Computational Methods for Inverse Problems in Geophysics

Russell J. Hewett
Mathematics & CMDA, Virginia Tech

Theory and Experience in Solving Inverse Problems in Geophysics Workshop
Uppsala University

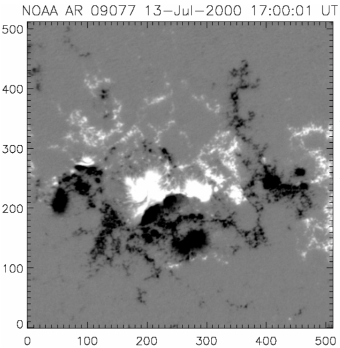
April 9, 2019

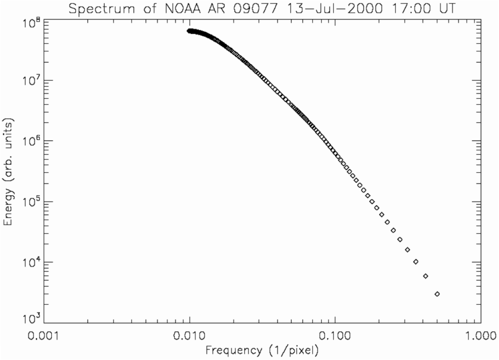
My Background

Research at the intersection of inverse problems, high-performance computation, and physics since ∼2003.

01-05 B.S. in Computer Science from Virginia Tech

- Thesis: Wavelet Analysis of Solar Active Regions
- Topics: Image processing and physical data extraction



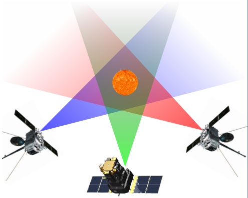


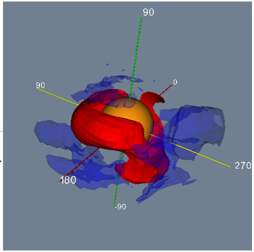
My Background

Research at the intersection of inverse problems, high-performance computation, and physics since ~2003.

05-11 Ph.D. in Computer Science with focus in Computational Science and Engineering from U. of Illinois

- Thesis: Numerical Methods for Solar Tomography in the STEREO Era
- Topics: Dynamic state estimation, constrained Kalman filtering, tomography of solar atmosphere, phase fields, image segmentation, geometrically constrained tomography, ray-tracing, tomography matrix calculation



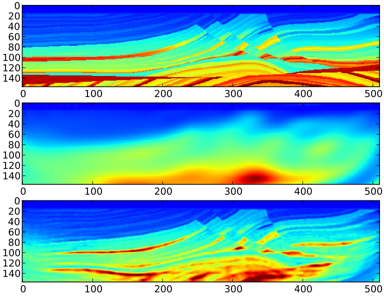


My Background

Research at the intersection of inverse problems, high-performance computation, and physics since ∼2003.

11-14 Postdoc in Mathematics (and Earth Science) at MIT

- Research: Seismic inversion, numerical optimization, wave equation solvers, numerical software design, www.pysit.org



My Background

Research at the intersection of inverse problems, high-performance computation, and physics since ∼2003.

14-18 Research Scientist and Project Manager at Total SA

- Research: High-performance seismic inversion, High-performance software design for CSE activities, geophysical inverse problems, uncertainty quantification, and machine learning

My Background

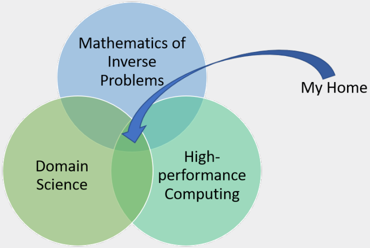
Research at the intersection of inverse problems, high-performance computation, and physics since ∼2003.

18- Assitant Professor in Mathematics and CMDA

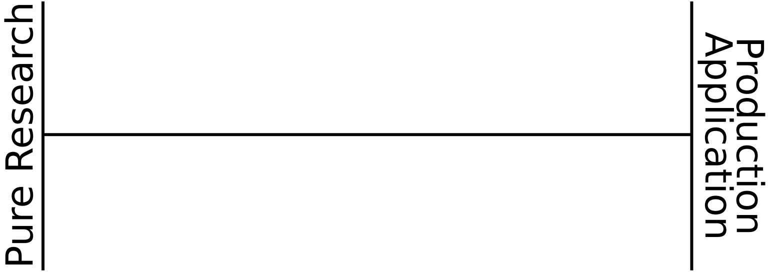
- Research: Machine Learning ∩ Physical Inverse Problems; Inverse Problems at Exascale

Characteristics of Interesting Problems

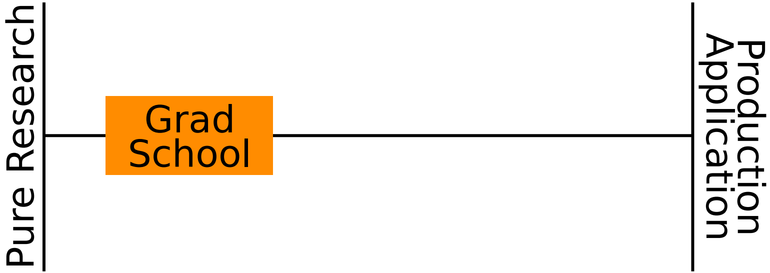
- Physics based
- Large, *noisy* real data sets
- Extremely large computation requirements



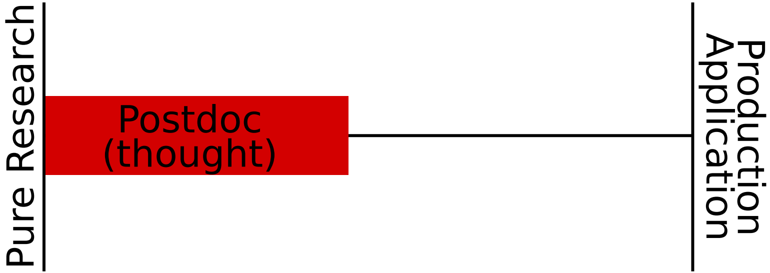
My Research Experience



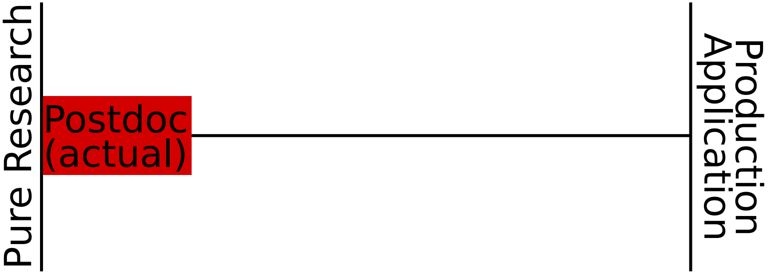
My Research Experience



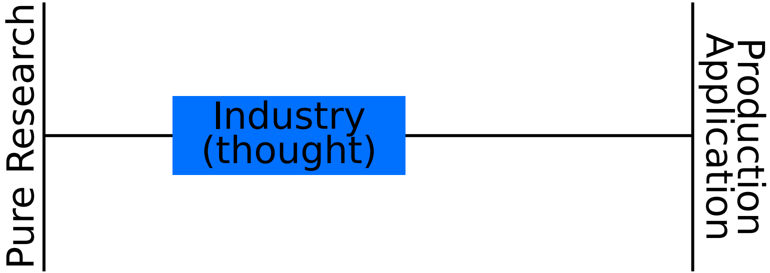
My Research Experience



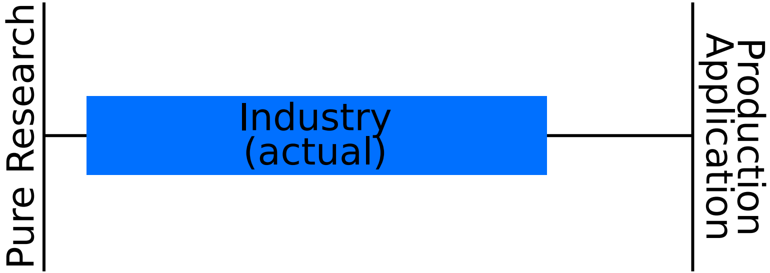
My Research Experience



My Research Experience



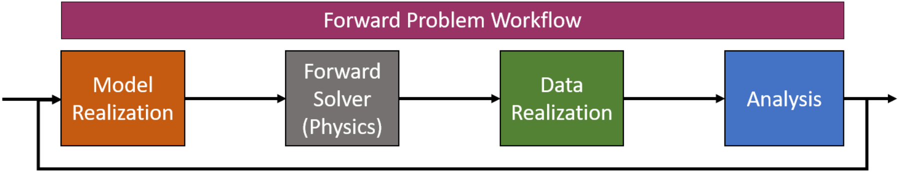
My Research Experience



Inverse Problems

Forward Problem

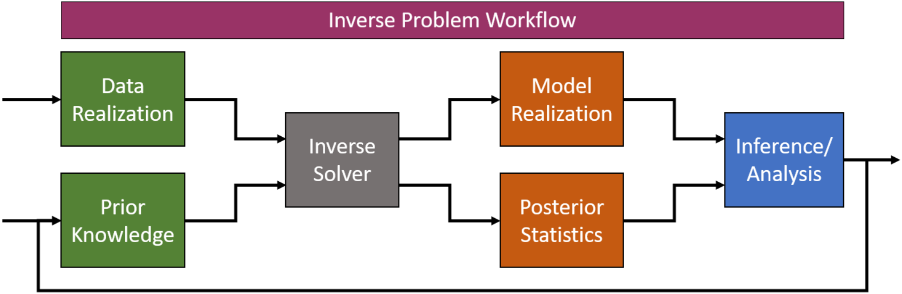
- Need high quality models, usually not empirical
- Need lots of compute
- One model $\Rightarrow$ One solution



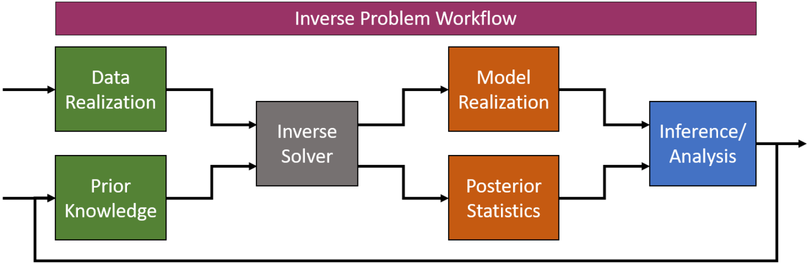
Inverse Problems

Inverse Problem

- Need high-quality data
- Need lots of compute (many fwd problems per inverse problem)
- One data set $\Rightarrow$ space of potential models



