if $|J[v_+] - J[v]| < \epsilon$, then $k \leftarrow k+1$, go to 2;
   else, $v \leftarrow v_+$, go to 3.

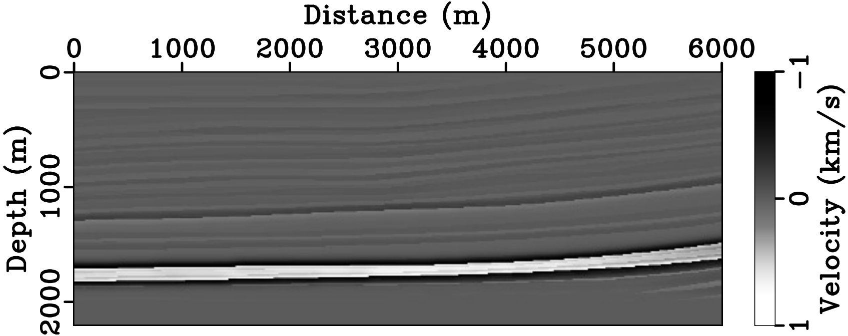
Lens model

2D constant density acoustic 2-8 order finite difference code

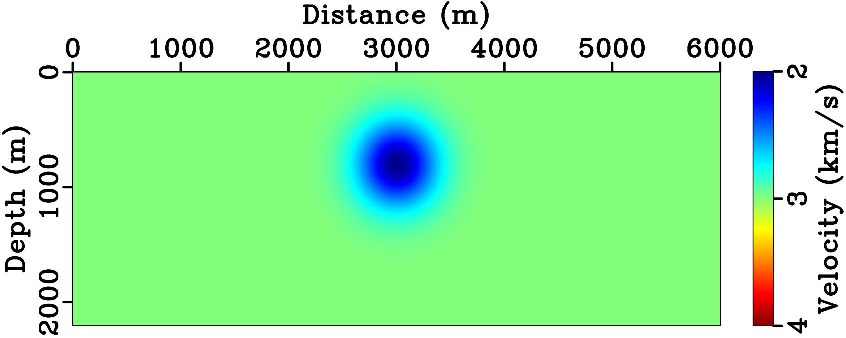
Steepest descent with quadratic backtrack line search

Parameter	Measurements
Source wavelet	bandpass $3 - 30\,Hz$
Source position $\mathbf{x_s}$	$x : 0 - 6\,km$ every $50\,m$, $z = 0\,m$
Receiver position $\mathbf{x_r}$	$x : 0 - 6\,km$ every $50\,m$, $z = 0\,m$
Space and time	$x = 6\,km$, $z = 2.2\,km$, $t = 2.4\,s$
Grid size	$dx = dh = dz = 12.5\,m$, $dt = 2\,ms$
Initial velocity	$v = 3.0\,km/s$
Maximum iter inner loop	20

Extended reflectivity r



True background velocity model v



Shot 61

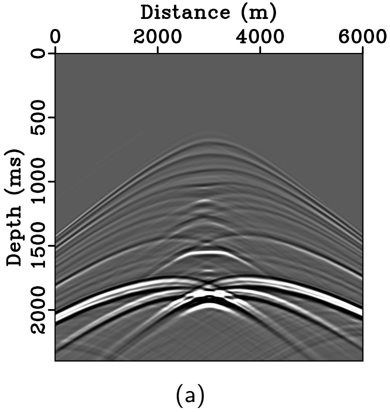


Figure : (a) Data of shot 61 at the center (b) bandpass $(3 - 7.5\,Hz)$ data

Initial background velocity model v_0

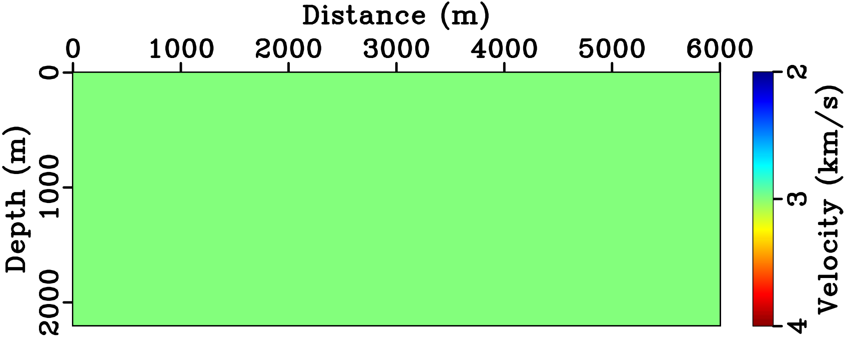
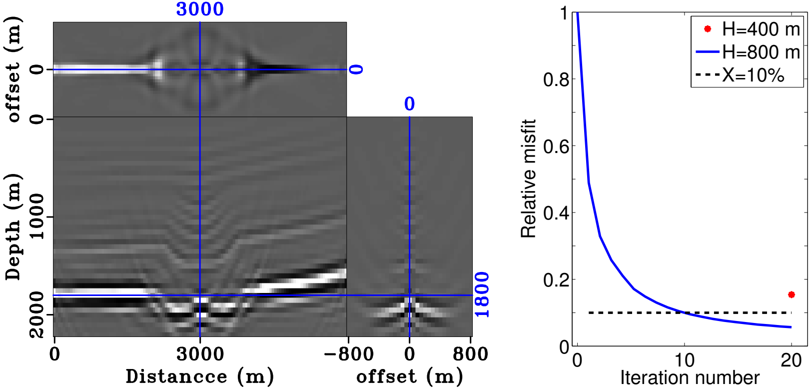


