Computational and data-driven methods for large-scale inverse problems
MAC-MIGS inaugural meeting

Tristan van Leeuwen
Oil & gas exploration

▷ acoustic waves

depth = 0.5 km

depth = 1 km
Seismology

elastic waves
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

\[ \pi_{\text{post}}(c|d) \propto \pi_l(d|c)\pi_{\text{prior}}(c). \]
The inverse problem

Bayesian approach leads to posterior of the form

\[ \pi_{\text{post}}(c|d) \propto \pi_l(d|c)\pi_{\text{prior}}(c). \]

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

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

Bayesian approach leads to posterior of the form

\[ \pi_{\text{post}}(c|d) \propto \pi_l(d|c)\pi_{\text{prior}}(c). \]

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

\[ \pi_{\text{post}}(c|d) \propto \prod_{i=1}^{m} \exp \left( -\frac{1}{2} \| F(c)q_i - d_i \|_2^2 \Sigma_d^{-1} \right) \exp \left( -\frac{1}{2} \| c - c_0 \|_2^2 \Sigma_c^{-1} \right). \]

MAP estimate:

\[ \hat{c} = \max_c \pi(c|d). \]
The inverse problem

Main tasks:

- Identify appropriate prior and likelihood
- 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 $\|(−Δ)^\alpha c\|_p^p$:

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

Stronger regularization can be achieved by *shape regularization*

$$c(x) = \begin{cases} 
  c_1 & x \in \Omega \\
  c_0 & x \notin \Omega 
\end{cases}.$$
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
  $\Omega = \{x \mid \phi(x) > 0\}$.
- Express $c(x) = H(-\phi(x))c_0 + H(\phi(x))c_1$ and solve for $\phi$. 
The level-set method

Seismic Full-waveform inversion

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

---

A dual formulation\textsuperscript{2}

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

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

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 $\hat{c} = H(F^*\hat{\mu})$.

---

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

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

Cast as PDE-constrained optimization problem

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

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

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

- 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
Constraint relaxation

Constraints may be too stringent

- bad initialization
- model-errors

---

Constraint relaxation

Constraints may be too stringent

- bad initialization
- model-errors

Include constraint as a penalty

\[
\min_{c,u} \sum_{i=1}^{m} \left\| Pu_i - d_i \right\|_2^2 \Sigma_d^{-1} + \left\| L(c) u_i - q_i \right\|_2^2 \Sigma_m^{-1} + \left\| c - c_0 \right\|_2^2 \Sigma_c^{-1}.
\]

---

Constraint relaxation

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

\[
\min_u \sum_{i=1}^{m} \|Pu_i - d_i\|_2^{2 \Sigma_d^{-1}} + \|L(c)u_i - q_i\|_2^{2 \Sigma_m^{-1}}.
\]

---

Constraint relaxation

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

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

This results in

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

with \( K = PG\Sigma_m G^* P^* \).

- requires solving a system with \( K \)
- if \( P \) is invertible, we get

\[
\| PG(c) q_i - d_i \|_{\Sigma_d}^2 \Sigma_d^{-1} + K(c)^{-1} = \| L(c) P^{-1} d_i - q_i \|_{\Sigma_m}^2.
\]

Constraint relaxation

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

$$\min_c \sum_{i=1}^{m} f_i(c) + g(c).$$
The optimization problems we’ve seen so far are of the form

\[
\min_c \sum_{i=1}^{m} f_i(c) + g(c).
\]

prototype algorithm:

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

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

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

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 \textit{approximate} gradients
Optimization

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

---

Optimization

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

**Stochastic:**

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

---


Optimization

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

Stochastic:\(^6\)

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

Deterministic:\(^7\)

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

---


Optimization

Stochastic optimization with increasing batch size

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

Quantify uncertainty of $\hat{c}$:

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

Estimating $\Sigma_d^{-1}$ 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 $\hat{\Sigma}_p$, can be efficiently approximated using randomized SVD
- covariance can be estimated as part of parameter-estimation:

$$\hat{\Sigma}_d = \sum_{i=1}^{m} (PG(c^{(k)})q_i - d_i)(PG(c^{(k)})q_i - d_i)^*$$

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

---

Covariance estimation

convergence history

- full
- rSVD (k=10)
- rSVD (k=20)

rel. error

iterations

0 5 10 15 20 25 30

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

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

---

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