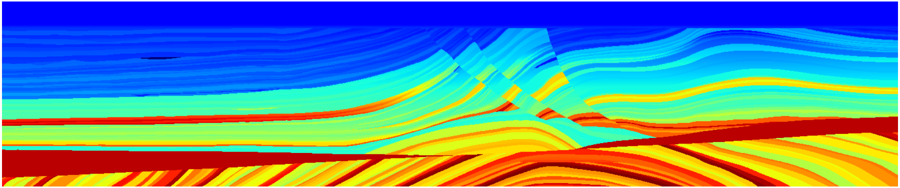
Subsurface Models



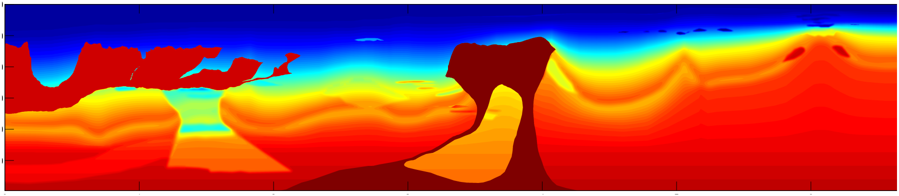
- Subsurface material parameters
  - P-wave velocity (acoustics)
  - S-wave velocity (elastics)
  - Density (both. . . neither?)
  - Anisotropy (VTI, TTI, orthorhombic)
- Subsurface reflectivity
  - Changes in material properties
  - Faults and fractures
- Temporal changes in the above (4D)

Subsurface Models

Marmousi 2 Velocity

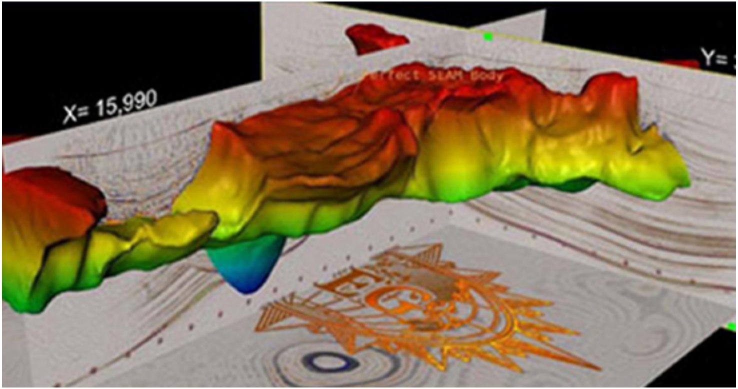


BP 2004 Velocity



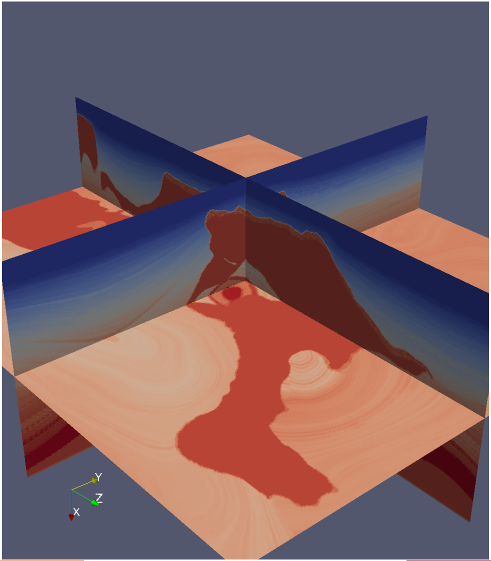
Subsurface Models

SEAM Phase I Velocity (Fehler; SEG)



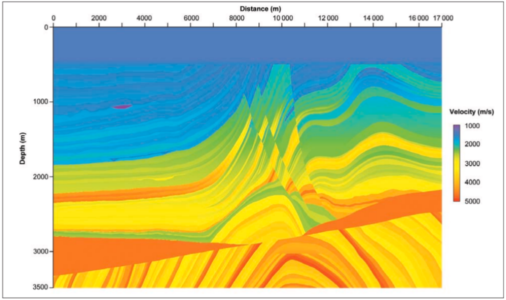
Subsurface Models

SEAM Phase I Velocity



Subsurface Models

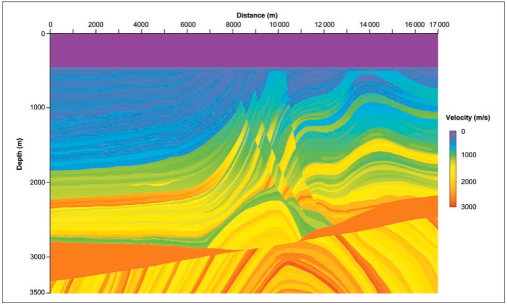
Marmousi 2



http://mcee.ou.edu/aaspi/publications/2006/martin_etal_TLE2006.pdf

Subsurface Models

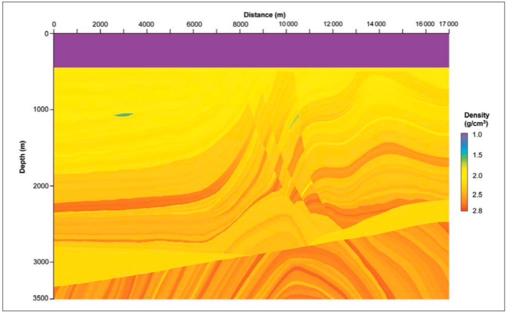
Marmousi 2



http://mcee.ou.edu/aaspi/publications/2006/martin_etal_TLE2006.pdf

Subsurface Models

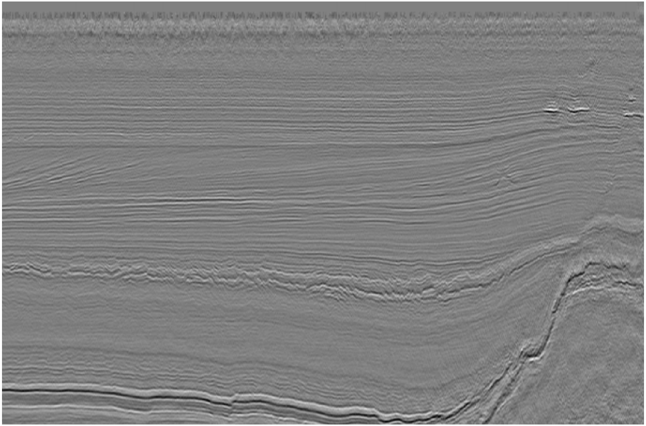
Marmousi 2



http://mcee.ou.edu/aaspi/publications/2006/martin_etal_TLE2006.pdf

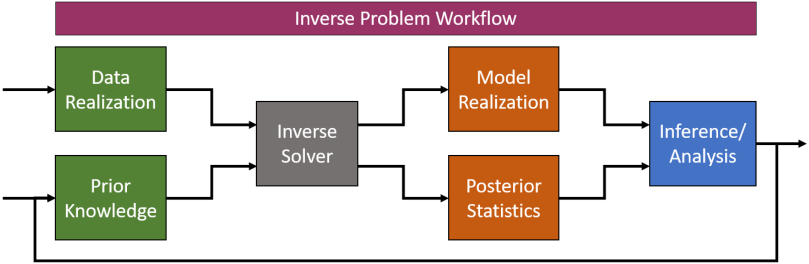
Subsurface Models

Netherlands Block F3 - Crossline 900



https://ghassanalregibdotcom.files.wordpress.com/2018/05/amir_aapg2018_slides.pdf

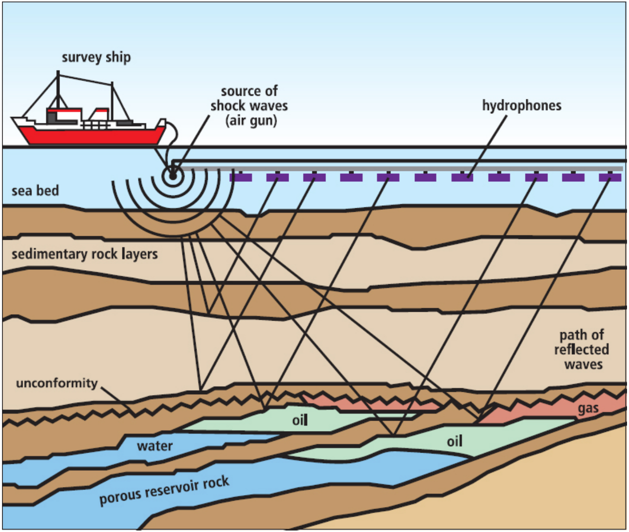
Geophysical Data



- Seismic Survey Design
  - Marine
  - Land
  - Exploration vs monitoring
- Seismic Sources
  - Airguns (marine)
  - Explosive (land)
  - Vibroseis (land, marine)
- Seismic Data Recorders
  - Hydrophones (marine)
  - Geophones (land)
  - DAS

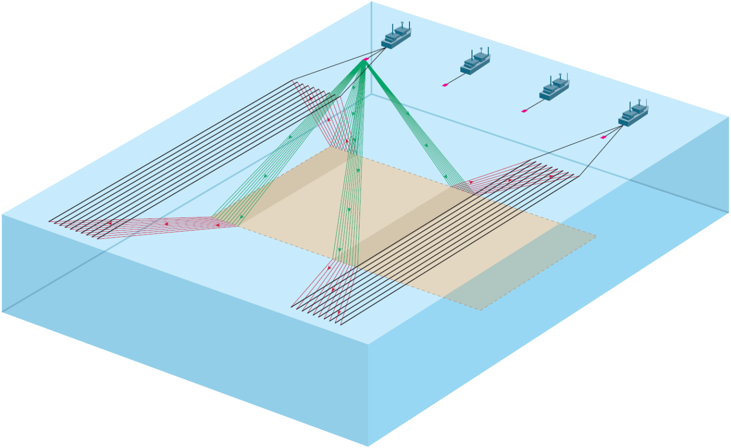
Exploration Seismic Data Acquisition: Design

Marine Survey



Exploration Seismic Data Acquisition: Design

Marine Survey

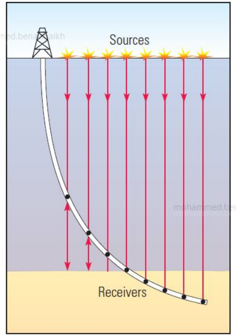


Wide-Azimuth

https://wiki.seg.org/wiki/Wide_azimuth#Wide-azimuth_acquisition_and_survey_design