Figure : Stage 1: $dh = dx = dz = 50\,m$, bandpass filter $3 - 7.5\,Hz$, $dt = 8\,ms$

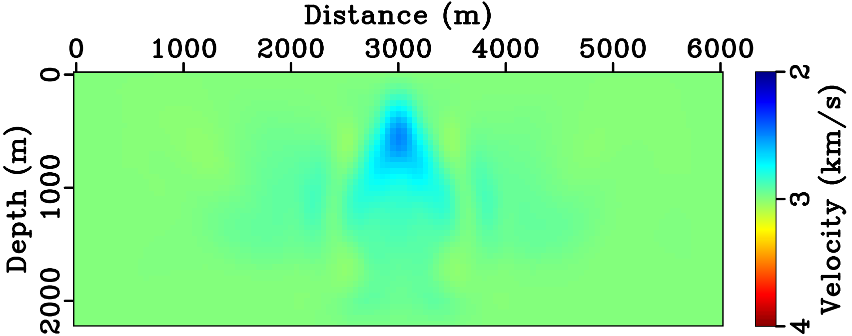
Inverted r_0 with initial v_0

Assume $v/v_{true} = 85\%$, $H = L|1 - \frac{v^2}{v_{true}^2}| \approx 800\, m$

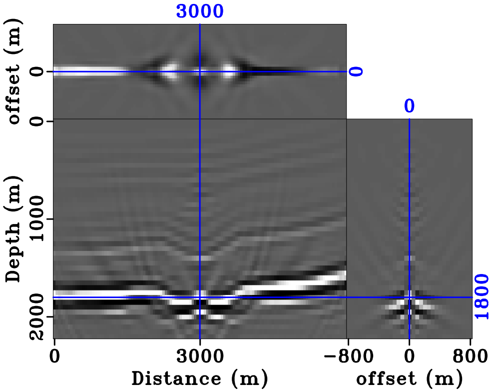


(a) (b)

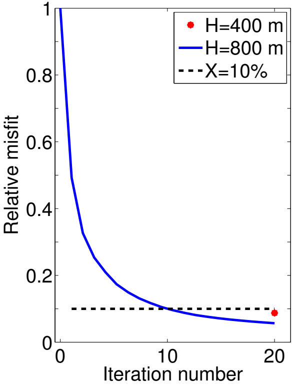
1st update v_1



r_1, $H: 800\ m \rightarrow 400\ m$

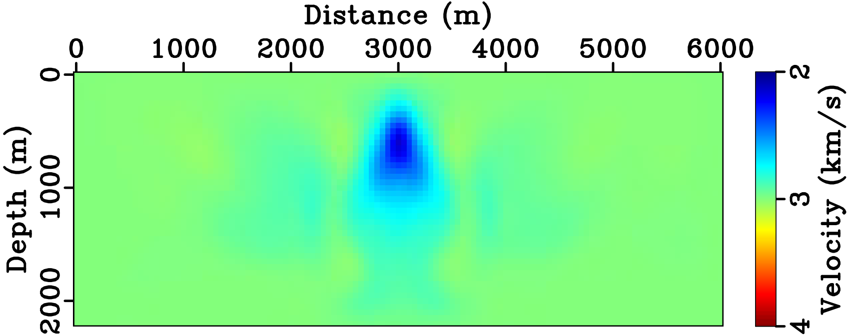


(c)



