High-dimensional EM imaging in geoscience and engineering accelerated by parallel black-box Fast Multipole Method

Jonghyun Harry Lee and Niels Grobbe

Collaboration with Ruoxi Wang, Chao Chen and Eric F. Darve

NSF's IUCRC CEMC-UHM IAB Meeting

11/8/2018

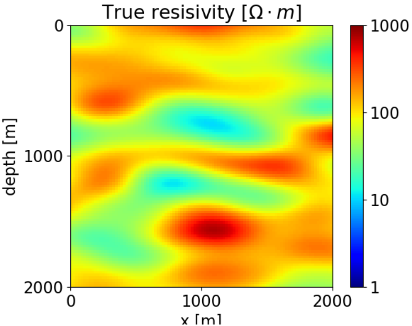
US Army Corps of Engineers

Cloud Credits for Research

Motivation

Researchers would like to know the structure of deep subsurface for optimal water resources management and scientific research.

Magnetotelluric (MT) survey is the one of a few options to image deep subsurface structure.

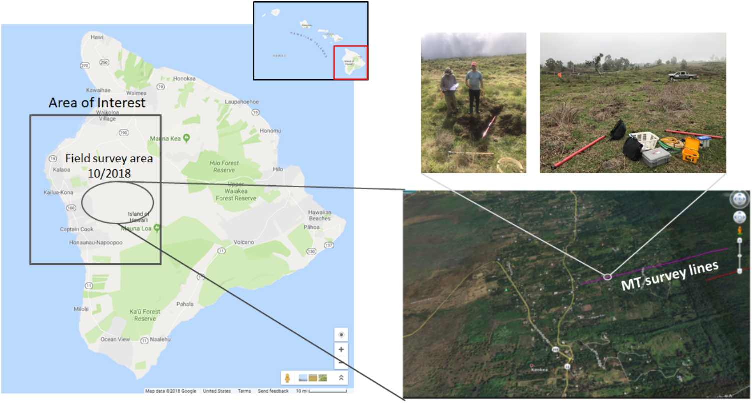


Synthetic "true" resisitivity field (reciprocal of electrical conductivity)

Motivation

Recently, the UHM geophysics and subsurface modeling teams visited a field site in Big Island and we would like to perform as many surveys as possible while we stay in the site.

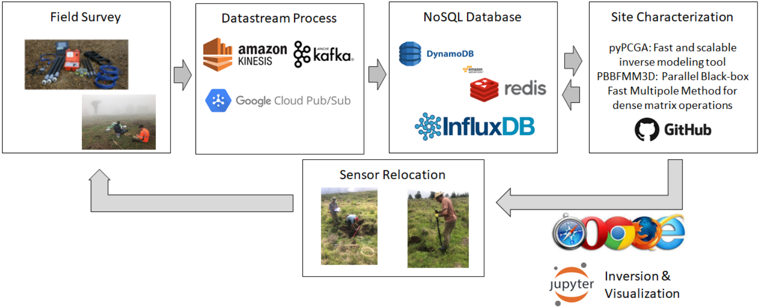
Can we make a real-time decision of which location we need to install MT instruments for better subsurface characterization?



Motivation

Q: Can we check site characterization results right away for a real-time decision of next MT survey locations?

A: **Yes! (during lunch time at latest)**



Real-time close-loop optimal MT survey design

EM in Geophysics

EM-based Geophysical method examples:

1. Magnetotelluric (MT): a passive geophysical investigation for deep subsurface regions
2. Ground-penetrating Radar (GPR): high resolution imaging for shallow regions, e.g., UXO detection
3. Marine Controlled-source ElectroMagnetic (CSEM): for offshore oil and gas exploration

- EM field generated by active or passive sources sensitive to subsurface conductivity variations in the region between the source and receivers
- For MT, the instrument measure naturally occuring, very low frequency EM waves (telluric currents) that penetrate into the earth
- Analysis of the variations in the electrical voltage and electromagnetic wave energy will enable us to determine the electrical resistivity of the rocks and to identify groundwater flow occurring at varying depths in the subsurface.

