Computational and data-driven methods for large-scale inverse problems

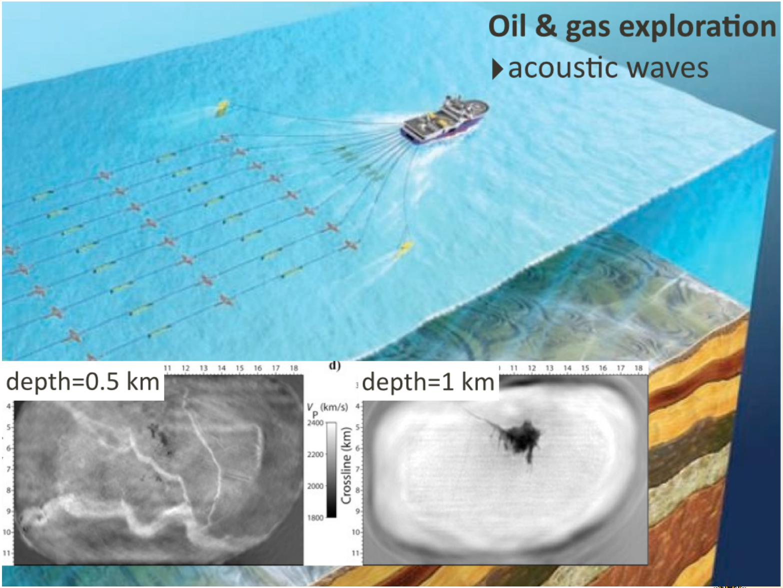
MAC-MIGS inaugural meeting

Tristan van Leeuwen

Universiteit Utrecht

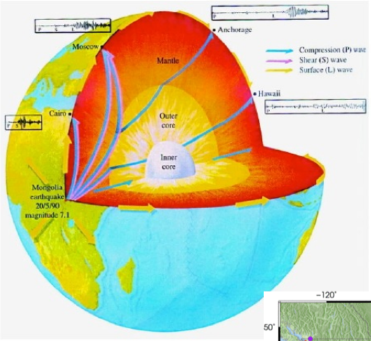
Oil & gas exploration

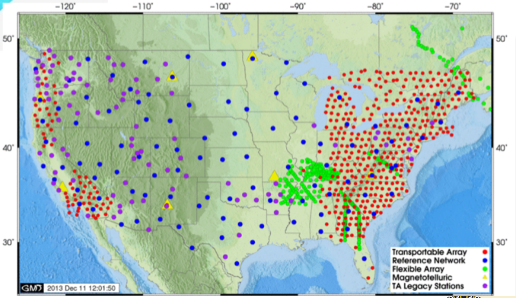
▸ acoustic waves



Seismology

- elastic waves





Universiteit Utrecht

Medical imaging

- ultrasound (sound)
- MRI (EM)
- Optical tomography (light)

Overview

- Inverse problems and PDE-constrained optimization

Selected topics:

- Geometric regularization

Overview

- Inverse problems and PDE-constrained optimization

Selected topics:

- Geometric regularization
- Constraint relaxation

Overview

- Inverse problems and PDE-constrained optimization

Selected topics:

- Geometric regularization
- Constraint relaxation
- Stochastic optimization

Overview

- Inverse problems and PDE-constrained optimization

Selected topics:

- Geometric regularization
- Constraint relaxation
- Stochastic optimization
- Uncertainty quantification

The inverse problem

Underlying physics is modeled by a PDE

$$L(c)u = q,$$

The inverse problem

Underlying physics is modeled by a PDE

$$L(c)u = q,$$

The measurements are a linear sampling of the state

$$d = Pu.$$

The inverse problem

Underlying physics is modeled by a PDE

$$L(c)u = q,$$

The measurements are a linear sampling of the state

$$d = Pu.$$

Introduce forward operator $F(c)q = PG(c)q$, where $G(c)q$ solves $L(c)u = q$.

The inverse problem

Given noisy data $d_i = F(c)q_i + \epsilon_i$ for $i = 1, \ldots, m$, find c.

