Seeing through Rock: Mathematics of Inverse Wave Propagation

William W. Symes

The Rice Inversion Project
Computational and Applied Mathematics
Rice University

SIAM Annual Meeting, July 2018

Inverse/imaging problems in seismology: from seismic data, deduce

- spatial distribution of rock mechanical parameters: wave velocities, density,...
- locations of faults and other structures - discontinuities in mechanical parameters

Agenda:

- how to make a seismic image (what's an image?)
- why it works, and how to improve it to an inversion $\approx$ data-predicting model (via iteration)
- remaining challenges (how to start iteration)

Agenda

Seismic Data, Physics, Simulation

Imaging and Asymptotic Inversion

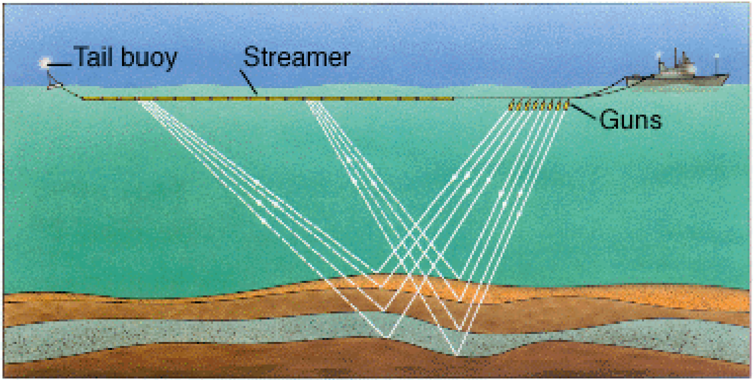
Linear least squares = Least squares migration'

Accelerating Linearized inversion (Least Squares Migration)

Accelerating Nonlinear (Full Waveform) Inversion

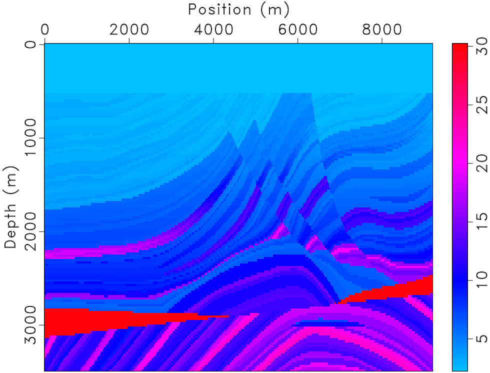
Autofocusing: estimating reference/initial model

Summary and Challenges



Marine streamer acquisition [thanks: Schlumberger]

Modeling mechanical parameter fields $C_{ijkl}, \rho, \ldots$: should allow at least discontinuities, cf. inspection of outcrops.



Left: Outcrop, Stuart I., WA (WS, 8/11). Right: Marmousi synthetic model (IFP 89): bulk modulus map κ, unit = GPa (density = 1 g/cm^3)

Acoustic model of seismic waves, a good if not great model:

$$\rho \frac{\partial \mathbf{v}}{\partial t} = -\nabla p; \; \frac{\partial p}{\partial t} = -\kappa \nabla \cdot \mathbf{v} + f; \; p, \mathbf{v} = 0, t << 0$$

p=pressure, $\mathbf{v}$=particle velocity, f=energy source

κ=bulk modulus, ρ=material density

.$\log \kappa, \log \rho \in L^{\infty}(\mathbf{R}^3) \Rightarrow$

- Unique causal weak solution $p, \mathbf{v}$ for causal $f \in L^2_{\mathrm{loc}}(\mathbf{R}_t, L^2(\mathbf{R}^3_{\mathbf{x}}))$
- Smooth in κ, ρ if f smooth in t

(Lions 68, 71, Stolk 00, Blazek et al. 13)

Energy source model: isotropic point radiators $f(\mathbf{x}, t; \mathbf{x}_s) = w(t)\delta(\mathbf{x} - \mathbf{x}_s)$ at source locations $\mathbf{x}_s$ (many!)

Receiver model: sample p at receiver locations $\mathbf{x}_r$ (many!), over time interval $[0, T]$

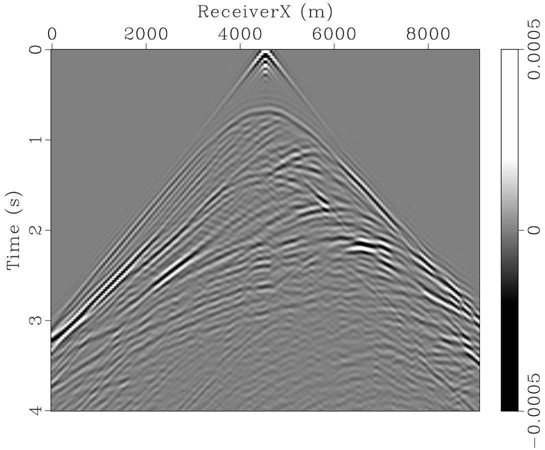
Acquisition manifold $(\mathbf{x}_s, \mathbf{x}_r) \in \Gamma \subset \mathbf{R}^3 \times \mathbf{R}^3$

Modeling (forward, prediction) operator $\mathcal{F}[\kappa] = p|_{\Gamma \times [0,T]}$ (this talk: fix ρ and ignore)

Numerical methods: Finite Difference, Finite Element (CG, DG, SEM,...),...

