Imaging with multi-source experiments

Eldad Haber

Outline

The problem

Standard solution techniques

Reformulation and some new insight

Numerical experiments

Extensions to moving sources/receivers

The problem

We consider an inverse problem solved by the optimization of

$$\begin{aligned} \min_m \quad & \frac{1}{2}\sum_j \|P^\top u_j - d_j\|^2 + R(m) \\ \text{s.t} \quad & A(m)u_j = Q_j \qquad j = 1, \ldots, N_s \end{aligned}$$

- $A(m)$ - a discretization of a parameter dependent differential operator
- Q_j - source
- P - observation matrix
- u_j - field
- d_j - data vector
- $R(m)$ - regularization

The problem

$$
\begin{aligned}
\min_{m} \quad & \frac{1}{2}\sum_{j} \|P^{\top}u_j - d_j\|^2 + R(m) \\
\text{s.t} \quad & A(m)u_j = Q_j \qquad j = 1, \ldots, N_s
\end{aligned}
$$

- The number of sources is LARGE
- The discretized PDE $A(m)$ is large and ill-conditioned
- Special structure - all sources share the same receivers

The problem

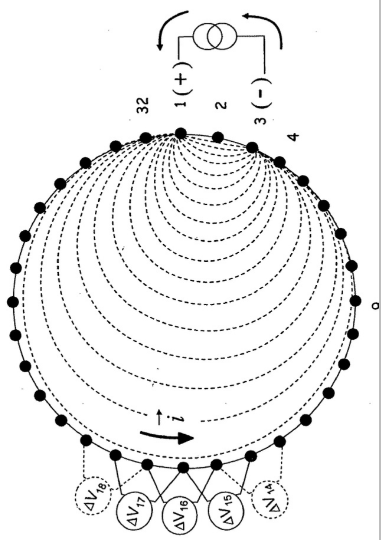
Medical examples include

- Electrical Impedance Tomography (EIT)
- Magnetic Induction Tomography (MIT)
- Microwave Imaging
- 3D Ultrasound

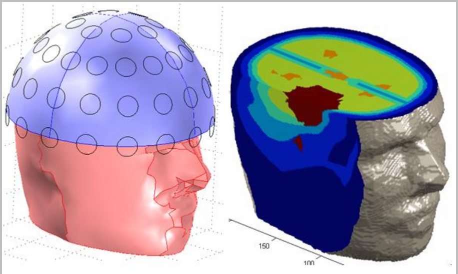
Geophysical examples include

- DC resistivity
- Electromagnetics
- Seismic imaging

Electrical Impedance Tomography



Magnetic Induction Tomography



Solution technique

$$\begin{aligned} \min_m \quad & \frac{1}{2}\sum_j \|P^\top u_j - d_j\|^2 + R(m) \\ \text{s.t} \quad & A(m)u_j = Q_j \qquad j = 1, \ldots, N_s \end{aligned}$$

Impossible to store all fields, use unconstrained approach [H. Oldenburg, Ascher 2000]

$$\min_m \quad \mathcal{J}(m) = \frac{1}{2}\sum_{j=1}^{N_s} \|P^\top A(m)^{-1} Q_j - d_j\|^2 + R(m)$$

Solution technique

$$\min_{m} \quad \mathcal{J}(m) = \frac{1}{2}\sum_{j=1}^{N_s} \|P^\top A(m)^{-1}Q_j - d_j\|^2 + R(m)$$

The gradient

$$\sum_{j=1}^{N_s} -G(m,u_j)^\top A(m)^{-\top} P(P^\top A(m)^{-1}Q_j - d_j) + \nabla R(m)$$

where $G(m,u_j) = \nabla_m (A(m)u_j)$

Solution technique

Computing the misfit and gradient

- Set misfit $= 0 \quad \nabla$misfit $= 0$
- For $j = 1, \ldots, N_s$
  - Solve $A(m)u_j = Q_j$
  - $r_j = P^\top u_j - d_j$
  - misfit $\leftarrow$ misfit $+ r_j^\top r_j$
  - Solve $A^\top \lambda = P r_j$
  - ∇misfit $\leftarrow \nabla$misfit $- G^\top \lambda$

Computation of misfit and its derivative require $2N_s$ solutions of the forward/adjoint problem.