MT Modeling

$$
\begin{aligned}
\nabla \times \mathbf{E} - i\omega\mu\mathbf{H} &= \mathbf{M}_s \\
\nabla \times \mathbf{H} - \sigma &= \mathbf{J}_s
\end{aligned}
$$

$\mathbf{E}$:= frequency-domain electric field
$\mathbf{H}$:= magnetic field
$\mathbf{M}_s$:= electric sources
$\mathbf{J}_s$:= magnetic sources

- quasi static approximation for low frequency geophysics applications (neglecting imaginary component - dielectric permittivity)
- MARE2DEM [Key, 2016] [1]: FEM solver with adaptive meshing refinement in terms of EM response accuracy at the receivers
- Support MPI for strong scaling with domain/data decomposition in different frequencies/receivers

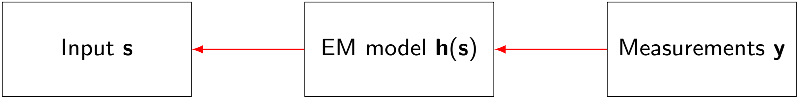
[1] `http://mare2dem.ucsd.edu/`

Inverse Problem

Input $\mathbf{s}$ $\longrightarrow$ EM model $\mathbf{h}(\mathbf{s})$ $\longrightarrow$ Output $\mathbf{y}$

- In the **forward problem**, given model parameters, $\mathbf{s}$, EM model (e.g., MARE2DEM) predicts the state of the system $\mathbf{y}$
- $\mathbf{s}$ is typically conductivity/resistivity, but could be other rock properties such as permeability, or boundary conditions
- $\mathbf{y}$ are EM responses/measurements

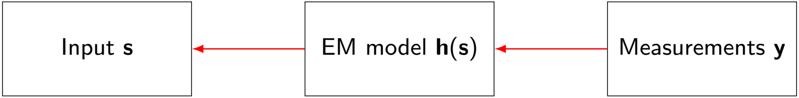
Inverse Problem



- In the **forward problem**, given model parameters, **s**, EM model (e.g., MARE2DEM) predicts the state of the system **y**
- **s** is typically conductivity/resistivity, but could be other rock properties such as permeability, or boundary conditions
- **y** are EM responses/measurements

- In the **inverse** problem, we use measurements of **y** to estimate **s**

Inverse Problem



- In the **forward problem**, given model parameters, **s**, EM model (e.g., MARE2DEM) predicts the state of the system **y**
- **s** is typically conductivity/resistivity, but could be other rock properties such as permeability, or boundary conditions
- **y** are EM responses/measurements

- In the **inverse** problem, we use measurements of **y** to estimate **s**
- MARE2DEM is now used to calculate the sensitivity of measurements to parameters, i.e., Jacobian

Inverse Problem in Hierarchical Bayesian Framework

Consider the measurement equation

$$y = h(s) + v \qquad v \sim \mathcal{N}(0, \mathbf{R})$$

$\mathbf{y} := n_{obs} \times 1$ noisy measurements	pressure, temperature
$h :=$ EM forward model	e.g., MARE2DEM
$v :=$ measurement and model error	uncertainty and error
$\mathbf{s} := n_{unknowns} \times 1$	permeability

$$s \sim \mathcal{N}(s_{prior}, \mathbf{Q}_{prior})$$

- Parameters are treated as random variables in a statistical framework (e.g., Gelman, Calin, and Stern, 2013; Kitanidis, 2010, Kitanidis, 1995)
- Use covariances **Q** and **R** to represent variability and uncertainty
- Objective: **A best estimate of unknowns and corresponding uncertainty at each grid cell of the numerical model, given a set of measurements**