2D Marmousi example:

- 180 sources $x_s \in [240, 8832]], z_s = 12$ m
- 382 receivers $x_r \in [12, 9156], z_r = 12$ m
- $w =$ indef. integral of bandpass filter [2.5,5,25,30] Hz
- staggered FD scheme, order (2,8):
  - space grid: $767(x) \times 291(z)$, $\Delta x = \Delta z = 12$ m
  - recorded time grid: $n_t = 1001$, $\Delta t = 4$ ms (interp. from internal simulation grid)



Simulated ("synthetic") shot record $= \mathcal{F}[\kappa](\cdot,\cdot;\mathbf{x}_s)$

Shot 90 (of 180), $x_s = 4560$ m, $z_s = 12$ m

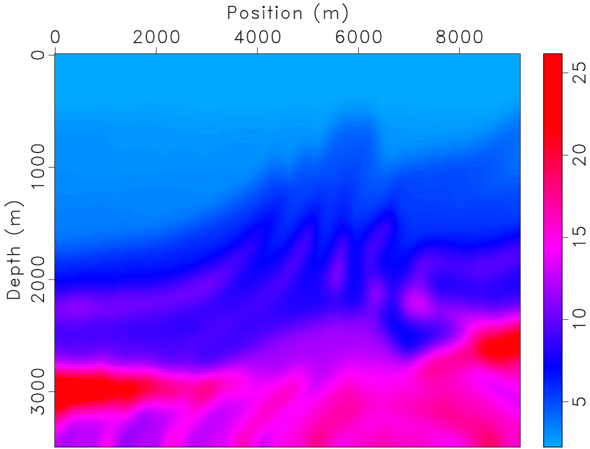
Basic inverse problem of seismology, acoustics version, as nonlinear least squares: given data d, find κ to minimize

$$J_{\rm FWI}[\kappa] = \frac{1}{2}\|\mathcal{F}[\kappa] - d\|^2$$

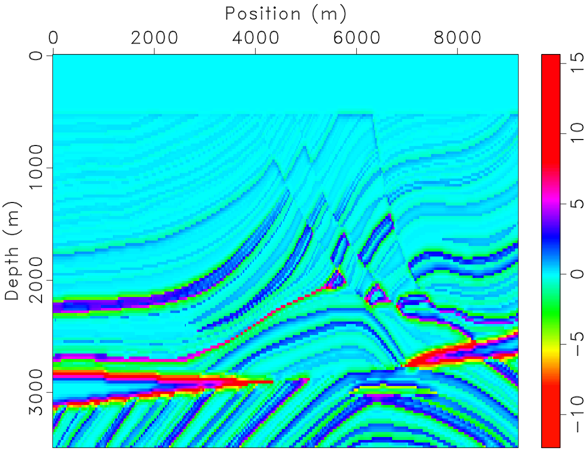
$=$ "Full waveform inversion" (FWI): feasibility, then commoditization over last 15 yrs - much promise, many challenges - largest number of sessions at 2017 SEG

Most practical industry data processing $\approx$ solution $\delta\kappa$ of linearized least squares

$$J_{\rm LSM}[\kappa_0, \delta\kappa] = \frac{1}{2}\|D\mathcal{F}[\kappa_0]\delta\kappa - \delta d\|^2$$



Background bulk modulus map κ_0 (density = 1 g/cm^3)
transparent (geometric optics), determines time of travel



Perturbation bulk modulus map $\delta\kappa$ (density perturbation $= 0$ g/cm^3), creates reflections

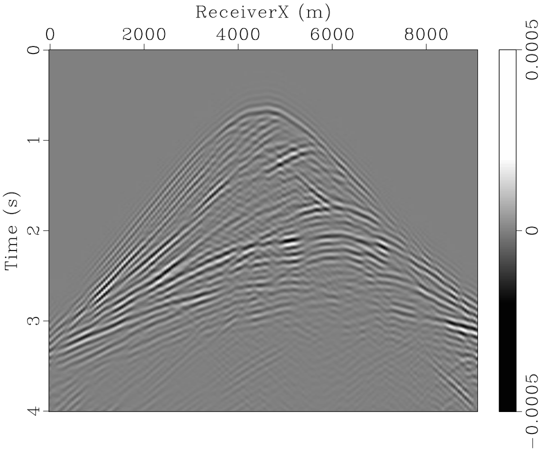
$D\mathcal{F}[\kappa_0]\delta\kappa = \delta p|_{\Gamma\times[0,T]}$, where

$$\rho_0 \frac{\partial \delta \mathbf{v}}{\partial t} = -\nabla \delta p$$

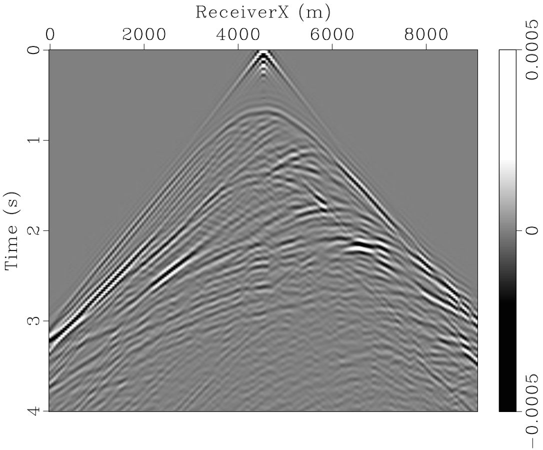
$$\frac{\partial \delta p}{\partial t} = -\kappa_0 \nabla \cdot \delta \mathbf{v} - \delta\kappa \nabla \cdot \mathbf{v}_0$$