For large scale problems difficult if not impossible

Solution technique

- Set misfit $= 0$ $\quad \nabla$misfit $= 0$
- For $j = 1, \ldots, N_s$
  - Solve $\overbrace{A(m)u_j = Q_j}^{\text{expensive!}}$
  - $r_j = P^\top u_j - d_j$
  - misfit $\leftarrow$ misfit $+ r_j^\top r_j$
  - Solve $\overbrace{A^\top \lambda = P r_j}^{\text{expensive}}$
  - ∇misfit $\leftarrow \nabla$misfit $- G^\top \lambda$

Solution technique

Current methods to deal with multiple rhs

- Factor the system if possible [Pratt, 2000, H. & Oldenburg, 2006]
- Almost factor the system (ILU, domain decomposition with large domains) [Ascher & van den Doel, 2009]
- Recycle right hand sides [Kilmer & de Sturler 2006]

Issues - complexity, storage

Solution technique

For the computation of a Gauss-Newton step similar calculations are needed.

Typically, avoid Gauss-Newton and use L-BFGS, nonlinear CG and steepest descent (storage).

Converges can be slow

A different point of view

The difficulty: computing the misfit.

Can we do this cheaper?

A different point of view

The difficulty: computing the misfit.
Can we do this cheaper?

Recall that

$$
\begin{aligned}
\text{misfit} &= \frac{1}{2}\sum_j \|P^\top A(m)^{-1}Q_j - d_j\|^2 \\
&= \frac{1}{2}\|P^\top A(m)^{-1}Q - D\|_F^2 = \\
&= \frac{1}{2}\mathrm{trace}\left((P^\top A(m)^{-1}Q - D)^\top (P^\top A(m)^{-1}Q - D)\right) \\
&= \frac{1}{2}\mathbf{E}_w\|(P^\top A(m)^{-1}Q - D)w\|^2
\end{aligned}
$$

where w is a random variable with

$$\mathbf{E}(w) = 0 \quad \mathsf{Cov}(w) = I$$

A different point of view

The original (deterministic) optimization problem is therefor equivalent to the (stochastic) optimization problem

$$\widehat{m} = \arg\min_m \quad \frac{1}{2}\sum_{j=1}^{N_s} \|P^\top A(m)^{-1}Q_j - d_j\|^2 + R(m)$$

$$= \arg\min_m \quad \frac{1}{2}\mathbf{E}_w\|P^\top A(m)^{-1}Qw - Dw\|^2 + R(m)$$

A different point of view

The original (deterministic) optimization problem is therefor equivalent to the (stochastic) optimization problem

$$\begin{aligned}\widehat{m} = \arg\min_m \quad & \frac{1}{2}\sum_{j=1}^{N_s} \|P^\top A(m)^{-1}Q_j - d_j\|^2 + R(m) \\ = \arg\min_m \quad & \frac{1}{2}\mathbf{E}_w\|P^\top A(m)^{-1}Qw - Dw\|^2 + R(m)\end{aligned}$$

So What?

A different point of view

$$\widehat{m} = \arg\min_{m} \quad \frac{1}{2}\mathbf{E}_w \|P^\top A(m)^{-1} Qw - Dw\|^2 + R(m)$$

- This is a stochastic optimization problem [Shapiro 09] and has been treated extensively in the literature
- We can capitalize on the structure of the problem to obtain cheap algorithms

Main point - Given a realization w_i a **Single** PDE solve is required to evaluate misfit$(m; w_i)$

A different point of view

$$\widehat{m} = \arg\min_{m} \quad \frac{1}{2}\mathbf{E}_w \|P^\top A(m)^{-1}Qw - Dw\|^2 + R(m)$$

A different point of view

$$\widehat{m} = \arg\min_{m} \quad \frac{1}{2}\mathbf{E}_w \|P^\top A(m)^{-1} Qw - Dw\|^2 + R(m)$$

Two methods for stochastic optimization [Shapiro 05]

SAA - Sample Average Approximation
Discretize the Expectation THEN optimize

SA - Stochastic Approximation
Optimize AND Discretize

Stochastic Optimization

SAA - Sample Average Approximation
Approximate the expectation using Monte-Carlo

$$\min_{m} \sum_{j} \frac{1}{2}\|P^\top A(m)^{-1}Qw_j - Dw_j\|^2 + R(m)$$

Stochastic Optimization

SAA - Sample Average Approximation
Approximate the expectation using Monte-Carlo

$$\min_m \sum_j \frac{1}{2}\|P^\top A(m)^{-1}Qw_j - Dw_j\|^2 + R(m)$$

SA - Stochastic approximation
for $j = 1, \ldots$

$$\begin{aligned} \hat{s} &= \mathrm{arg(aprox)min} \frac{1}{2}\|P^\top A(m_j+s)^{-1}Qw_j - Dw_j\|^2 + R(s) \\ & s_{j+1} = \mathrm{average}(s_{1:j}, \hat{s}) \end{aligned}$$

end

SAA - Sample Average Approximation

Approximate the expectation using Monte-Carlo

$$\min_m \quad \mathcal{J}(m; w) = \sum_j \frac{1}{2} \|P^\top A(m)^{-1} Q w_j - D w_j\|^2 + R(m)$$

- How to pick w?
- How many w's?