Inverse Problem in Hierarchical Bayesian Framework

Consider the measurement equation

$$y = h(s) + v \qquad v \sim \mathcal{N}(0, \mathbf{R})$$

Using Bayes' rule, the posterior pdf is

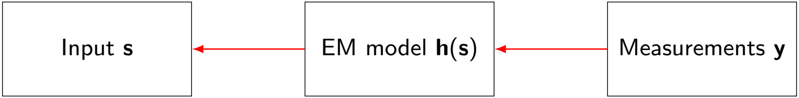
$$p(s|y) \quad \propto \quad \underbrace{p(y|s)}_{\text{Data misfit}} \; \underbrace{p(s)}_{\text{Prior}}$$

- Data misfit - How well the model reproduces data
- Prior - Prior knowledge of unknown field structure

Best estimate is obtained by maximizing the likelihood of **s** given a set of measurements **y**, using GN optimization:

$$p(\mathbf{s}) \sim \exp\left(\underbrace{-\frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{s}))^\top \mathbf{R}^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{s}))}_{likelihood} \underbrace{-\frac{1}{2}(\mathbf{s} - \mathbf{s}_{prior})^\top \mathbf{Q}_{\text{prior}}^{-1} (\mathbf{s} - \mathbf{s}_{prior})}_{prior} \right)$$

Inverse Problem: the challenges for large systems



For large-scale systems:

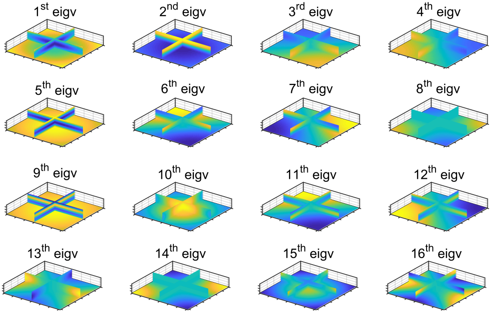
- Typically many unknowns, few measurements $n_{obs} \ll n_{unknowns}$
- Requires $\mathcal{O}(min(n_{obs}, n_{unknowns}))$ EM model runs or more
- $\mathcal{O}(n^2)$ or $\mathcal{O}(n^3)$ matrix computation and storage costs

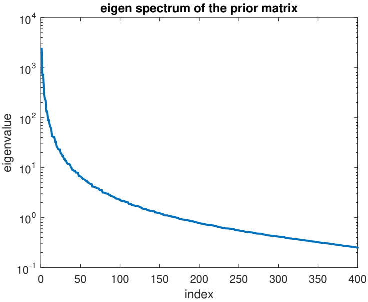
Therefore:

- Fast Linear Algebra is necessary to reduce computation and storage
  - Matrix-matrix, matrix-vector multiplications
- Allowable number of forward model runs

pyPCGA: Fast and Scalable Inversion Tool in Python[2]

Principal **C**omponent **G**eostatistical **A**pproach:
A computationally efficient algorithm for geostatistical (spatially distributed unknown) inversion based on the dimension reduction of Hierarchical Bayesian inverse solution through the optimal compression of prior covariance and Jacobian-free evaluation of sensitivity.[1]





[1] Lee and Kitanidis, 2014, Lee, 2016
[2] https://github.com/jonghyunharrylee/pyPCGA

pyPCGA: Fast and Scalable Inversion Tool in Python

Compression of the covariance matrix reduces the number of matrix-vector multiplications to $\mathcal{O}(n_{pc})$:

$$\mathbf{Q}_{prior} = \mathbf{U}_{1:n_{pc}} \mathbf{\Sigma}_{1:n_{pc}} \mathbf{U}_{1:n_{pc}}^{T}$$

Calculation of sensitivity matrix requires EM model runs, black-box style, using the finite difference approach:

$$\mathbf{Hs} = \frac{h\left(\mathbf{s} + \Delta s\right) - h\left(\mathbf{s}\right)}{\Delta s}$$