Exploration Seismic Data Acquisition: Design

Vertical Seismic Profile (VSP) Survey

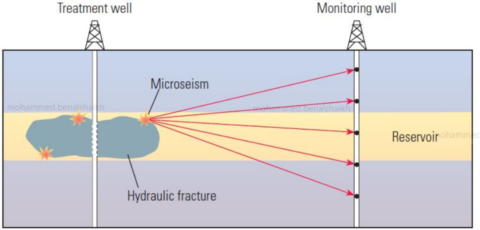


Walk-above

https://wiki.seg.org/wiki/Borehole_geophysics

Exploration Seismic Data Acquisition: Design

Vertical Seismic Profile (VSP) Survey

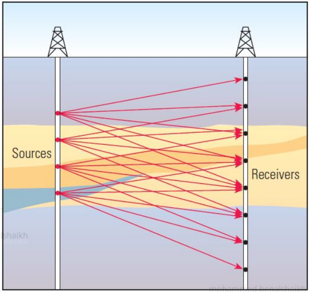


Microseismic

https://wiki.seg.org/wiki/Borehole_geophysics

Exploration Seismic Data Acquisition: Design

Vertical Seismic Profile (VSP) Survey

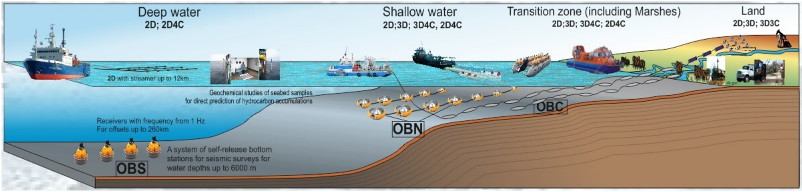


Cross-well

https://wiki.seg.org/wiki/Borehole_geophysics

Exploration Seismic Data Acquisition: Design

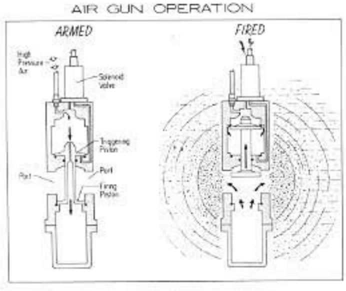
OBC/OBN/OBS Survey



http://marinegeos.com/services/1-seismic-survey

Exploration Seismic Data Acquisition: Sources

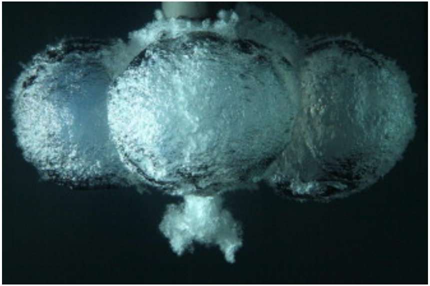
Marine: Airgun



https://woodshole.er.usgs.gov/operations/sfmapping/airgun.htm, Hutchinson and Detrick, 1984

Exploration Seismic Data Acquisition: Sources

Marine: Airgun



https://www.sciencedirect.com/science/article/pii/S0894177714000648

Exploration Seismic Data Acquisition: Sources

Marine: Airgun

https://www.soundingsonline.com/features/atlantic-boaters-may-soon-encounter-seismic-blasting-offshore-drilling

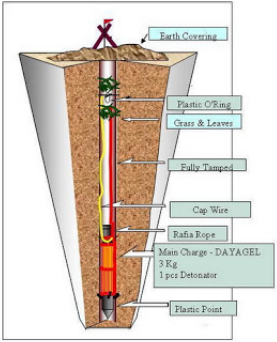
Exploration Seismic Data Acquisition: Sources

Land: VibroSeis

https://pixabay.com/en/vibrator-vibroseis-seismic-survey-863296/

Exploration Seismic Data Acquisition: Sources

Land: Explosive



https://vickybabel10.blogspot.com/2011/03/procedure-seismic-preloading.html

Exploration Seismic Data Acquisition: Sources

Marine: VibroSeis

https://library.seg.org/doi/pdf/10.1190/segam2016-13762702.1

Exploration Seismic Data Acquisition: Receivers

Land: Geophone

http://web.mit.edu/12.000/www/finalpresentation/experiments/geology.html

Exploration Seismic Data Acquisition: Receivers

Land: Geophone

http://web.mit.edu/12.000/www/finalpresentation/experiments/geology.html

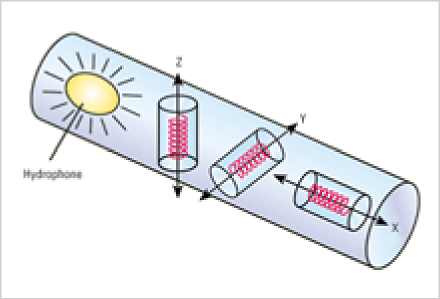
Exploration Seismic Data Acquisition: Receivers

Marine: Hydrophone

https://woodshole.er.usgs.gov/operations/sfmapping/hydrophone.htm

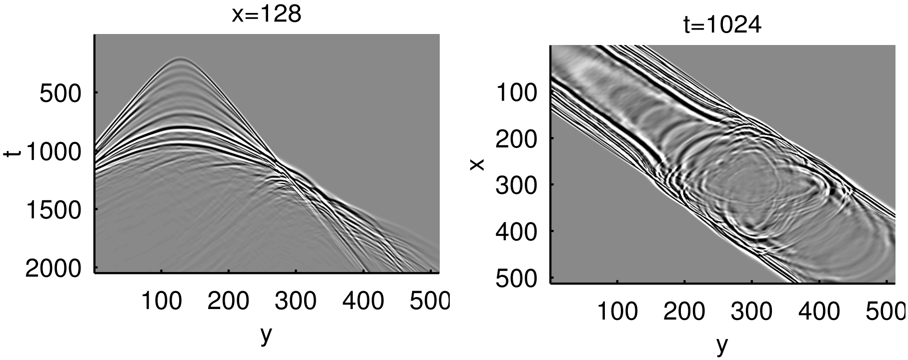
Exploration Seismic Data Acquisition: Receivers

4-Component Sensors (4C)



https://www.glossary.oilfield.slb.com/Terms/sym/4c_seismic_data.aspx

Exploration Seismic Data Acquisition



Geophysical Inverse Problems

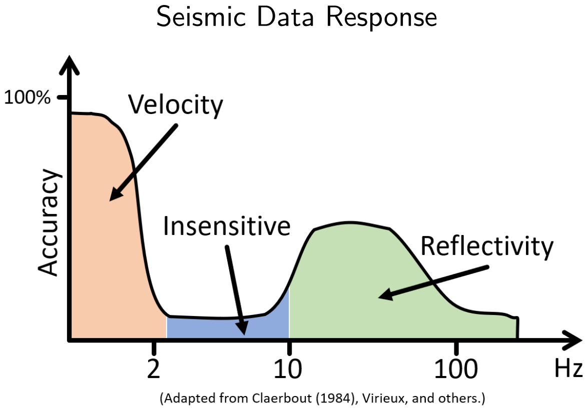
I am considering only the seismic inverse problem.

Regimes I have neglected:

- CSEM (Controlled Source Electromagnetic)
- Gravity
- LIDAR
- etc.

Geophysical Inverse Problems

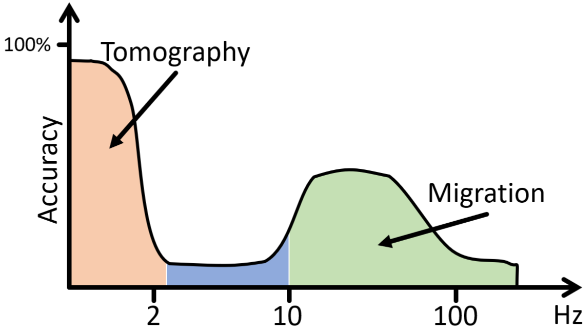
I will discuss only the full-waveform inversion problem and I will stay in an "exploration" context.



Geophysical Inverse Problems

I will discuss only the full-waveform inversion problem and I will stay in an "exploration" context.

Seismic Data Response

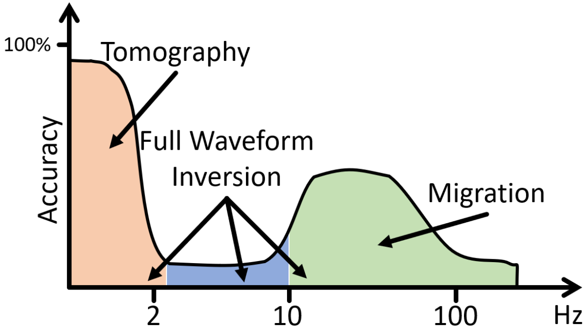


(Adapted from Claerbout (1984), Virieux, and others.)

Geophysical Inverse Problems

I will discuss only the full-waveform inversion problem and I will stay in an "exploration" context.

Seismic Data Response



(Adapted from Claerbout (1984), Virieux, and others.)

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

d: data

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

d: data

k: knowledge

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

d: data

k: knowledge

$\$$: money

(Exploration) Seismic Inverse Problem

$$\max \boldsymbol{\$}(d, k, \$)$$

d: data

k: knowledge

$\$$: money

$\boldsymbol{\$}$: more money

(Exploration) Seismic Inverse Problem

$$\max \boldsymbol{\$}(d, k, \$)$$

Subproblems:

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

Subproblems:

- $\arg\max_{d} \ \mathbf{\$}(d, k, \$)$
  - Find better data
  - Engineers and Analysts

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

Subproblems:

- $\arg\max_{d} \mathbf{\$}(d, k, \$)$
  - Find better data
  - Engineers and Analysts
- $\arg\max_{k} \mathbf{\$}(d, k, \$)$
  - Find better knowledge or use knowledge better
  - Research Scientists

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

Subproblems:

- $\arg\max_{d} \mathbf{\$}(d, k, \$)$
  - Find better data
  - Engineers and Analysts
- $\arg\max_{k} \mathbf{\$}(d, k, \$)$
  - Find better knowledge or use knowledge better
  - Research Scientists
- $\arg\min_{\$} \max \mathbf{\$}(d, k, \$)$
  - Do all this, but cheaper
  - Managers

(Exploration) Seismic Inverse Problem

$$\max \mathbf{\$}(d, k, \$)$$

Subproblems:

- $\arg\max_{d} \mathbf{\$}(d, k, \$)$
  - Find better data
  - Engineers and Analysts
- $\arg\max_{k} \mathbf{\$}(d, k, \$)$
  - Find better knowledge or use knowledge better
  - Research Scientists
- $\arg\min_{\$} \max \mathbf{\$}(d, k, \$)$
  - Do all this, but cheaper
  - Managers