$\Rightarrow$ use same FD scheme to approximate $D\mathcal{F}[\kappa_0]\delta\kappa$

Abbreviation: $F\delta\kappa = D\mathcal{F}[\kappa_0]\delta\kappa$



Linearized shot record $F\delta\kappa(\cdot,\cdot;\mathbf{x}_s)$, $x_s = 4560$ m



Simulated shot record - $\mathcal{F}[\kappa](\cdot,\cdot;\mathbf{x}_s)$, $x_s = 4560$ m

Agenda

Seismic Data, Physics, Simulation

Imaging and Asymptotic Inversion

Linear least squares = Least squares migration'

Accelerating Linearized inversion (Least Squares Migration)

Accelerating Nonlinear (Full Waveform) Inversion

Autofocusing: estimating reference/initial model

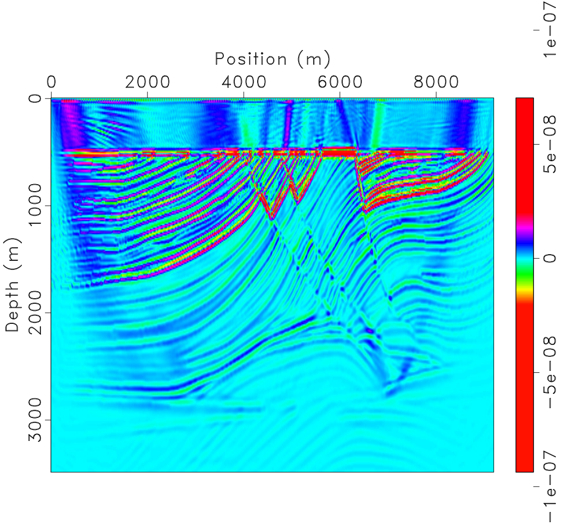
Summary and Challenges

Basic principle of seismic imaging: to form image of subsurface, "cook" data to resemble linearized data δd, then apply *transposed linearized modeling operator*

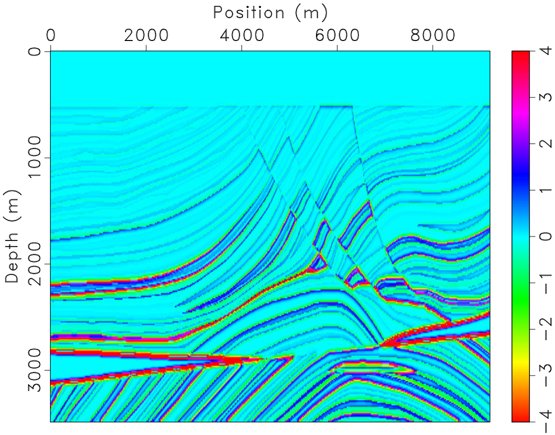
$$\text{image} = F^T \delta d$$

Computation of F^T via numerical solution of wave equations: *adjoint state method* (Chavent & Lemmonier 74, Plessix 06)

$\Rightarrow$ data as energy source in backwards-in-time simulation: *Reverse Time Migration* ("RTM")



$F^T \delta d$ = RTM image



$\delta\kappa =$ model perturbation

Why is $F^T \delta d$ an image?

"Image of $\delta\kappa$" $=$ "has (approximately some of) same singularities as $\delta\kappa$"

High frequency asymptotic analysis 80's-90's (Beylkin, Rakesh, Bleistein, Burridge, Spencer, de Hoop, Lambaré, Jin, Nolan, ten Kroode, Smit, Verdel, Stolk,...):

- some limitations $\Rightarrow F \approx$ *Fourier integral operator*
- more limitations $\Rightarrow F^T F \approx$ *pseudodifferential* operator ("ΨDO") - 2D: generic (Stolk 00)
- $\Rightarrow$ if $\delta d = F \delta\kappa$ then singularities of $F^T \delta d = F^T F \delta\kappa \subset$ singularities of $\delta\kappa$ (singularity $=$ *wave front set*)

Zhang & Bleistein 03, 05, Zhang et al. 09, ten Kroode 12,...: Explicit computation of *principal symbol* of $F^T F$

$$F^T F \delta\kappa(\mathbf{x}) \approx \text{const.} \;\; \times \int d\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} \frac{(\ldots)}{\cos\theta_r \cos\theta_s} \delta\hat{\kappa}(\mathbf{k})$$

- $\cos\theta_{s,r}$ = wave/ray angle of incidence at source, receiver
- "(...)" = explicit filters, rational functions of κ_0, ρ_0

NB: construction requires *extension* of F - extended model depends on artificial space variables, scattering angle,..., explicit computation of extended $F^T F$ (const.=1)