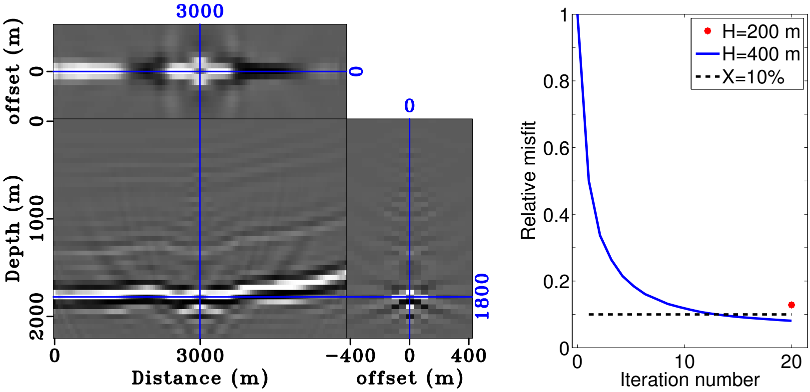
(d)

v_2



Distance (m)
0
1000
2000
3000
4000
5000
6000
Depth (m)
0
1000
2000
Velocity (km/s)
2
3
4

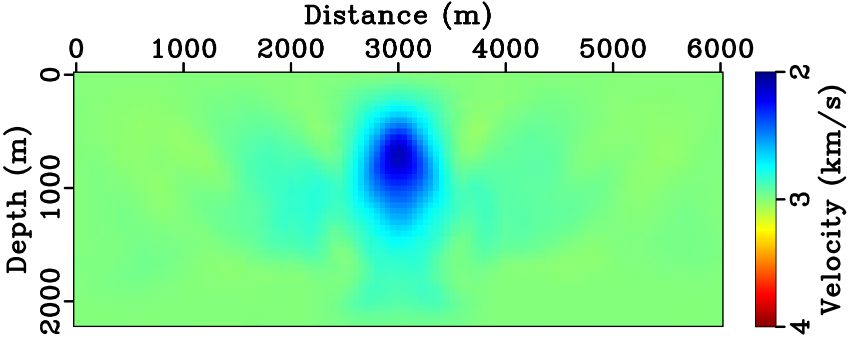
r_2



3000
0
0
1800
offset (m)
0
0
Depth (m)
1000
2000
0
3000
−400
0
400
Distance (m)
offset (m)
1
0.8
0.6
0.4
0.2
0
Relative misfit
H=200 m
H=400 m
X=10%
0
10
20
Iteration number

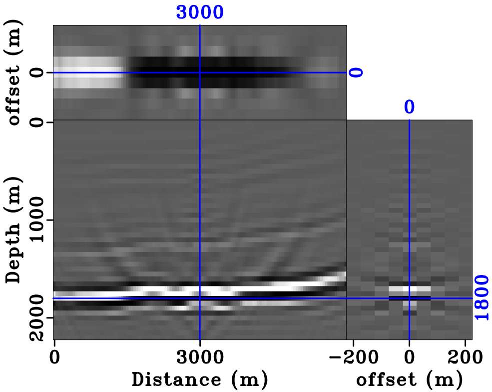
(e) (f)

v_5



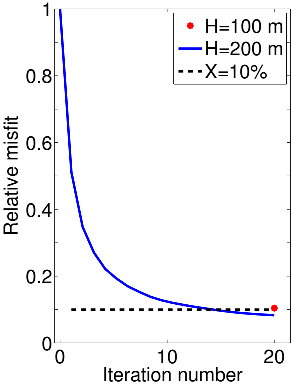
Distance (m)
0
1000
2000
3000
4000
5000
6000
Depth (m)
0
1000
2000
Velocity (km/s)
2
3
4

r_5



3000
0
offset (m)
0
0
0
Depth (m)
1000
2000
1800
0
3000
−200
0
200
Distance (m)
offset (m)

(g)



H=100 m
H=200 m
X=10%
Relative misfit
1
0.8
0.6
0.4
0.2
0
0
10
20
Iteration number

(h)

Stage 2, v_7

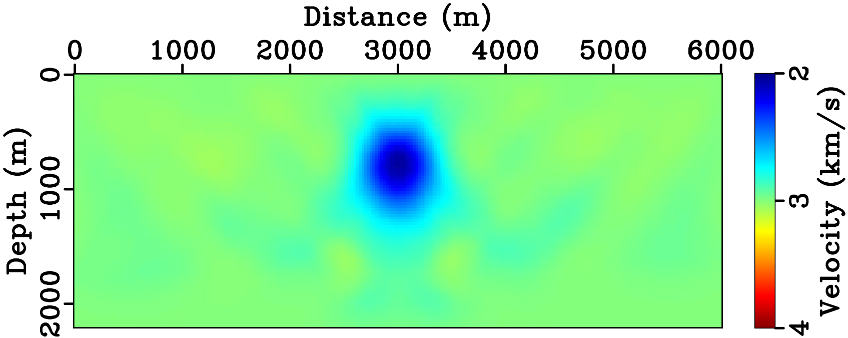
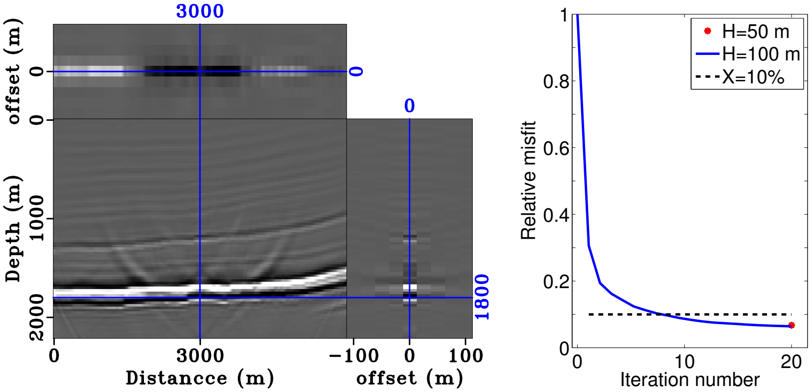