Seismic Inverse Problem

Full Waveform Inversion Objective

$$J(m) = \|d(t) - \mathcal{F}(m(x))\|_2^2$$

$d(t)$: Data
$m(x)$: Unknown physical coefficients
$\mathcal{F}$: Modeling operator

Seismic Inverse Problem

FWI Objective: "Complete" Version

$$J(m, \boldsymbol{f}) = \sum_{s \in \mathcal{S}} \left\{ \| g(d_s) - g(S_s \mathcal{F}_s(\mathcal{R}_s(m), f_s)) \| + T_f(f_s) \right\} + T_m(m)$$

Seismic Inverse Problem

Full Waveform Inversion Objective

$$J(m) = \|d(t) - \mathcal{F}(m(x))\|_2^2$$

$d(t)$: Data
$m(x)$: Unknown physical coefficients
$\mathcal{F}$: Modeling operator

Seismic Inversion

Full Waveform Inversion Objective

$$J(m) = \|d(t) - \mathcal{F}(m(x))\|_2^2$$

$d(t)$: Data
$m(x)$: Unknown physical coefficients
$\mathcal{F}$: Modeling operator

Full Waveform Inversion Objective

$$J(m) = \|d(t) - \mathcal{F}(m(x))\|_2^2$$

$d(t)$: Data
$m(x)$: Unknown physical coefficients
$\mathcal{F}$: Modeling operator

Seismic Inversion

Full Waveform Inversion Problem

$$\arg\min_{m} \; J(m) = \|d(t) - \mathcal{F}(m(x))\|_2^2$$
$$\text{s.t. } L[m]u = f$$
$$\text{for } \mathcal{F}(m) = u$$

- $L[m]u = f$ is a wave equation
- $\mathcal{F}$ operator solves wave equations

- PDE constrained optimization!
  - Time-domain, frequency-domain, Laplace-domain, etc.

Modeling Operators

Example: Second Order Isotropic Acoustics (w/ Constant Density)

- $m(x) = 1/c^2(x)$, where $c(x)$ is p-wave velocity
- L is self adjoint
  - For continuous at least...
  - Up to BCs

$$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad (m_0\partial_{tt} - \triangle)u_0 = f$$

$$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad (m_0\partial_{tt} - \triangle)\delta u = -\delta m \partial_{tt} u_0$$

$$F_{m_0}^* r = \delta m \quad \Leftrightarrow \quad \begin{array}{l} \langle \delta m, -\langle q, \partial_{tt} u_0 \rangle_T \rangle_\Omega \\ \text{s.t.} \quad (m_0\partial_{tt} - \triangle)q = r \end{array}$$

$$\delta m = -\langle q, \partial_{tt} u_0 \rangle_T = -\int_0^T q(x,t)\partial_{tt} u_0(x,t)dt$$

Optimization Setup

$J(m) = \frac{1}{2}||d - \mathcal{F}(m)||_2^2$

- Objective function evaluation
- Computation: Solve wave equation
- Cost: ~ 1

Optimization Setup

$J(m) = \frac{1}{2}||d - \mathcal{F}(m)||_2^2$

- Objective function evaluation
- Computation: Solve wave equation
- Cost: ~ 1

$\nabla J(m_0) = -F_{m_0}^{*}(d - \mathcal{F}(m_0))$

- Objective gradient evalutation
- Computation: Adjoint state method
- Cost: $\sim 2+$

Optimization Setup

$$J(m) = \frac{1}{2}||d - \mathcal{F}(m)||_2^2$$

- Objective function evaluation
- Computation: Solve wave equation
- Cost: ~ 1

$$\nabla J(m_0) = -F_{m_0}^*(d - \mathcal{F}(m_0))$$

- Objective gradient evalutation
- Computation: Adjoint state method
- Cost: $\sim 2+$

$$D^2 J \delta m = F_{m_0}^* F_{m_0} \delta m - < D^2 \mathcal{F} \delta m, d - \mathcal{F}(m_0) >$$

- Objective Hessian application
- Computation: 2^{nd}-order adjoint state method
- Cost: $\sim 4+$

Modeling Operators

$$\mathcal{F}(m_0) = u_0 \qquad \Leftrightarrow \quad L[m_0]u_0 = f$$

- Forward modeling
- Cost: 1 wave solve

Modeling Operators

$$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad L[m_0]u_0 = f$$

- Forward modeling
- Cost: 1 wave solve

$$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad L[m_0]\delta u = -\frac{\delta L}{\delta m}[\delta m]u_0$$

- Linear forward modeling (Born)
- Cost: 2 wave solves
- Impossible to form as matrix!

Modeling Operators

$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad L[m_0]u_0 = f \qquad F \in \mathbb{R}^{m \times n}$

- Forward modeling
- Cost: 1 wave solve

$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad L[m_0]\delta u = -\frac{\delta L}{\delta m}[\delta m]u_0$

- Linear forward modeling (Born)
- Cost: 2 wave solves
- Impossible to form as matrix!

Modeling Operators

$$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad L[m_0]u_0 = f$$

- Forward modeling
- Cost: 1 wave solve

$$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad L[m_0]\delta u = -\frac{\delta L}{\delta m}[\delta m]u_0$$

- Linear forward modeling (Born)
- Cost: 2 wave solves
- Impossible to form as matrix!

$F \in \mathbb{R}^{m \times n}$

3D Survey

- 10×1000 rcv (small)
- 8s recording (short)
- 8ms sampling (long)
- m =10,000,000 samples

Modeling Operators

$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad L[m_0]u_0 = f$

- Forward modeling
- Cost: 1 wave solve

$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad L[m_0]\delta u = -\frac{\delta L}{\delta m}[\delta m]u_0$

- Linear forward modeling (Born)
- Cost: 2 wave solves
- Impossible to form as matrix!

$F \in \mathbb{R}^{m \times n}$

3D Survey

- 10×1000 rcv (small)
- 8s recording (short)
- 8ms sampling (long)
- $m =$10,000,000 samples

3D Modeling

- $876 \times 1001 \times 750$ dof
- $n =$ 657,657,000 dof

Modeling Operators

$$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad L[m_0]u_0 = f$$

- Forward modeling
- Cost: 1 wave solve

$$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad L[m_0]\delta u = -\frac{\delta L}{\delta m}[\delta m]u_0$$

- Linear forward modeling (Born)
- Cost: 2 wave solves
- Impossible to form as matrix!

$F \in \mathbb{R}^{m \times n}$

3D Survey

- 10×1000 rcv (small)
- 8s recording (short)
- 8ms sampling (long)
- $m =$10,000,000 samples

3D Modeling

- $876 \times 1001 \times 750$ dof
- $n =$ 657,657,000 dof

Matrix Size

- IEEE single precision. . .
- $\sim$ 23.4PB Storage
- n wave solves
- $\sim$18 years
  @ 100k/day

Modeling Operators

$$\mathcal{F}(m_0) = u_0 \quad \Leftrightarrow \quad L[m_0]u_0 = f$$

- Forward modeling
- Cost: 1 wave solve

$$F_{m_0}\delta m = \delta u \quad \Leftrightarrow \quad L[m_0]\delta u = -\frac{\delta L}{\delta m}[\delta m]u_0$$

- Linear forward modeling (Born)
- Cost: 2 wave solves
- Impossible to form as matrix!

$$F^*_{m_0} r = \delta m \quad \Leftrightarrow \quad \begin{array}{c} \left\langle q, -\frac{\delta L}{\delta m}[\delta m]u_0 \right\rangle_{\Omega\times T} \\ \text{s.t.} \quad L^*[m_0]q = r \end{array}$$

- Adjoint modeling
- "Migration" or imaging operator
- Cost: $1+$ wave solves

$F \in \mathbb{R}^{m\times n}$

3D Survey

- 10×1000 rcv (small)
- 8s recording (short)
- 8ms sampling (long)
- $m =$10,000,000 samples

3D Modeling

- $876 \times 1001 \times 750$ dof
- $n =$ 657,657,000 dof

Matrix Size

- IEEE single precision…
- $\sim$ 23.4PB Storage
- n wave solves
- $\sim$18 years @ 100k/day

Solving FWI

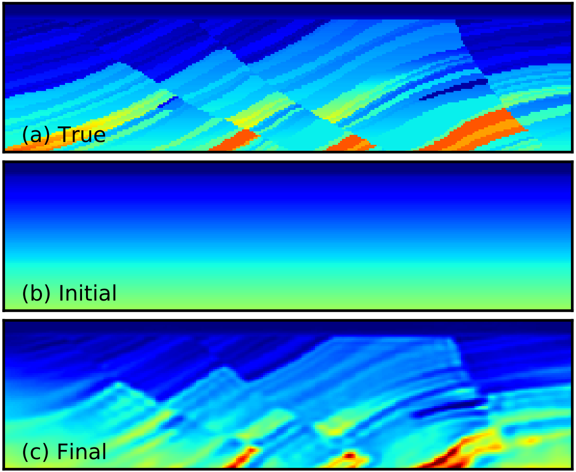
- Solve with gradient-based optimization

Generalized Gradient Scheme

1. Given m_0.
2. While $i < \texttt{MaxIter}$
   - 2.1 $g_i = \nabla J(m_i)$
   - 2.2 $s_i = h(m_i, g_i)$
   - 2.3 $\alpha = \texttt{LineSearch}(m_i, g_i, s_i)$
   - 2.4 $m_{i+1} = m_i + \alpha s_i$

- Gradient descent?
- L-BFGS?
- Hessian/Quasi-Newton schemes?