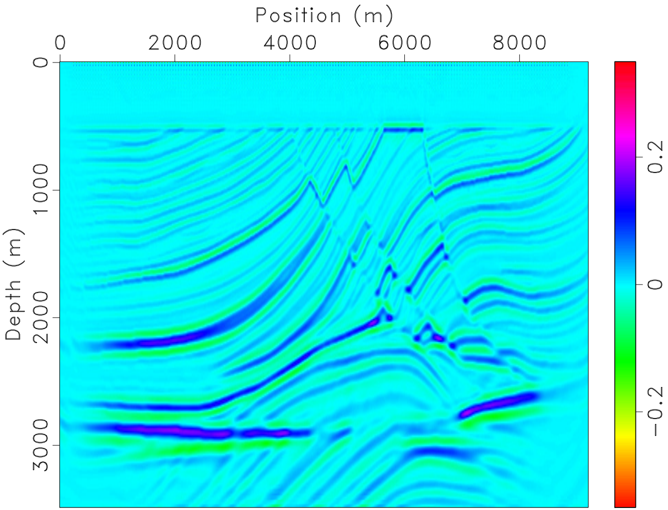
Calculus of ΨDOs $\Rightarrow$ cancel factors in integrand using computable filters, multiplication operators

$\Rightarrow$ approximate inverse $F^\dagger$ modulo scale, low frequency error = *true/preserved amplitude* RTM ("TARTM")

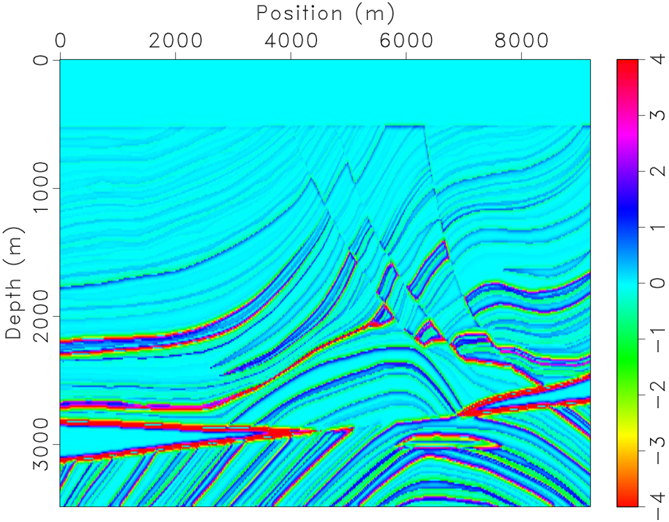
Hou & S. 15, 17: factorization

$$F^\dagger = W_m^{-1} F^T W_d,$$

W_m, W_d simple explicit filters: $W_d = -|f|^{-3}\frac{\partial}{\partial z_s}\frac{\partial}{\partial z_r}$, $W_m^{-1} = 32\rho^{-1}\kappa^3|k_z|$ - column-by-column ("trace-by-trace") action, negligible add'l cost beyond RTM



$F[\kappa_0]^\dagger d$ = true amplitude RTM



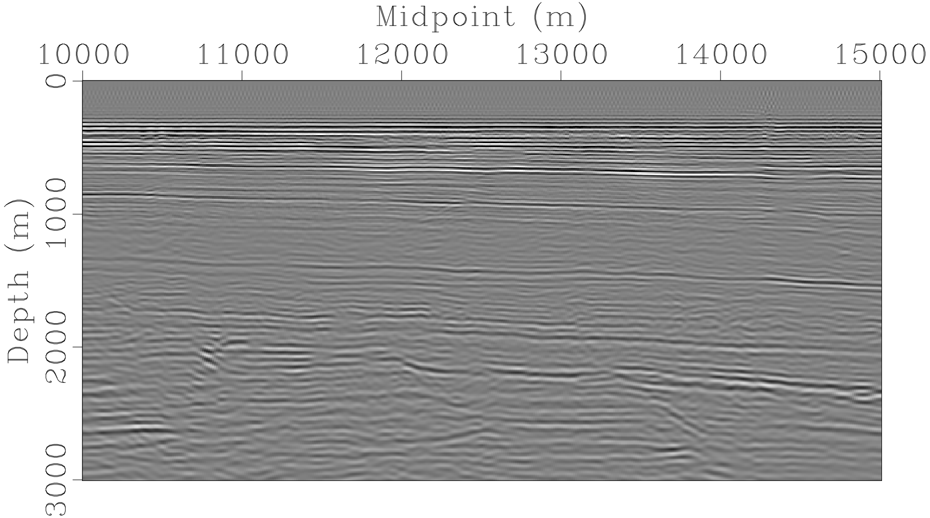
$\delta\kappa =$ model perturbation

Application to field data: Mobil "Viking Graben" survey - marine seismic line, Norwegian sector of North Sea