Any distribution with $\mathbf{E}(w) = 0$ and $\mathsf{Cov}(w) = I$ has

$$\mathbf{E}\left(w^\top H w\right) = \mathsf{trace}(H)$$

Choose the distribution such that [Hutchinson 93]

$$\mathsf{Var}\left(w^\top H w\right) \rightarrow \min$$

$$w = \mathsf{rand}(\pm 1)$$

SAA - Sample Average Approximation

Approximate the expectation using Monte-Carlo

$$\min_m \quad \mathcal{J}(m;w) = \sum_j \frac{1}{2}\|P^\top A(m)^{-1}Qw_j - Dw_j\|^2 + R(m)$$

- How to pick w?
- How many w's?

The number of w's depends on the variability of the unbiased estimator

$$\mathbf{E}(\mathcal{J}(m;w) \approx \frac{1}{N}\sum_{j=1}^{N}\mathcal{J}(m;w_j)$$

and the accuracy we would like to obtain.

SAA - Example

Generate $A(m)$ by discretizing the PDE

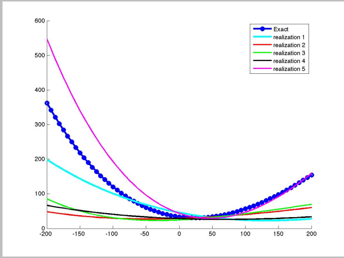
$$\nabla \cdot \exp(m) \nabla u$$

Assume 1089 sources (right hand sides) and 1089 receivers
Look at

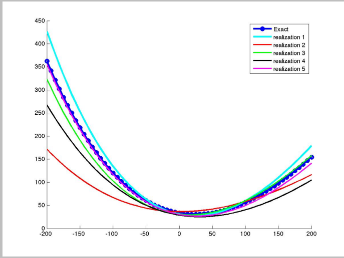
$$f(\alpha) = \frac{1}{2N} \sum_j \|P^\top A(m+\alpha s)^{-1} Q - D\|_F^2$$

SAA - Sample Average Approximation

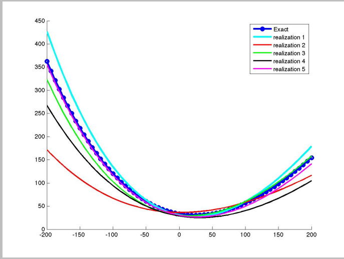
1 vector



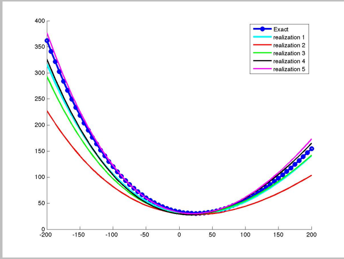
10 vectors



25 vectors



50 vectors



SAA - Sample Average Approximation

Controlling the quality of the approximation can be done by repeating the minimization with different samples

For our problems we have found that a small sample size may be sufficient [Also Golub & Bai, 99, Golub & von Matt 98]

Advantage of SAA - separate the stochastic part from the optimization

Disadvantage of SAA - number of realization may be too large

SA - Stochastic Approximation

Stochastic approximation
for $j = 1, \ldots$

$$\begin{aligned} \hat{s} &= \arg(\text{aprox})\min \frac{1}{2}\|P^\top A(m_j + s)^{-1} Q w_j - D w_j\|^2 + R(s) \\ & s_{j+1} = \text{average}(s_{1:j}, \hat{s}) \end{aligned}$$

end

Questions

- How approximate?
- What methods can be used?
- Convergence?

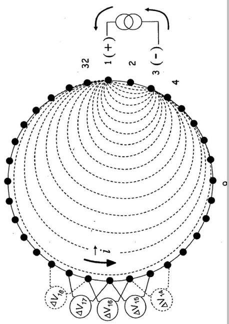
SA - Stochastic Approximation

- Proof that it works only on various flavors of steepest descent (recent work [Nemirovski and Shaoiro])
- Observed in practice - works well for L-BFGS and Gauss-Newton [Schraudolph, Yu & Gunter 10]
- Much interest in machine learning (online algorithms)

An illustrative example

Model problem - Electrical Impedance Tomography

$$\min_{m} \quad \|P^\top (G^\top S(m) G)^{-1} Q - D\|_F^2 + \frac{\alpha}{2}\|Gm\|^2$$



An illustrative example

- Assume 1089 sources and 1089 receivers.
- Mesh $32 \times 32 \times 32$
- Number of unknowns (fields) $35,684,352$
- Computation of full forward, roughly, 2 hours

Use standard, SAA and SA to solve the problem.