Solving FWI



- 50 L-BFGS iterations w/ Locally 1D Time Solver (L. Zepeda, RJH, M. Rao, L. Demanet (SEG 2013))

Real-world Utility

- Interpreters don't actually use inverted material parameters
- They still want to look at seismic "images"
- Found by "migration" or imaging algorithms
    - Kirchoff migration
    - (One-way) wave equation migration
    - Reverse-time migration
    - Linearized inversion

Real-world Utility

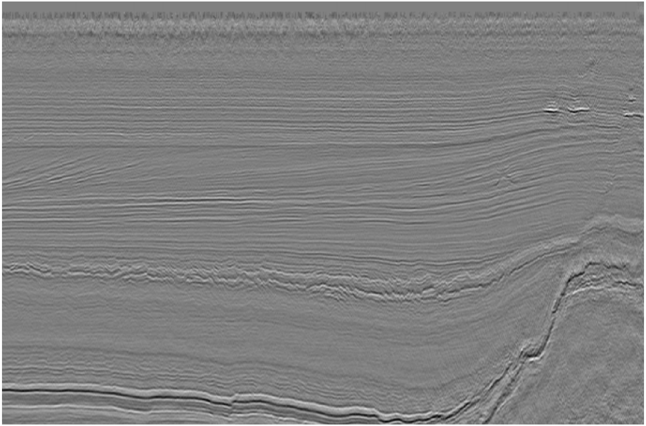
- Interpreters don't actually use inverted material parameters
- They still want to look at seismic "images"
- Found by "migration" or imaging algorithms
  - Kirchoff migration
  - (One-way) wave equation migration
  - Reverse-time migration
  - Linearized inversion
- Migration is essentially back-propagation
- Or, we can consider it as a gradient calculation in FWI

Real-world Utility

- Interpreters don't actually use inverted material parameters
- They still want to look at seismic "images"
- Found by "migration" or imaging algorithms
  - Kirchoff migration
  - (One-way) wave equation migration
  - Reverse-time migration
  - Linearized inversion
- Migration is essentially back-propagation
- Or, we can consider it as a gradient calculation in FWI
- In any case, it is computed at much higher-frequency
- Looking at reflectivity, not material parameters

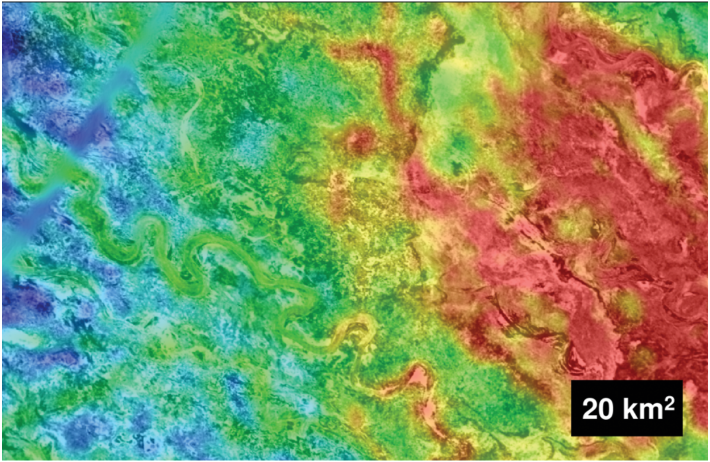
Seismic Image

Netherlands Block F3 - Crossline 900



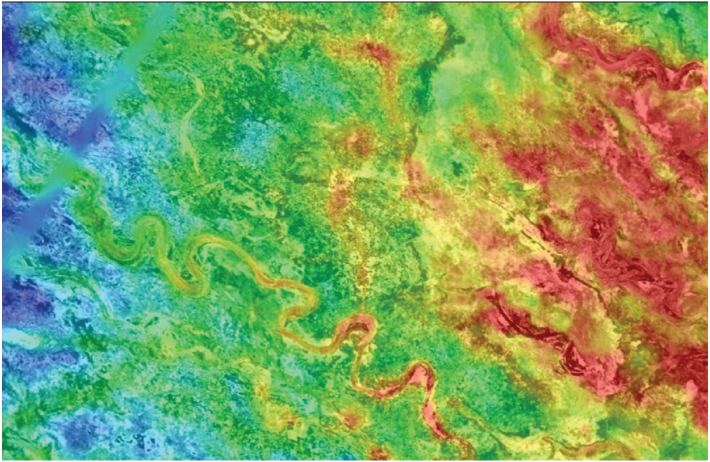
https://ghassanalregibdotcom.files.wordpress.com/2018/05/amir_aapg2018_slides.pdf

Real World Results



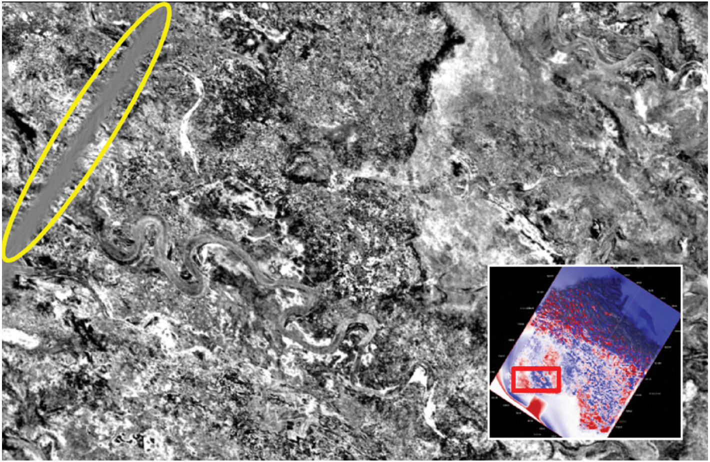
"An offshore Gabon full-waveform inversion case study," Xiao, et al., *Interpretation*, November 2016. Data from CGG.

Real World Results



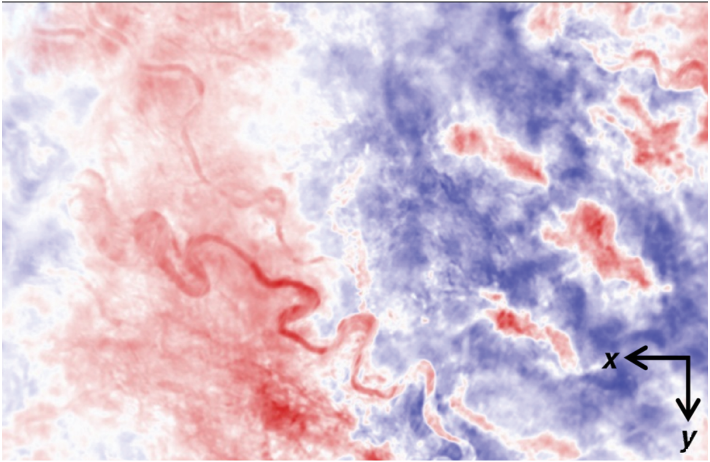
"An offshore Gabon full-waveform inversion case study," Xiao, et al., *Interpretation*, November 2016. Data from CGG.

Real World Results



"An offshore Gabon full-waveform inversion case study," Xiao, et al., *Interpretation*, November 2016. Data from CGG.

Real World Results

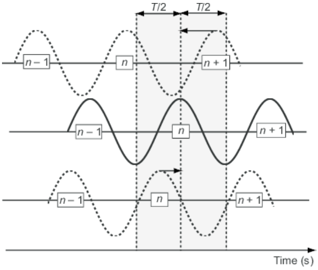


"An offshore Gabon full-waveform inversion case study," Xiao, et al., *Interpretation*, November 2016. Data from CGG.

Challenges for FWI

- Due to the physical formulation
- Due to mathematical formulation
- Due to computational requirements
- Due to business decisions

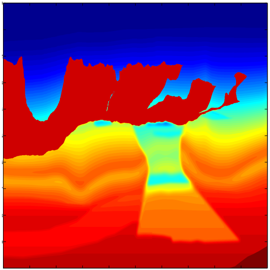
Cycle Skipping



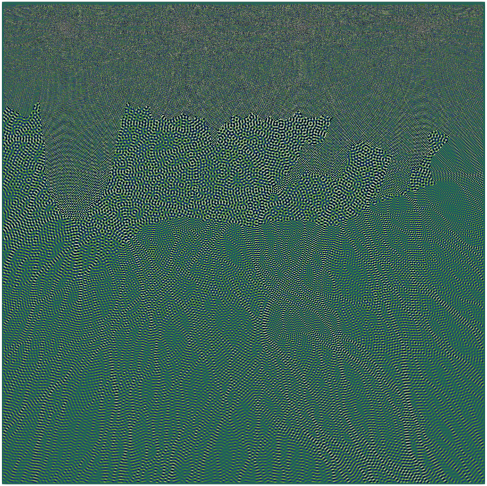
Jean Virieux

- Due to the physical and mathematical formulations
- Manifests as global nonconvexity
- Partially resolved by working in frequency domain (or in Laplace domain)
- Time-domain wave equation becomes Helmholtz equation

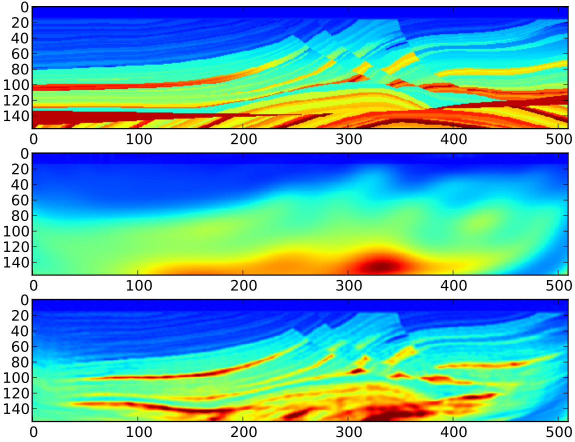
FWI In Frequency Domain



- max. 512 wavelengths in domain
- PML width: 2.5 wavelengths
- 64×64 domain decomposition



FWI In Frequency Domain



- Frequency continuation

Realistic and Practical Physics

- I have shown the *idea* behind FWI using constant density acoustic physics
- Of course, the earth is not constant density, nor acoustic
- Massive increase in computational – and software – costs
- Does better physics drive the need for more compute?
- Or is more compute driving the availability of better physics?