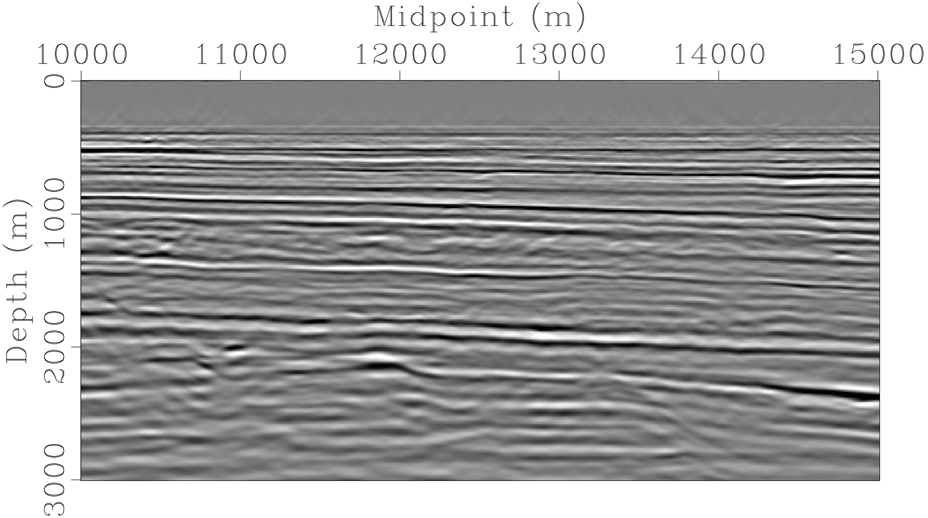
Released for 1994 SEG Annual Meeting post-convention workshop, described in Foster & Keys: "Comparison of Seismic Inversion Methods on a Single Real Dataset" (SEG 98)

Preprocessing ("parabolic Radon demultiple") - cook to resemble linearized data

This example: RTM vs. TARTM, 200 shots near Well B



Viking Graben: RTM image of shots 269-508 - Automatic Gain Control ("AGC") $\Rightarrow$ amplitudes are meaningless



Viking Graben: asymptotic inverse / TARTM image - reasonable fit to well log (Hou & S. *Geophysics* 18)

Agenda

Seismic Data, Physics, Simulation

Imaging and Asymptotic Inversion

Linear least squares = Least squares migration'

Accelerating Linearized inversion (Least Squares Migration)

Accelerating Nonlinear (Full Waveform) Inversion

Autofocusing: estimating reference/initial model

Summary and Challenges

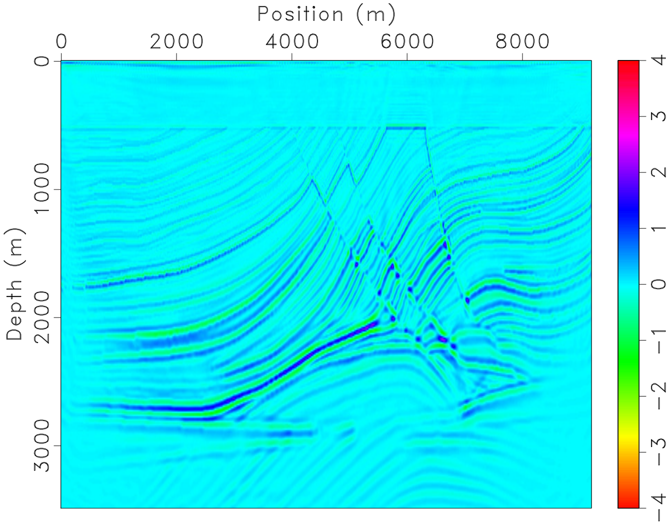
Alternative approach to improve RTM: solve linear least squares

$$J_{\mathrm{LSM}}[\delta\kappa] = \frac{1}{2}\|F\delta\kappa - \delta d\|^2$$

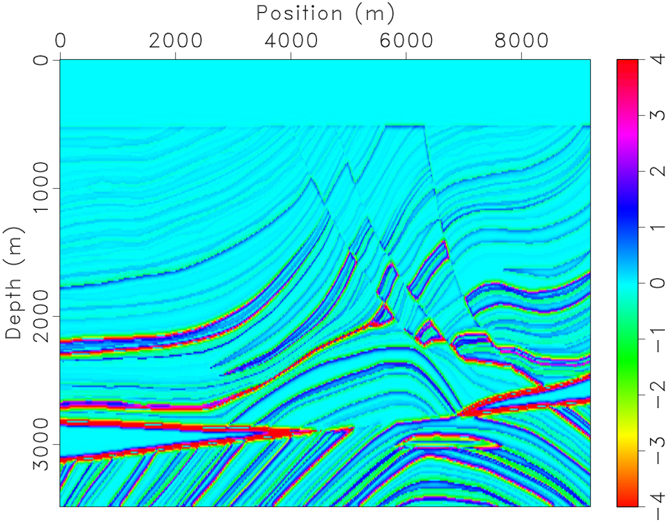
"Linearized inversion" (Lailly et al. 89, Chavent & Plessix 99), Least Squares Migration (LSM - Nemeth & Schuster 99, Kuehl & Sacchi 04,....)

Equivalent to linear system - too large for Gaussian Elimination, must use iterative method

Standard choice: conjugate gradient method or equivalent, each iteration = 1 F + 1 F^T (RTM) - expensive!!!