The inverse problem

Given noisy data $d_i = F(c)q_i + \epsilon_i$ for $i = 1, \ldots, m$, find c.

Problem is typically ill-posed:

- existence
- uniqueness
- stability

The inverse problem

Bayesian approach leads to posterior of the form

$$\pi_{\mathrm{post}}(c|d) \propto \pi_{\mathrm{l}}(d|c)\pi_{\mathrm{prior}}(c).$$

The inverse problem

Bayesian approach leads to posterior of the form

$$\pi_{\text{post}}(c|d) \propto \pi_{\text{l}}(d|c)\pi_{\text{prior}}(c).$$

With $\epsilon_i \sim \mathcal{N}(0, \Sigma_d)$ and $c \sim \mathcal{N}(c_0, \Sigma_c)$ we have

$$\pi_{\text{post}}(c|d) \propto \prod_{i=1}^{m} \exp\left(-\tfrac{1}{2}\|F(c)q_i - d_i\|^2_{\Sigma_d^{-1}}\right) \exp\left(-\tfrac{1}{2}\|c - c_0\|^2_{\Sigma_c^{-1}}\right).$$

Universiteit Utrecht

The inverse problem

Bayesian approach leads to posterior of the form

$$\pi_{\text{post}}(c|d) \propto \pi_{\text{l}}(d|c)\pi_{\text{prior}}(c).$$

With $\epsilon_i \sim \mathcal{N}(0, \Sigma_d)$ and $c \sim \mathcal{N}(c_0, \Sigma_c)$ we have

$$\pi_{\text{post}}(c|d) \propto \prod_{i=1}^{m} \exp\left(-\frac{1}{2}\|F(c)q_i - d_i\|^2_{\Sigma_d^{-1}}\right) \exp\left(-\frac{1}{2}\|c - c_0\|^2_{\Sigma_c^{-1}}\right).$$

MAP estimate:

$$\widehat{c} = \max_c \pi(c|d).$$

Universiteit Utrecht

The inverse problem

Main tasks:

- Identify appropriate prior and likelyhood
- Solve large-scale non-linear optimization problem to get MAP estimate
- Quantify uncertainty

Geometric regularization

Popular choices for regularization include $\|(-\Delta)^\alpha c\|_p^p$:

- Tikhonov regularization ($p = 2$, smooth)
- Total Variation ($p = 1$, piecewise polynomial)

Geometric regularization

Popular choices for regularization include $\|(-\Delta)^\alpha c\|_p^p$:

- Tikhonov regularization ($p = 2$, smooth)
- Total Variation ($p = 1$, piecewise polynomial)

Stronger regularization can be achieved by *shape regularization*

$$c(x) = \begin{cases} c_1 & x \in \Omega \\ c_0 & x \notin \Omega \end{cases}.$$

Universiteit Utrecht

Geometric regularization

Popular choices for regularization include $\|(-\Delta)^{\alpha}c\|_p^p$:

- Tikhonov regularization ($p = 2$, smooth)
- Total Variation ($p = 1$, piecewise polynomial)

Stronger regularization can be achieved by *shape regularization*

$$c(x) = \begin{cases} c_1 & x \in \Omega \\ c_0 & x \notin \Omega \end{cases}.$$

- Represent the shape in terms of a level-set function $\Omega = \{x \,|\, \phi(x) > 0\}$.
- Express $c(x) = H(-\phi(x))c_0 + H(\phi(x))c_1$ and solve for ϕ.

The level-set method[1]

Seismic Full-waveform inversion

- Use smooth approximation of H and express ϕ in terms of RBFs
- Use quasi-Newton method to solve for coefficients
- Use alternating optimization when background is unknown

[1] Kadu, A., van Leeuwen, T., & Mulder, W. A. (2017). Salt Reconstruction in Full-Waveform Inversion With a Parametric Level-Set Method. IEEE Transactions on Computational Imaging, 3(2), 305–315.

A dual formulation[2]

For discrete *linear* inverse problems $Fc = d$, with $c \in \{-1, 1\}^n$ the primal problem is given by

$$\min_{\phi} \|FH(\phi) - d\|^2.$$

[2] Kadu, A., & Van Leeuwen, T. (2019). A convex formulation for binary tomography. IEEE Transactions on Computational Imaging, 1–1.