Realistic and Practical Physics

Physical Models

$$\partial_{tt}u = a\triangle u + f$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–

Realistic and Practical Physics

Physical Models

$$\partial_t \begin{bmatrix} p \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} & -c\nabla\cdot \\ a\nabla & \end{bmatrix} \begin{bmatrix} p \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} g \\ \boldsymbol{h} \end{bmatrix}$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–
Iso-Aco (1st)	4	2	1x-2x

Realistic and Practical Physics

Physical Models

$$\partial_t \begin{bmatrix} \tilde{\boldsymbol{\sigma}} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} & -\boldsymbol{A}^T \boldsymbol{D}^* \boldsymbol{R} \\ a\boldsymbol{R}^T \boldsymbol{D} \boldsymbol{A} & \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\sigma}} \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{h} \end{bmatrix}$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–
Iso-Aco (1st)	4	2	1x-2x
TTI-Aco (1st)	5 (+?)	4	3x

Realistic and Practical Physics

Physical Models

$$\partial_t \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} & -\mathbb{C}\boldsymbol{D}^* \\ a\boldsymbol{D} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{h} \end{bmatrix}$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–
Iso-Aco (1st)	4	2	1x-2x
TTI-Aco (1st)	5 (+?)	4	3x
Iso-Ela (1st)	9	3	$\sim$ 40x

Realistic and Practical Physics

Physical Models

$$\partial_t \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} & -\mathbb{C}\boldsymbol{D}^* \\ a\boldsymbol{D} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{h} \end{bmatrix}$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–
Iso-Aco (1st)	4	2	1x-2x
TTI-Aco (1st)	5 (+?)	4	3x
Iso-Ela (1st)	9	3	$\sim$ 40x
VTI-Ela (1st)	9	8	$\sim$ 40x

Realistic and Practical Physics

Physical Models

$$\partial_t \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} & -\mathbb{C}\boldsymbol{D}^* \\ a\boldsymbol{D} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{h} \end{bmatrix}$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–
Iso-Aco (1st)	4	2	1x-2x
TTI-Aco (1st)	5 $(+?)$	4	3x
Iso-Ela (1st)	9	3	$\sim$ 40x
VTI-Ela (1st)	9	8	$\sim$ 40x
TTI-Ela (1st)	9	36	$>$ 100x

Realistic and Practical Physics

Physical Models

$$\partial_t \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} & -\mathbb{C}\boldsymbol{D}^* \\ a\boldsymbol{D} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{h} \end{bmatrix}$$

Physics	Solutions	Parameters	Computation
Iso-Aco (const. ρ) (2nd)	1	1	–
Iso-Aco (1st)	4	2	1x-2x
TTI-Aco (1st)	5 (+?)	4	3x
Iso-Ela (1st)	9	3	$\sim$ 40x
VTI-Ela (1st)	9	8	$\sim$ 40x
TTI-Ela (1st)	9	36	$>$ 100x
Elastic (1st)	9	81	$>$ 100x

Software Considerations

- I have greatly simplified the problem
- Only considering one simple PDE and an "academic" objective function
- My personal results in this talk are only embarassingly parallel
- Essentially data parallel in shot record
- There is still model parallelism to exploit, which requires high-end HPC software
- In addition to. . .

Software Considerations

FWI Objective

$$J(m) = \|d - \mathcal{F}(m)\|_2^2$$

$d(t)$:	Data
$m(x)$:	Unknown physical coefficients
$\mathcal{F}$:	Modeling operator

Software Considerations

FWI Objective

$$J(m) = \sum_{s \in \mathcal{S}} \|d_s - \mathcal{F}_s(m)\|_2^2$$

$d_s(t)$: Data for shot s
$m(x)$: Unknown physical coefficients
$\mathcal{F}_s$: Modeling operator for shot s

Software Considerations

FWI Objective

$$J(m) = \sum_{s \in \mathcal{S}} \| d_s - S_s \mathcal{F}_s(m) \|_2^2$$

$d_s(t)$: Data for shot s
$m(x)$: Unknown physical coefficients
$\mathcal{F}_s$: Modeling operator for shot s

S_s: Data sampling operator for shot s

Software Considerations

FWI Objective

$$J(m) = \sum_{s \in \mathcal{S}} \| d_s - S_s \mathcal{F}_s(m) \|$$

$d_s(t)$: Data for shot s
$m(x)$: Unknown physical coefficients
$\mathcal{F}_s$: Modeling operator for shot s

S_s: Data sampling operator for shot s
$\|\cdot\|$: (Arbitrary) residual norm

Software Considerations

FWI Objective

$$J(m, \boldsymbol{f}) = \sum_{s \in \mathcal{S}} \| d_s - S_s \mathcal{F}_s(m, f_s) \|$$

- $d_s(t)$: Data for shot s
- $m(x)$: Unknown physical coefficients
- $\mathcal{F}_s$: Modeling operator for shot s

- S_s: Data sampling operator for shot s
- $\|\cdot\|$: (Arbitrary) residual norm
- $f_s(x, t)$: Unknown seismic source function for shot s

Software Considerations

FWI Objective

$$J(m, \boldsymbol{f}) = \sum_{s \in \mathcal{S}} \| g(d_s) - g(S_s \mathcal{F}_s(m, f_s)) \|$$

$d_s(t)$:	Data for shot s
$m(x)$:	Unknown physical coefficients
$\mathcal{F}_s$:	Modeling operator for shot s
S_s:	Data sampling operator for shot s
$\|\cdot\|$:	(Arbitrary) residual norm
$f_s(x, t)$:	Unknown seismic source function for shot s
g:	Data filter and interpolation function

Software Considerations

FWI Objective

$$J(m, \boldsymbol{f}) = \sum_{s \in \mathcal{S}} \| g(d_s) - g(S_s \mathcal{F}_s(\mathcal{R}_s(m), f_s)) \|$$

$d_s(t)$:	Data for shot s
$m(x)$:	Unknown physical coefficients
$\mathcal{F}_s$:	Modeling operator for shot s
S_s:	Data sampling operator for shot s
$\|\cdot\|$:	(Arbitrary) residual norm
$f_s(x,t)$:	Unknown seismic source function for shot s
g:	Data filter and interpolation function
$\mathcal{R}_s$:	Model restriction and filter operator for shot s

Software Considerations

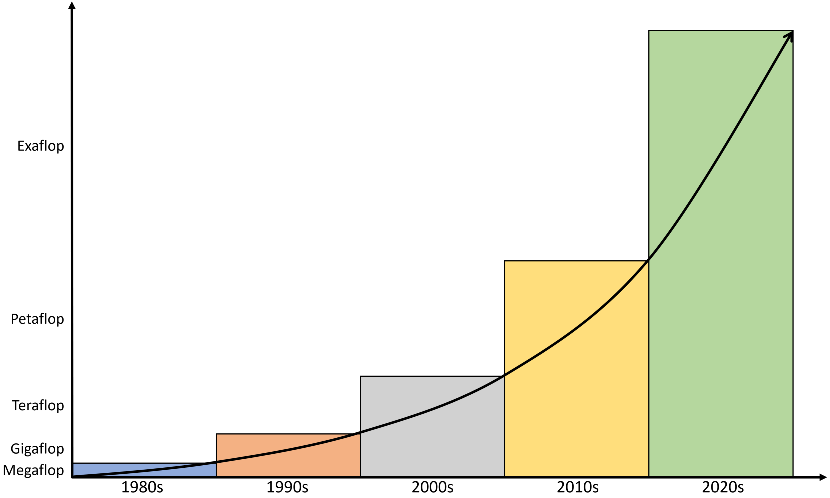
FWI Objective

$$J(m, \boldsymbol{f}) = \sum_{s \in \mathcal{S}} \{\|g(d_s) - g(S_s \mathcal{F}_s(\mathcal{R}_s(m), f_s))\| + T_f(f_s)\} + T_m(m)$$

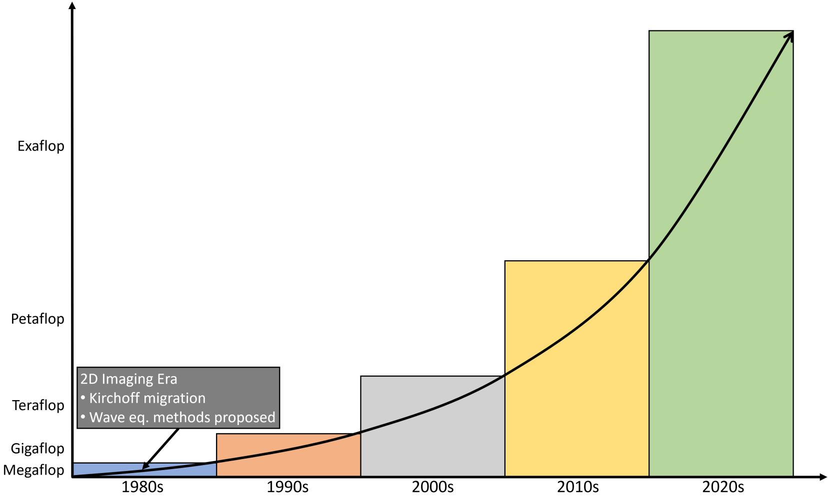
$d_s(t)$: Data for shot s
$m(x)$: Unknown physical coefficients
$\mathcal{F}_s$: Modeling operator for shot s

S_s: Data sampling operator for shot s
$\|\cdot\|$: (Arbitrary) residual norm
$f_s(x,t)$: Unknown seismic source function for shot s
g: Data filter and interpolation function
$\mathcal{R}_s$: Model restriction and filter operator for shot s
T: Regularization