LSM - 20 CG iterations, data residual $\approx 40\%$



$\delta\kappa$ = model perturbation

Good news: compared to RTM image, LSM image has

- major reflectors still in right places
- more balanced amplitudes (that is, output is more similar to $\delta\kappa$)
- attenuates acquisition footprint, low frequency refraction noise
- accommodates any wave physics

SEG 17: multiple sessions on LSM (elastic, acoustic, Q, case studies,...)

Bad news:

- \$\$\$\$: every iteration costs 1 linearized forward map (F) and 1 RTM (F^T)

Agenda

- Seismic Data, Physics, Simulation
- Imaging and Asymptotic Inversion
- Linear least squares = Least squares migration'
- **Accelerating Linearized inversion (Least Squares Migration)**
- Accelerating Nonlinear (Full Waveform) Inversion
- Autofocusing: estimating reference/initial model
- Summary and Challenges

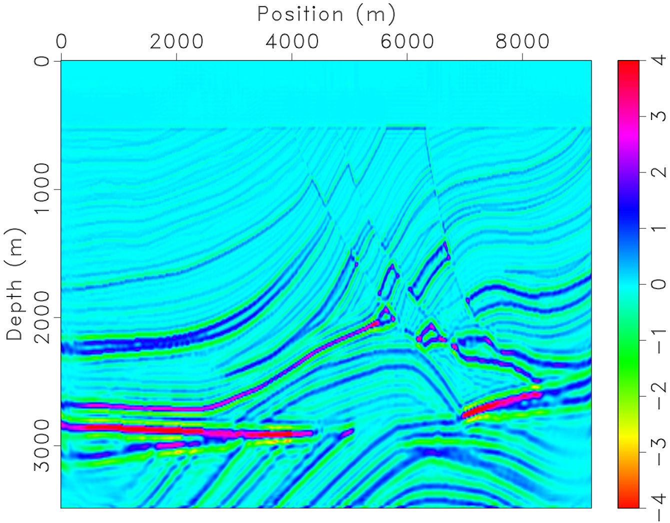
Recall

$$F^\dagger = W_m^{-1} F^T W_d, \;\; F^\dagger F \approx \text{ const. } I$$

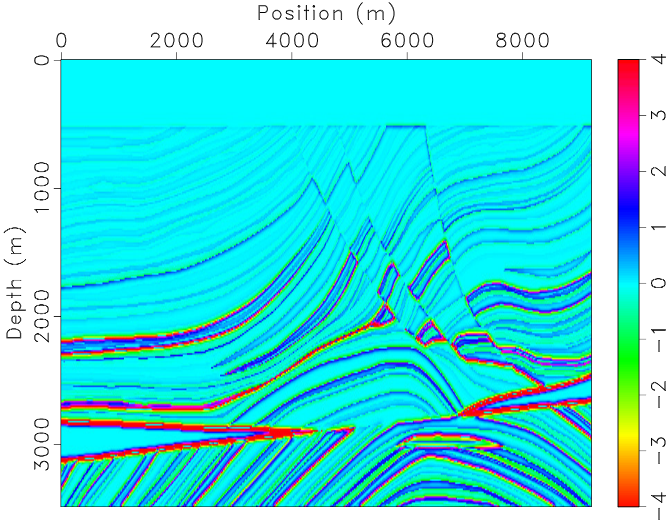
Some restrictions $\Rightarrow$ W_m, W_d are SPD $\Rightarrow$ $F^\dagger = W_m^{-1} F^T W_d$ is transpose of F in weighted norms $\Rightarrow$ F unitary modulo scale

Replace scalar products, transpose in CG with (W_m, W_d)-weighted versions $\sim$ preconditioned CG

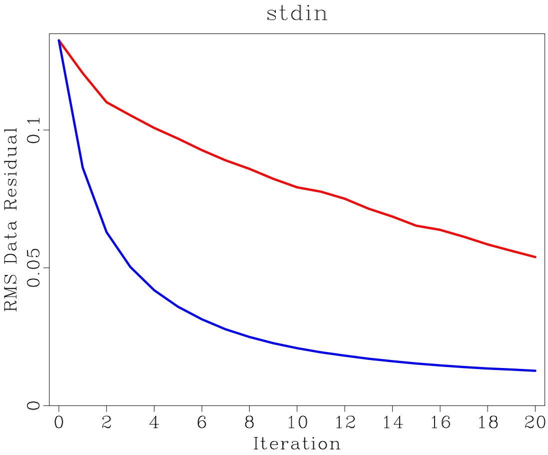
$\Rightarrow$ much faster convergence (Hou & S. SEG 16, EAGE 16, *Geophysics* 17)



LSM - 20 iterations PCG, data error < 10%



$\delta\kappa$ = model perturbation



Iteration number vs. data error: Red = CG, Blue = PCG

Agenda

Seismic Data, Physics, Simulation

Imaging and Asymptotic Inversion

Linear least squares = Least squares migration'

Accelerating Linearized inversion (Least Squares Migration)

Accelerating Nonlinear (Full Waveform) Inversion

Autofocusing: estimating reference/initial model

Summary and Challenges

LSM: given δd and κ_0, find $\delta\kappa$ to minimize $\sum_{\mathbf{x}_s,\mathbf{x}_r,t} |F[\kappa_0]\delta\kappa - \delta d|^2$

Full Waveform Inversion ("FWI"): given d, find κ to minimize $\sum_{\mathbf{x}_s,\mathbf{x}_r,t} |\mathcal{F}[\kappa] - d|^2$

LSM is linearized FWI: why not apply same tricks to FWI?