A dual formulation[2]

For discrete *linear* inverse problems $Fc = d$, with $c \in \{-1, 1\}^n$ the primal problem is given by

$$\min_{\phi} \|FH(\phi) - d\|^2.$$

The corresponding *dual* problem is given by

$$\min_{\mu} \|FF^{\dagger}(\mu - d)\|_2^2 + \|F^{*}\mu\|_1.$$

The result is obtained by $\widehat{c} = H(F^{*}\widehat{\mu})$.

[2] Kadu, A., & Van Leeuwen, T. (2019). A convex formulation for binary tomography. IEEE Transactions on Computational Imaging, 1–1.

A dual formulation

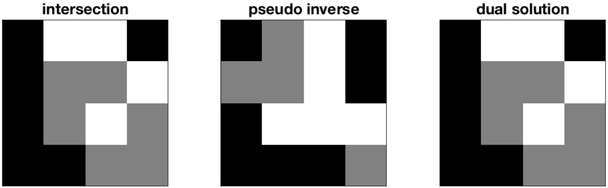
Discrete tomography

- Measurements consist of sums along rows, columns and diagonal.
- This particular problem instance has two solutions

A dual formulation

Discrete tomography

- Measurements consist of sums along rows, columns and diagonal.
- This particular problem instance has two solutions



PDE-constrained optimization

Cast as PDE-constrained optimization problem

$$\min_{c,u} \sum_{i=1}^{m} \|Pu_i - d_i\|^2_{\Sigma_d^{-1}} + \|c - c_0\|^2_{\Sigma_c^{-1}}, \quad \text{s.t.} \quad L(c)u_i = q_i.$$

PDE-constrained optimization

Cast as PDE-constrained optimization problem

$$\min_{c,u} \sum_{i=1}^{m} \|Pu_i - d_i\|^2_{\Sigma_d^{-1}} + \|c - c_0\|^2_{\Sigma_c^{-1}}, \quad \text{s.t.} \quad L(c)u_i = q_i.$$

- *All-at-once*: apply Newton's method to KKT system
- *Reduced*: eliminates constraints and solves non-linear least-squares problem

Adjoint-state method

$$\min_c \sum_{i=1}^{m} \|PG(c)q_i - d_i\|^2_{\Sigma_d^{-1}} + \|c - c_0\|^2_{\Sigma_c^{-1}}.$$

- Gradient and Hessian mat-vecs can be computed by solving forward and adjoint PDEs
- Dependence on c can be very non-linear, initialization is important

Universiteit Utrecht

Constraint relaxation[3]

Constraints may be too stringent

- bad initialization
- model-errors

[3]Leeuwen, T. van, & Herrmann, F. J. (2016). A penalty method for PDE-constrained optimization in inverse problems. Inverse Problems, 32(1), 015007.

Constraint relaxation[3]

Constraints may be too stringent

- bad initialization
- model-errors

Include constraint as a penalty

$$\min_{c,u} \sum_{i=1}^{m} \|Pu_i - d_i\|^2_{\Sigma_d^{-1}} + \|L(c)u_i - q_i\|^2_{\Sigma_m^{-1}} + \|c - c_0\|^2_{\Sigma_c^{-1}}.$$

[3] Leeuwen, T. van, & Herrmann, F. J. (2016). A penalty method for PDE-constrained optimization in inverse problems. Inverse Problems, 32(1), 015007.