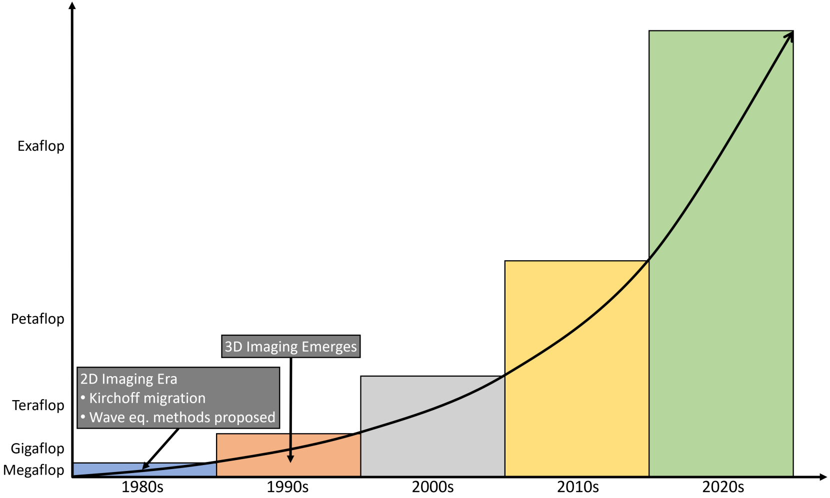
HPC and Subsurface Inverse Problems



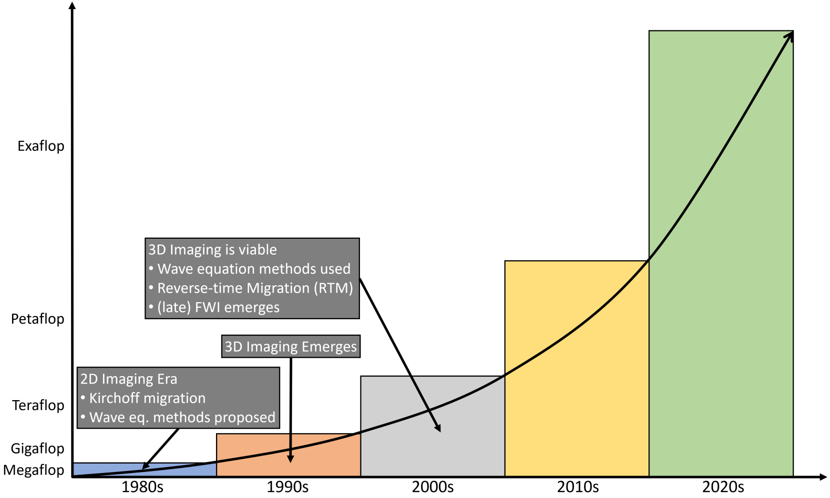
HPC and Subsurface Inverse Problems



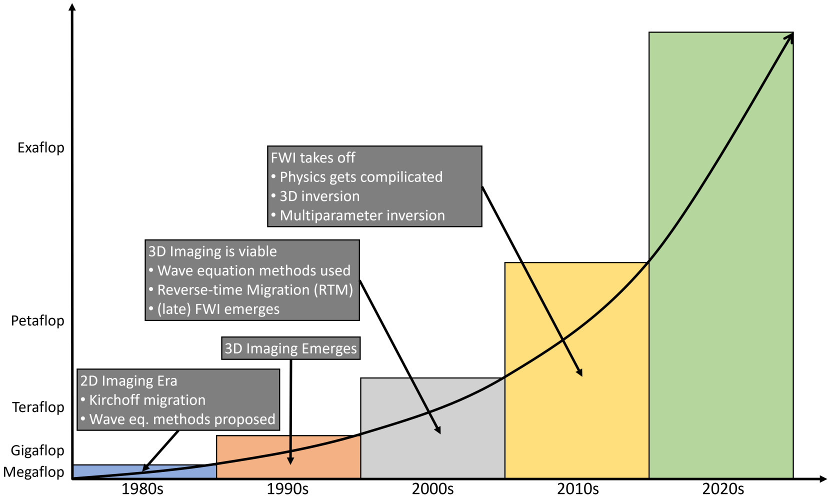
HPC and Subsurface Inverse Problems



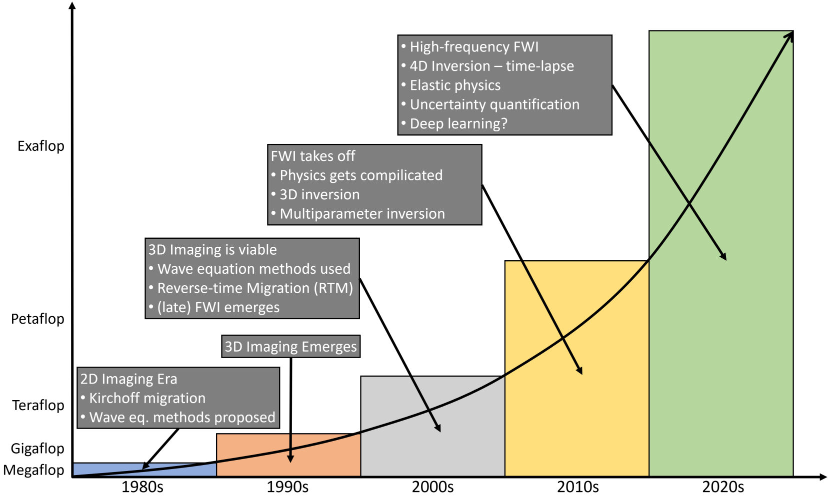
HPC and Subsurface Inverse Problems



HPC and Subsurface Inverse Problems

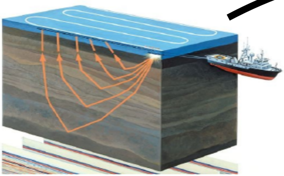


HPC and Subsurface Inverse Problems



Scale: Business Considerations

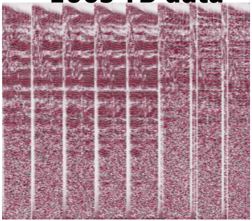
100s M$



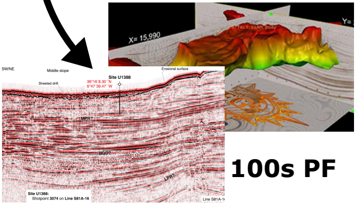
1000s km^2 surface acquisition

Acquire Seismic Data

100s TB data



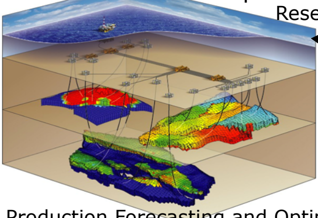
Seismic Depth Imaging



100s PF

???s Gboe ???? M$ Investment

Reservoir Appraisal



Production Forecasting and Optimization

Current / Future Challenges: Data and Acquisition

Field data

- Exploit high redundancy in field data (compression)
- "Chain of custody" in field data
- Real data is noisy!
- Multicomponent data (more than just pressure)

Current / Future Challenges: Data and Acquisition

Field data

- Exploit high redundancy in field data (compression)
- "Chain of custody" in field data
- Real data is noisy!
- Multicomponent data (more than just pressure)

Novel Acquisitions

- Marine: Coil surveys, multiple sources
- Marine vibrators
- Pseudo-random source/receiver geometry
- Survey refinement

Current / Future Challenges: PDEs and Inverse Problems

Inverse Problems

- Source function is unknown
- Meshing complex topography (seafloor, salts)
- Uncertainty quantification

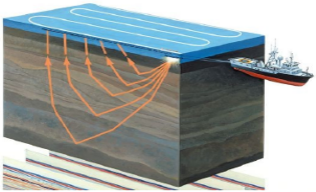
PDEs

- Boundary conditions!
- Beyond finite difference

Current / Future Challenges: PDEs and Inverse Problems

Inverse Problems

- Source function is unknown
- Meshing complex topography (seafloor, salts)
- Uncertainty quantification



PDEs

- Boundary conditions!
- Beyond finite difference

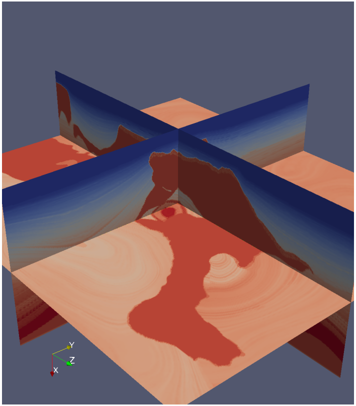
Current / Future Challenges: PDEs and Inverse Problems

Inverse Problems

- Source function is unknown
- Meshing complex topography (seafloor, salts)
- Uncertainty quantification

PDEs

- Boundary conditions!
- Beyond finite difference



Current / Future Challenges: PDEs and Inverse Problems

Inverse Problems

- Source function is unknown
- Meshing complex topography (seafloor, salts)
- Uncertainty quantification

PDEs

- Boundary conditions!
- Beyond finite difference

Current / Future Challenges: PDEs and Inverse Problems

Inverse Problems

- Source function is unknown
- Meshing complex topography (seafloor, salts)
- Uncertainty quantification

PDEs

- Boundary conditions!
- Beyond finite difference

- Perfectly Matched Layers
- Anisotropy
- High-order ABCs

Current / Future Challenges: PDEs and Inverse Problems

Inverse Problems

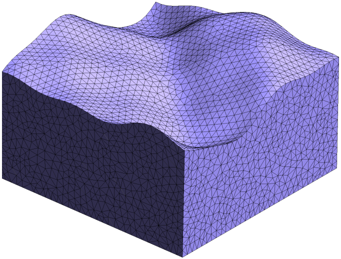
- Source function is unknown
- Meshing complex topography (seafloor, salts)
- Uncertainty quantification

PDEs

- Boundary conditions!
- Beyond finite difference

Example: Discontinuous Galerkin Solvers

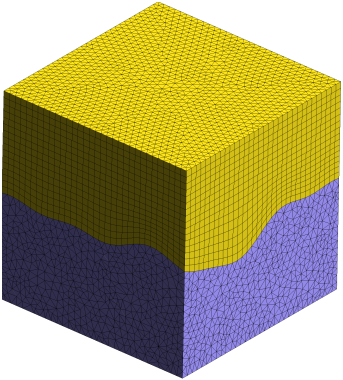
- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



J. Chan, Z. Wang, RJH, T. Warburton (2016)

Example: Discontinuous Galerkin Solvers

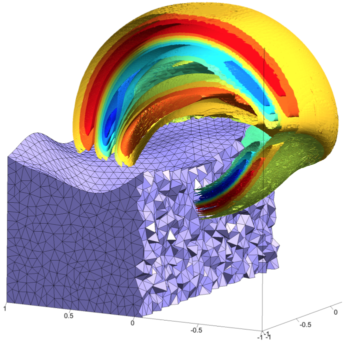
- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



J. Chan, Z. Wang, RJH, T. Warburton (2016)

Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



J. Chan, Z. Wang, RJH, T. Warburton (2016)

Example: Discontinuous Galerkin Solvers

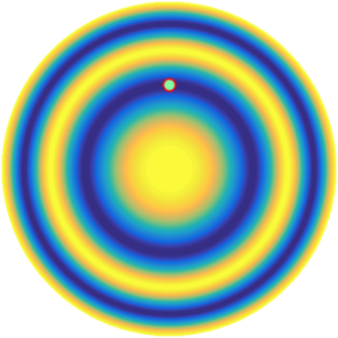
- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)

Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)

Example: Discontinuous Galerkin Solvers

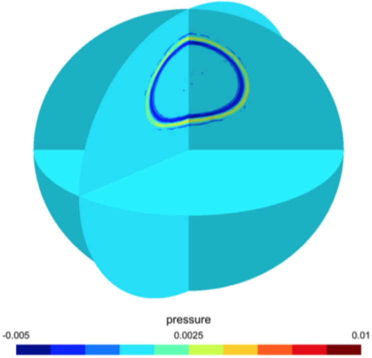
- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



J. Chan, RJH, T. Warburton (2016)

Example: Discontinuous Galerkin Solvers

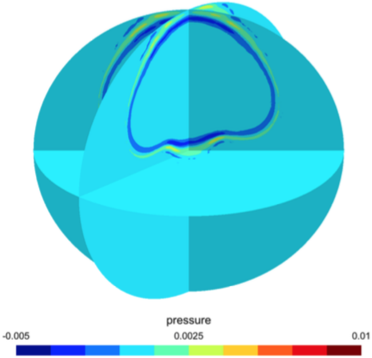
- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



J. Chan, RJH, T. Warburton (2016)

Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



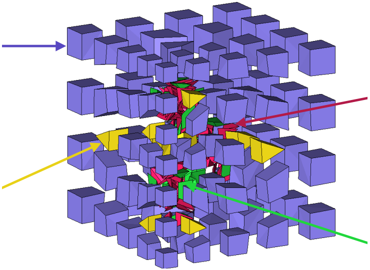
J. Chan, RJH, T. Warburton (2016)

Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)

Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)

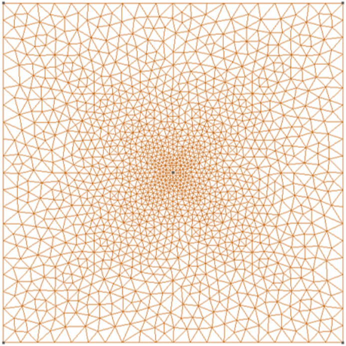


Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)

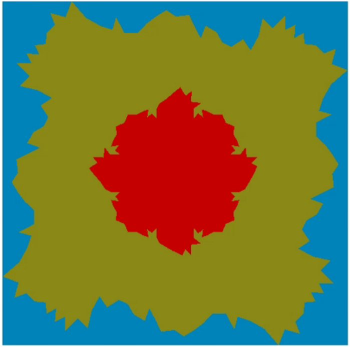
Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



Example: Discontinuous Galerkin Solvers

- Advantages
  - Can resolve complex topography
  - Natural mapping to accelerators
  - Many tuning parameters
- Challenges
  - Achieving peak performance (beating finite difference)
  - Application in inverse problems
  - Absorbing boundaries
  - Many tuning parameters
- Strategies
  - Hybrid meshing (& adjoints)
  - h- and p-refinement (& adjoints)
  - Local time stepping (& adjoints)
  - Optimal basis functions (& adjoints)



Current / Future Challenges

Optimization for large-scale PDE constrained IPs

- Uncertainty quantification without access to the Hessian
- What parameters can be inverted for?
  - Seismic data is not sensitive to density
  - Shear velocity only visible through S-P conversion
  - What about Thomson parameters?
- Joint inversion with CSEM, gravity, etc.