Both LSM and FWI require kinematically accurate model *a priori*: LSM makes no updates to background, FWI needs initial model predicting travel times to $< 1/2$ wavelength (Pratt 98)

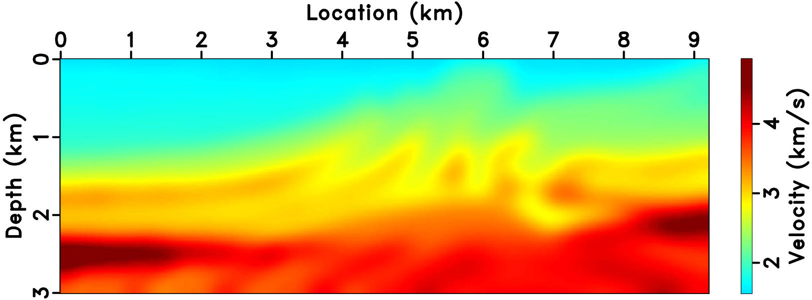
Standard industry choice: gradient descent (= steepest descent)

- compute FWI gradient $g = F[\kappa]^T(\mathcal{F}[\kappa] - d)$
- $\kappa \leftarrow \kappa - \alpha g$, α chosen by (very short!) line search

Gauss-Newton FWI algorithm (Ghattas et al 03, 09; Métivier et al 12, 14): replace F^T with $(F^T F)^{-1} F^T$ = LSM solution - fewer iterations, more reliable convergence, but requires inner iteration, much more expensive

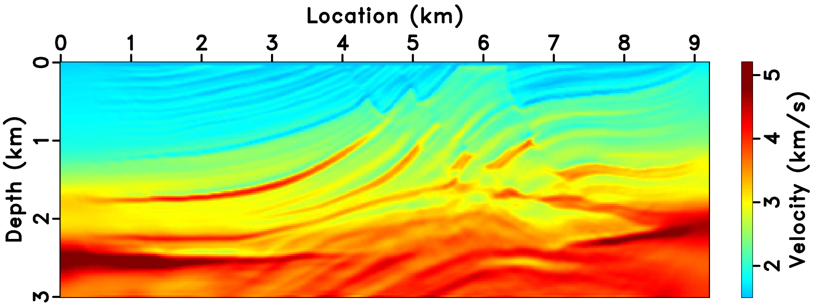
Hou & S. SEG 16: replace F^T by approximate inverse: search direction = $F[\kappa]^\dagger(\mathcal{F}[\kappa] - d)$ = A(pproximate) G(auss) N(ewton)

Cost of AGN step $\approx$ cost of FWI gradient descent step



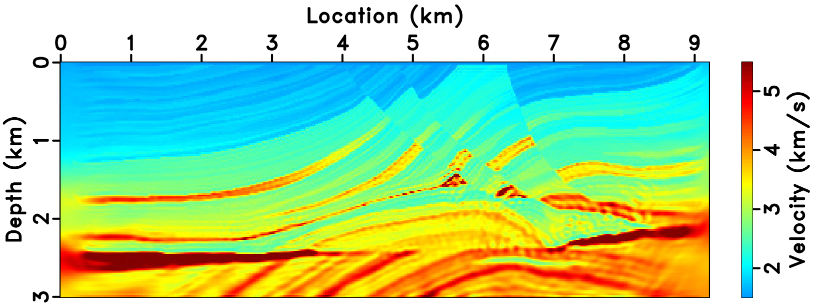
Initial model = smoothing of Marmousi model

Hou & S. SEG 16



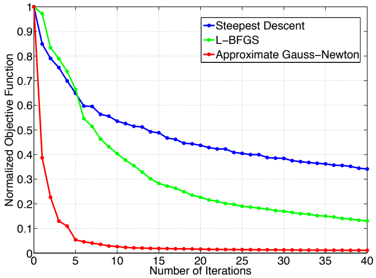
Velocity after 1 AGN iteration

Hou & S. SEG 16



Velocity after 40 AGN iterations (overkill!)

Hou & S. SEG 16



Data residual vs. Iteration: AGN (red), gradient descent (blue), L-BFGS (green)

Hou & S. SEG 16

Agenda

Seismic Data, Physics, Simulation

Imaging and Asymptotic Inversion

Linear least squares = Least squares migration'

Accelerating Linearized inversion (Least Squares Migration)

Accelerating Nonlinear (Full Waveform) Inversion

Autofocusing: estimating reference/initial model

Summary and Challenges

"It all depends on v(x,y,z)."

- J. Claerbout

Imaging/Inversion success depends on background (LSM) or initial (FWI) model (for acoustics: κ or $v = \sqrt{\kappa/\rho}$)

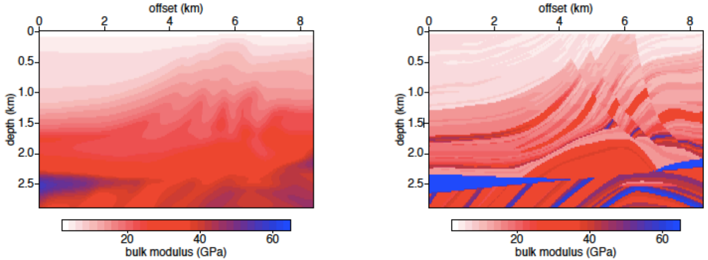
How to obtain? First idea: just use FWI!