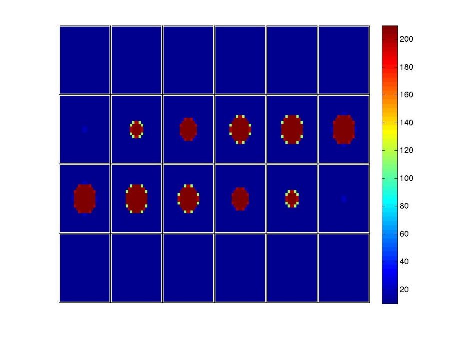
An illustrative example

Method	# iterations	# rhs/iter	Cost (pde solves)
Standard	37	1089	80,293
SSA	45	10	2757
SA	453	1	923

Computational saving of factor 100

$$\|m_{\mathsf{SAA}} - m_{\mathsf{Standard}}\|^2/\|m_{\mathsf{Standard}}\|^2 = 2.1 \times 10^{-2}$$
$$\|m_{\mathsf{SA}} - m_{\mathsf{Standard}}\|^2/\|m_{\mathsf{Standard}}\|^2 = 3.2 \times 10^{-2}$$

Inversion parameters

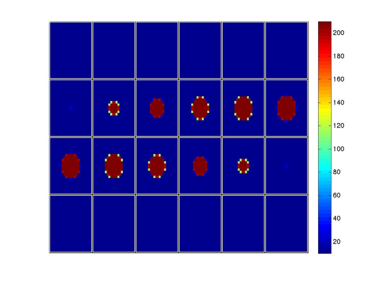


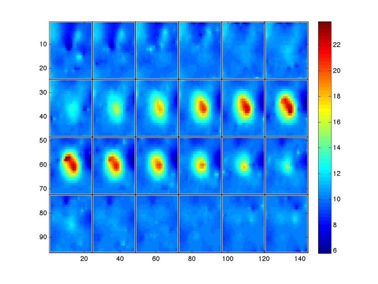
$\alpha = 10^{-4}$

Starting model - $m = 10^{-2}$S/m

Converges - solution does not change between iterations

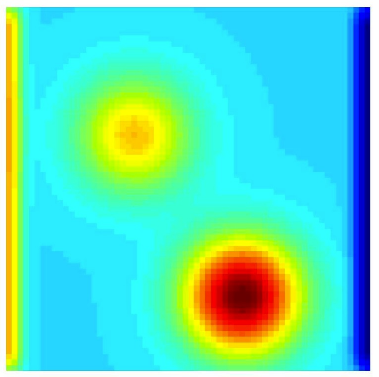
Recovered solution



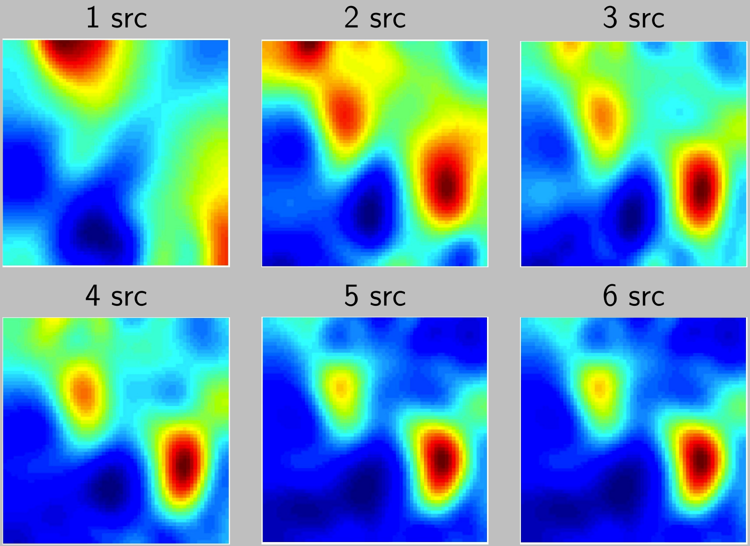


Sequential SAA

How to choose the size of the random batch?
Use continuation in batch-size
Example - EIT in 2D



Sequential SAA



Conclusions

- Develop a new point of view for multi-source data
- Can solve the problem in a fraction of the cost of the original problem
- Key - stochastic trace estimators and stochastic optimization
- Applications in other parameter estimation problems with many sources