Current / Future Challenges

HPC & Software engineering

- HPC portable codes
- Math operations obscured deep in legacy codes – software architecture issues
- Domain experts are not expert computational scientists
- Efficient automatic adjoints and derivatives – numerical correctness vs HPC correctness
- Integration with rapidly developing exascale software
- Incredible cost of future algorithms

- CPU only is not viable
- “All” new machines are GPU machines

Current / Future Challenges

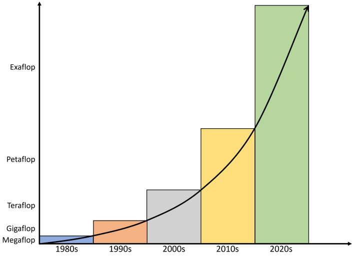
HPC & Software engineering

- HPC portable codes
- Math operations obscured deep in legacy codes – software architecture issues
- Domain experts are not expert computational scientists
- Efficient automatic adjoints and derivatives – numerical correctness vs HPC correctness
- Integration with rapidly developing exascale software
- Incredible cost of future algorithms

Current / Future Challenges

HPC & Software engineering

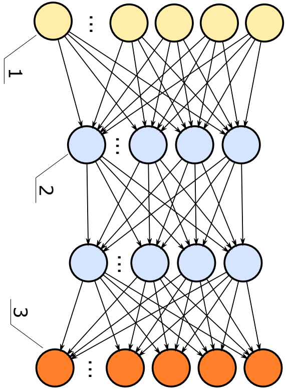
- HPC portable codes
- Math operations obscured deep in legacy codes – software architecture issues
- Domain experts are not expert computational scientists
- Efficient automatic adjoints and derivatives – numerical correctness vs HPC correctness
- Integration with rapidly developing exascale software
- Incredible cost of future algorithms



Current and Future Challenges

Machine Learning / Deep Learning

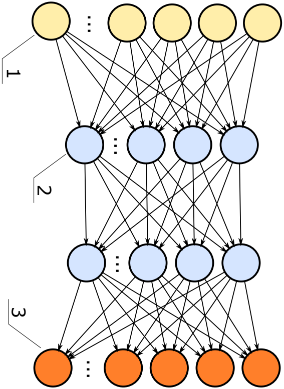
- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



Current and Future Challenges

Machine Learning / Deep Learning

- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



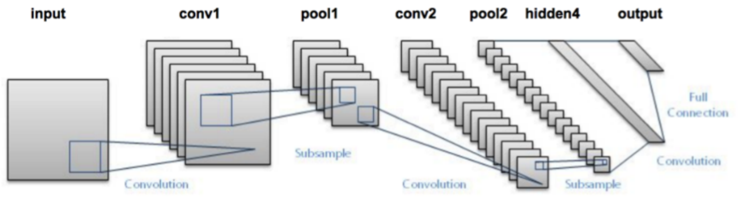
Adjoint States: Least-Squares

$$
\begin{aligned}
J(\boldsymbol{x}) &= \tfrac{1}{2}||\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}||_2^2 \\
\nabla J(\boldsymbol{x}_0) &= -\boldsymbol{A}^*(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x_0}) \\
&= -\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x_0})
\end{aligned}
$$

Adjoint States: Deep Learning

$$
\begin{aligned}
J(\boldsymbol{w},\boldsymbol{b}) &= \tfrac{1}{2}||d - \mathcal{F}(\boldsymbol{w},\boldsymbol{b}; I)||_2^2 \\
&= \tfrac{1}{2}||d - \mathcal{O}_6(\boldsymbol{w}_6,\boldsymbol{b}_6; \mathcal{C}_5(\ldots; \mathcal{P}_4(\ldots; \mathcal{C}_3(\ldots; \mathcal{P}_2(\boldsymbol{w}_2,\boldsymbol{b}_2; \mathcal{C}_1(\boldsymbol{w}_1,\boldsymbol{b}_1; I))))))||_2^2
\end{aligned}
$$

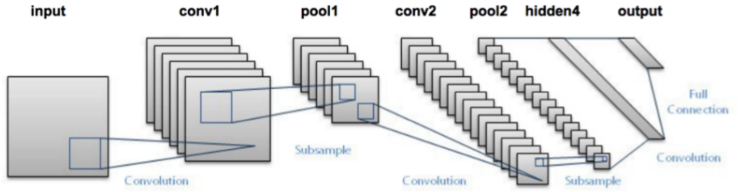
LeNet-5



Adjoint States: Deep Learning

$$\begin{aligned}
J(\boldsymbol{w}, \boldsymbol{b}) &= \tfrac{1}{2}||d - \mathcal{F}(\boldsymbol{w}, \boldsymbol{b}; I)||_2^2 \\
&= \tfrac{1}{2}||d - \mathcal{O}_6(\boldsymbol{w}_6, \boldsymbol{b}_6; \mathcal{C}_5(\ldots; \mathcal{P}_4(\ldots; \mathcal{C}_3(\ldots; \mathcal{P}_2(\boldsymbol{w}_2, \boldsymbol{b}_2; \mathcal{C}_1(\boldsymbol{w}_1, \boldsymbol{b}_1; I))))))||_2^2 \\
\nabla J(\boldsymbol{w}_0, \boldsymbol{b}_0) &= -F^*(d - \mathcal{F}(\boldsymbol{w}, \boldsymbol{b}; I)) \\
&= -\boldsymbol{C}_1^* \boldsymbol{P}_2^* \boldsymbol{C}_3^* \boldsymbol{P}_4^* \boldsymbol{C}_5^* \boldsymbol{O}_6^* (d - \mathcal{F}(\boldsymbol{w}, \boldsymbol{b}; I))
\end{aligned}$$

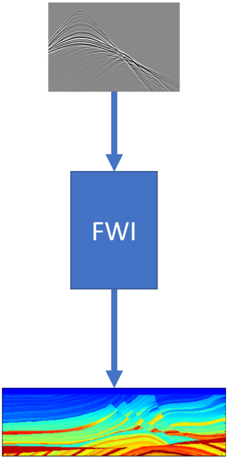
LeNet-5



Current and Future Challenges

Machine Learning / Deep Learning

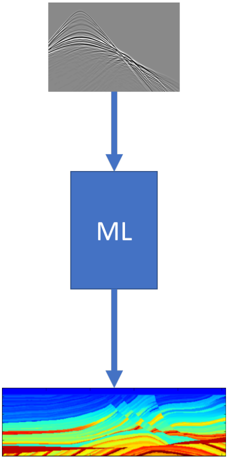
- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



Current and Future Challenges

Machine Learning / Deep Learning

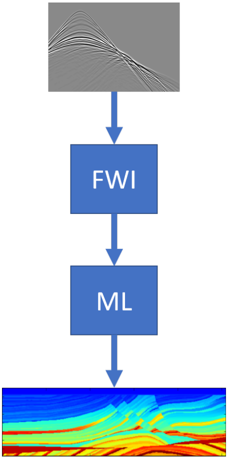
- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



Current and Future Challenges

Machine Learning / Deep Learning

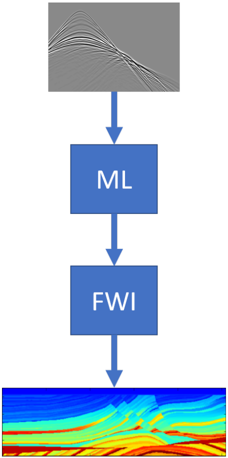
- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



Current and Future Challenges

Machine Learning / Deep Learning

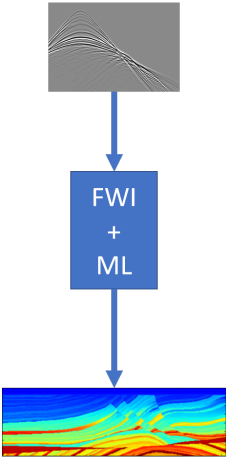
- Flavor of the ~~month~~ ~~year~~ decade?
- “Universal function approximation”
- “High-dimensional interpolation problem” (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



Current and Future Challenges

Machine Learning / Deep Learning

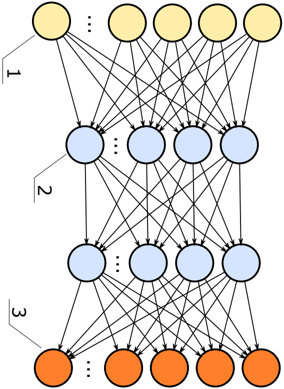
- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints
- Availability of training data?



Current and Future Challenges

Machine Learning / Deep Learning

- Flavor of the ~~month~~ ~~year~~ decade?
- "Universal function approximation"
- "High-dimensional interpolation problem" (Mallat)
- Training is inverse problem, backpropagation is adjoint state method!
- Autoencoders/generative models: path to UQ?
- Integrating physical (PDE) constraints



- Availability of training data?

Summary

Computational Science R&D Flow

physics → math → algorithm → HPC computation

In all of these aspects. . .

- Scale of exploration seismic inverse problems is astounding
- As academics we need to think bigger. . .
  - . . . and train students to think bigger!
- There are lots of interesting and challenging computational, mathematical, and numerical problems arising from seismic inverse problems

Summary

- Give FWI a shot:

  www.pysit.org

- Try the 1D FWI development exercise!