Verdict: disaster (Gauthier, Tarantola, & Virieux 86)

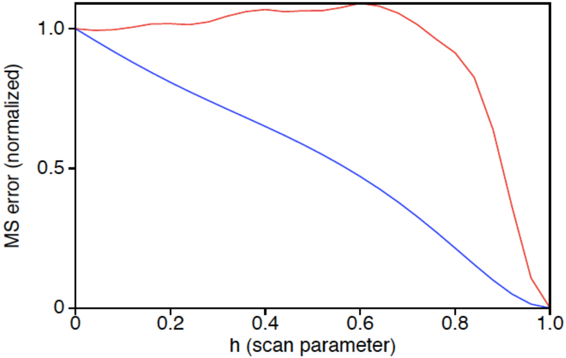
Visualizing the shape of the objective: *scan* from model m_0 to model m_1

$$f(h) = J_{\mathrm{FWI}}[(1-h)m_0 + hm_1]$$

Expl: data = simulation of Marmousi data (Versteeg & Grau 91), with bandpass filter source.



Marmousi bulk modulus: smoothed m_0, original m_1



Red: [2,5,40,50] Hz data. Blue: [2,4,8,12] Hz data

Diameter(domain of convexity) $\sim$ longest wavelength w/ good S/N

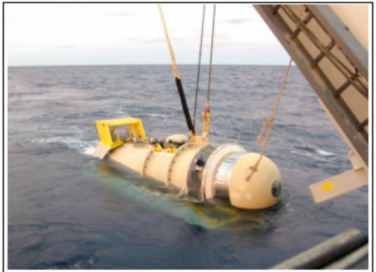


Figure 1: BP's Wolfspar® prototype low-frequency seismic source being retrieved during operational testing at the Gulf of Mexico systems integration test in 2014.

So collect low frequency data... (Thanks: Dellinger et al., SEG 16)

However - how low is “low”?

Zombie inversions: Plessix et al 10, successful inversion with lowest good freq = 1.5 Hz, failed inversion with 2.0 Hz

$\Rightarrow$ major theme in FWI research: how to make FWI robust against lack of low-frequency data, or equivalently lack of sufficiently accurate initial guess

$O(10)$ conceptually distinct approaches suggested in last decade

MS151 $=$ a sampling of attacks on the robustness problem

- Y. Yang, B. Engquist: replace L^2 norm with Wasserstein metric - larger region of convexity for complex, localized signals (also Métivier et al.)
- A. Mamonov et al.: extract projection of Green's function onto sampled time snapshots (reduced-order model) via block Cholesky, remove nonlinear effects, use other techniques developed for linearized data
- J. Zhai: strong restrictions on material parameter variation (piecewise constant) permit application of BC method, application to anisotropic elasticity
- WWS: model extension $=$ add non-physical parameters to κ etc., leverage data redundancy, extend F invertibly - "good" model trivializes (focuses) extension (very old idea, many variants)

also:

- traveltime inversion ("tomography")
- hybrid traveltime-waveform objectives, dissection of FWI gradient
- other data domains (Laplace, Fourier,...)
- Marchenko inversion (Green's function reconstruction based on reciprocity)
- band extrapolation via event identification (Demanet & Li, Warner et al.)
- band extrapolation via flux-corrected transport (Kalita & Alkhalifah)
- neural nets (see MS158, also geo literature)
- etc. etc.

see IPAM Spring Program 17 WS 2

Agenda

Seismic Data, Physics, Simulation

Imaging and Asymptotic Inversion

Linear least squares = Least squares migration'

Accelerating Linearized inversion (Least Squares Migration)

Accelerating Nonlinear (Full Waveform) Inversion

Autofocusing: estimating reference/initial model

Summary and Challenges

- How to see through rock: apply adjoint of linearized modeling operator F to "cooked" seismic data (RTM), iterate to improve data fit (LSM), accommodate nonlinear physics (FWI)
- Why it works: various restrictions $\Rightarrow$ $F \approx$ unitary with good choice of norms
- $\Rightarrow$ accelerate convergence of Krylov methods
- Key difficulty: how to chose background (LSM), initiate iteration (FWI)

Many ideas for estimating background/initial model, final verdict not yet in - very active field of research, see MS151 for sampling

Also incorporate higher fidelity physics in acceleration technology (beyond acoustics - elasticity, viscoelasticity,...)

Thanks

- TRIP students and postdocs: Jie Hou, Lei Fu, Yin Huang, Raanan Dafni, Kirk Blazek, Christiaan Stolk, many others
- TRIP sponsors for their long-term support
- Shell for support of Jie Hou and Raanan Dafni
- Fons ten Kroode, Hervé Chauris, Biondo Biondi,...
- Madagascar, Seismic Unix
- Rice University Research Computing Support Group
- Texas Advanced Computing Center, University of Texas, Austin
- SIAM 18 Organizers