Constraint relaxation[4]

A *reduced-space* approach involves solving a state-estimation problem

$$\min_u \sum_{i=1}^{m} \|Pu_i - d_i\|^2_{\Sigma_d^{-1}} + \|L(c)u_i - q_i\|^2_{\Sigma_m^{-1}}.$$

[4] Leeuwen, T. Van. (2019). A note on extended full waveform inversion, arXiv:1904.00363, 2019

Constraint relaxation[4]

A *reduced-space* approach involves solving a state-estimation problem

$$\min_{u} \sum_{i=1}^{m} \|Pu_i - d_i\|^2_{\Sigma_d^{-1}} + \|L(c)u_i - q_i\|^2_{\Sigma_m^{-1}}.$$

This results in

$$\min_{c} \sum_{i=1}^{m} \|PG(c)q_i - d_i\|^2_{\Sigma_d^{-1}+K(c)^{-1}},$$

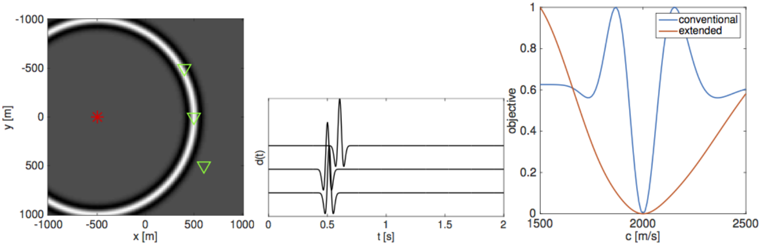
with $K = PG\Sigma_m G^* P^*$.

- requires solving a system with K
- if P is invertible, we get $\|PG(c)q_i - d_i\|^2_{\Sigma_d^{-1}+K(c)^{-1}} = \|L(c)P^{-1}d_i - q_i\|^2_{\Sigma_m^{-1}}$.

[4] Leeuwen, T. Van. (2019). A note on extended full waveform inversion, arXiv:1904.00363, 2019

Universiteit Utrecht

Constraint relaxation[5]



[5] Leeuwen, T. Van. (2019). A note on extended full waveform inversion, arXiv:1904.00363, 2019

Universiteit Utrecht

Optimization

The optimization problems we've seen so far are of the form

$$\min_c \underbrace{\sum_{i=1}^m f_i(c)}_{f(c)} + g(c).$$

Optimization

The optimization problems we've seen so far are of the form

$$\min_c \underbrace{\sum_{i=1}^m f_i(c)}_{f(c)} + g(c).$$

prototype algorithm:

$$c^{(k+1)} = \text{prox}_{\alpha g}\left(c^{(k)} - \alpha \nabla f(c^{(k)})\right).$$

Universiteit Utrecht

Optimization

The optimization problems we've seen so far are of the form

$$\min_c \underbrace{\sum_{i=1}^m f_i(c)}_{f(c)} + g(c).$$

prototype algorithm:

$$c^{(k+1)} = \text{prox}_{\alpha g}\left(c^{(k)} - \alpha \nabla f(c^{(k)})\right).$$

- evaluation of objective and gradient requires 2m PDE-solves
- guarantee convergence when using *approximate* gradients

Optimization

Convergence depends on error $e^{(k)} = \nabla f(c^{(k)}) - \nabla \widetilde{f}(c^{(k)})$.

[6] Aravkin, A., Friedlander, M. P., Herrmann, F. J., Leeuwen, T., & van Leeuwen, T. (2012). Robust inversion, dimensionality reduction, and randomized sampling. Mathematical Programming, 134(1), 101–125.

Optimization

Convergence depends on error $e^{(k)} = \nabla f(c^{(k)}) - \nabla \widetilde{f}(c^{(k)})$.

Stochastic:[6]

- error needs to be unbiased and variance needs to be small enough
- pick subset (batch) of experiments
- use linearity of PDE $\widetilde{f} = \|PG(c)\widetilde{q} - \widetilde{d}\|^2$ with $\widetilde{q} = \sum_i w_i q_i$ and $\mathbb{E}(ww^*) = I$.

[6] Aravkin, A., Friedlander, M. P., Herrmann, F. J., Leeuwen, T., & van Leeuwen, T. (2012). Robust inversion, dimensionality reduction, and randomized sampling. Mathematical Programming, 134(1), 101–125.

[7] van Leeuwen, T., & Herrmann, F. J. (2014). 3D Frequency-Domain Seismic Inversion with Controlled Sloppiness. SIAM Journal on Scientific Computing, 36(5), S192–S217.

Optimization

Convergence depends on error $e^{(k)} = \nabla f(c^{(k)}) - \nabla \widetilde{f}(c^{(k)})$.