Figure : Stage 2: $dh = dx = dz = 25\,m$, bandpass filter $3 - 15\,Hz$, $dt = 4\,ms$

Stage 2, r_7



3000
offset (m)
0
0
0
0
Depth (m)
1000
2000
1800
0
3000
Distancce (m)
−100
0
100
offset (m)
1
0.8
0.6
0.4
0.2
0
Relative misfit
H=50 m
H=100 m
X=10%
0
10
20
Iteration number

(a) (b)

Stage 3, v_8

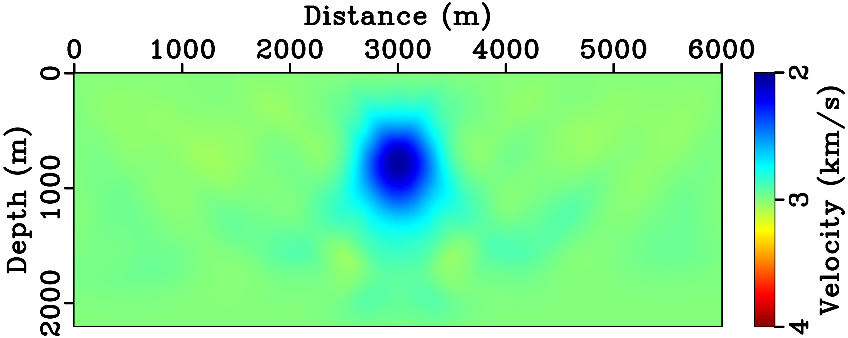
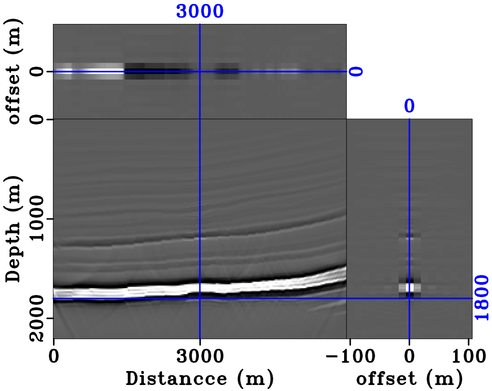


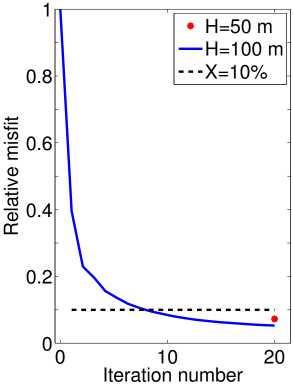
Figure : Stage 2: $dh = dx = dz = 12.5\,m$, original d and w, $dt = 2\,ms$

Stage 3, r_8



3000
offset (m)
0
0
0
0
Depth (m)
1000
2000
1800
0
3000
Distancce (m)
−100
0
100
offset (m)

(a)



1
0.8
0.6
0.4
0.2
0
Relative misfit
0
10
20
Iteration number
H=50 m
H=100 m
X=10%

(b)

Summary

Lens model, cost $\approx 7\%$ of finest gird and full offset

- Determine H adaptively using data fitting criterion
- Multiscale

Basic hypothesis: if H long enough, good data fit; adequate H deceases with velocity error.

Future work

Choose $\alpha \rightarrow$ my 2nd talk 14:45

Consider physics errors, variable density acoustics, field data

Acceleration: optimization methods, Preconditioner

Acknowledgments

The Rice Inversion Project (TRIP) sponsors

Rice University Research Computing Support Group (RCSG)

Texas Advanced Computing Center (TACC)

Madagascar, Seismic Unix

TRIP members

- Golub, G., and V. Pereyra, 1973, The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate: SIAM Journal on Numerical Analysis, **10**, 413–432.
- ———, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26.
- Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33.
- Shen, P., 2004, Wave equation migration velocity analysis by differential semblance optimization: PhD thesis, Rice University.