Stochastic:[6]

- error needs to be unbiased and variance needs to be small enough
- pick subset (batch) of experiments
- use linearity of PDE $\widetilde{f} = \|PG(c)\widetilde{q} - \widetilde{d}\|^2$ with $\widetilde{q} = \sum_i w_i q_i$ and $\mathbb{E}(ww^*) = I$.

Deterministic:[7]

- $\|e^{(k)}\|$ needs to converge to zero at least as fast as the error $\|c^{(k)} - \widehat{c}\|$
- approximate f using approximate PDE-solves

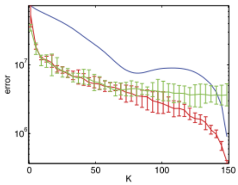
[6] Aravkin, A., Friedlander, M. P., Herrmann, F. J., Leeuwen, T., & van Leeuwen, T. (2012). Robust inversion, dimensionality reduction, and randomized sampling. Mathematical Programming, 134(1), 101–125.

[7] van Leeuwen, T., & Herrmann, F. J. (2014). 3D Frequency-Domain Seismic Inversion with Controlled Sloppiness. SIAM Journal on Scientific Computing, 36(5), S192–S217.

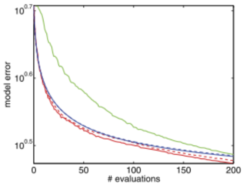
Optimization

Stochastic optimization with increasing batch size

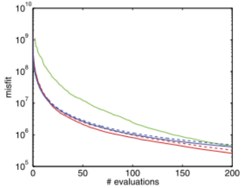
- subsets, natural order (blue)
- subsets, random order (red)
- mixing (green)



(a)



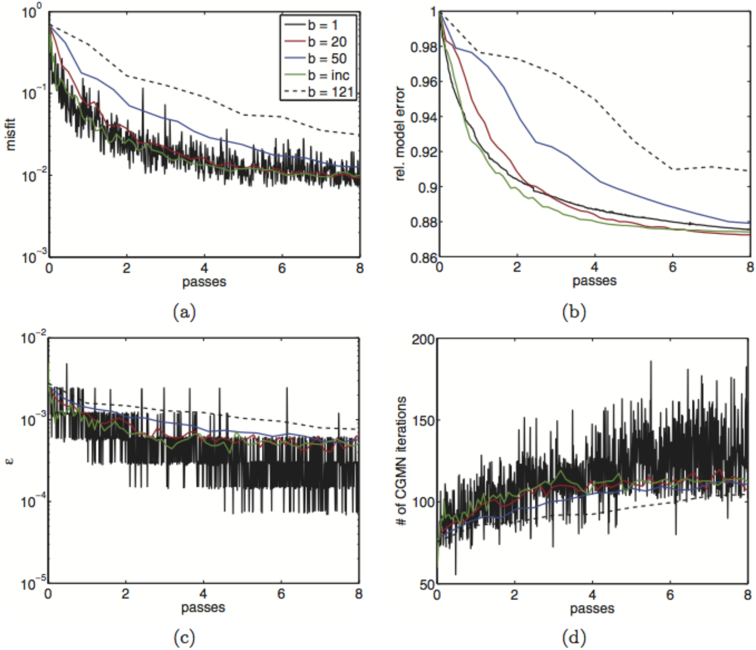
(b)



(c)

Universiteit Utrecht

Optimization



Universiteit Utrecht

Uncertainty quantification

Quantify uncertainty of $\widehat{c}$:

- use a local Gaussian approximation
- estimate *resolution* (psf) from Hessian

Universiteit Utrecht

Covariance estimation[8]

Estimating Σ_d^{-1} is crucial for UQ

$$\min_{c,\Sigma_d} \log|\Sigma_d| + \frac{1}{m}\sum_{i=1}^{m} \|PG(c)q_i - d_i\|^2_{\Sigma_d^{-1}}.$$

- closed-form expression available for $\widehat{\Sigma}_p$, can be efficiently approximated using randomized SVD
- covariance can be estimated as part of parameter-estimation:

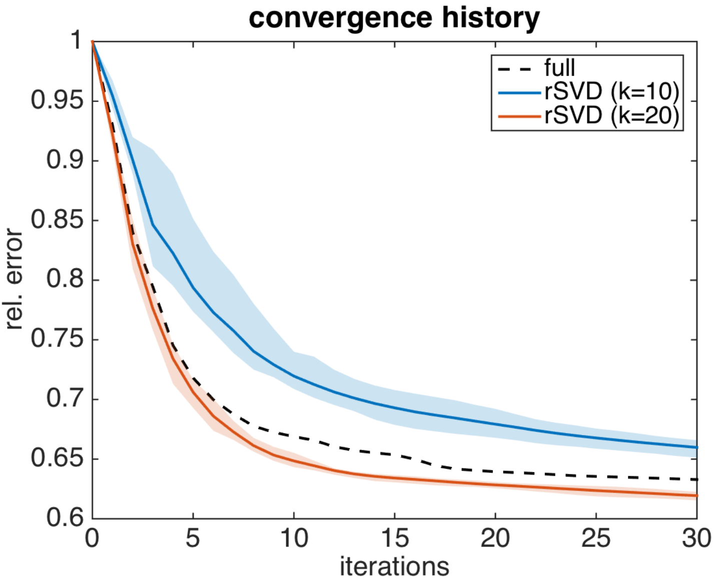
$$\widehat{\Sigma_d} = \sum_{i=1}^{m}(PG(c^{(k)})q_i - d_i)(PG(c^{(k)})q_i - d_i)^*$$

$$c^{(k+1)} = c^{(k)} - \alpha \sum_{i=1}^{m} J_i^* \widehat{\Sigma_d}^{-1}(PG(c)q_i - d_i).$$

[8] van Leeuwen, T. (2017). Joint parameter and state estimation for wave-based imaging and inversion. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6210–6214).

Universiteit Utrecht

Covariance estimation



Probing[9]

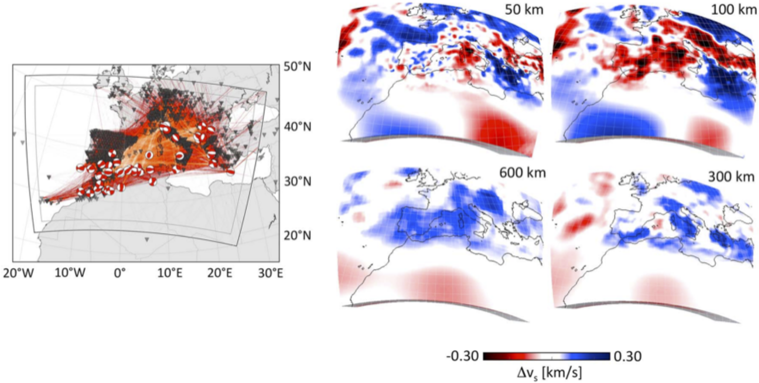
Assuming that the Hessian acts as a spatial convolution, $H = F^* \mathrm{diag}(\widehat{h}) F$, can we estimate the width of the kernel?

- Apply H to random vector $u = Hw$
- Compute auto-correlation $a = \mathrm{diag}\left(Fuu^* F^*\right)$
- In expectation we have $\mathbb{E}(a) = |\widehat{h}|^2$.

[9]Fichtner, A., & Leeuwen, T. van. (2015). Resolution analysis by random probing. Journal of Geophysical Research: Solid Earth, 120(8), 5549–5573.

Universiteit Utrecht

Probing[10]



[10] Fichtner, A., & Leeuwen, T. van. (2015). Resolution analysis by random probing. Journal of Geophysical Research: Solid Earth, 120(8), 5549–5573.

Wrap-up

Solving an inverse problem can be split in three main tasks

- modeling
- computation
- analysis and interpretation

Wrap-up

Solving an inverse problem can be split in three main tasks

- modeling
- computation
- analysis and interpretation

- Ideally, they form a feed-back loop and the data are used in all steps
- Interesting problems occur at the intersection of these steps

Universiteit Utrecht