Computations involving large matrices (**Q**, **H**) utlilize fast linear algebra that allows **fully parallelizable, fast matrix-vector multiplications**:

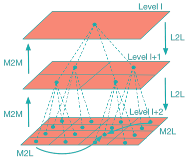
- Fast Fourier Transform (FFT) approach for regular grids[1]
- Fast Multipole Method (FMM) and Hierachical Matrices Approach for unstructured grids[2]

[1] `https://github.com/arvindks/kle`
[2] `https://github.com/ruoxi-wang/PBBFMM3D`

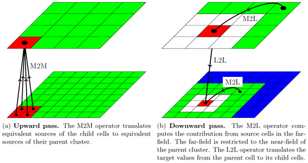
PBBFMM3D: Parallel Black-Box Fast Multipole Method for $\mathcal{O}(n)$ matrix operations

Typical covariance matrix-Jacobian column vector products requires $\mathcal{O}(n^2)$ and eigen-decomposition of covariance matrix needs $\mathcal{O}(n^3)$ computational costs.

To achieve linear scalability, we develop and use Parallel Black-Box Fast Multipole Method (PBBFMM3D)[1] leading to $\mathcal{O}(n_{unknowns} n_{pc})$ computation costs for truncated eigendecomposition (using randomized SVD)



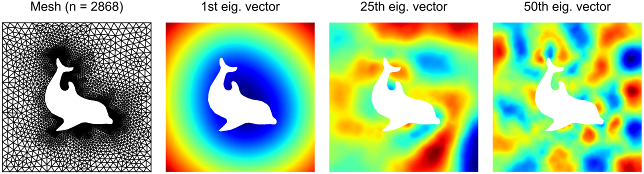
Hierarchical decomposition[1]



local/multipole operations[2]

[1] Wang, Chen, Lee, Darve, https://github.com/ruoxi-wang/PBBFMM3D
[2] Agullo et al., 2014

PBBFMM3D: Parallel Black-Box Fast Multipole Method for $\mathcal{O}(n)$ matrix operations

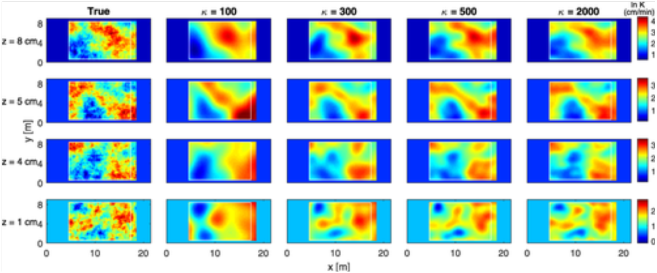


3 minutes for 1 M x 1 M covariance matrix eigenvalue decomposition with 100 modes on a 36 CPU core machine

pyPCGA: Advantages

Computational gain:

- Matrix computations scale linearly with number of unknowns
- $\sim \mathcal{O}(100)$ forward model runs for large domains (for $\gg 10^6$ unknowns)
- Parallelization further accelerates inversion



- **Linear scaling makes possible the inversion of domains with millions of unknowns and observations**[1].

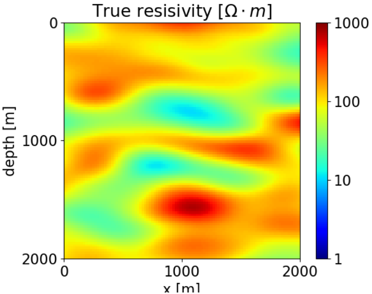
[1] Lee et al., 2016

pyPCGA: Advantages

Method	Adjoint-based method	PCGA
# of simulation runs	$n_{obs}+1$	$n_{pc}+1$
matrix multiplication	$\mathcal{O}(n_{obs} n_{unknowns})$	$\mathcal{O}(n_{unknowns} n_{pc})$
storage	$\mathcal{O}(n_{obs} n_{unknowns})$	$\mathcal{O}(n_{obs} n_{pc})$

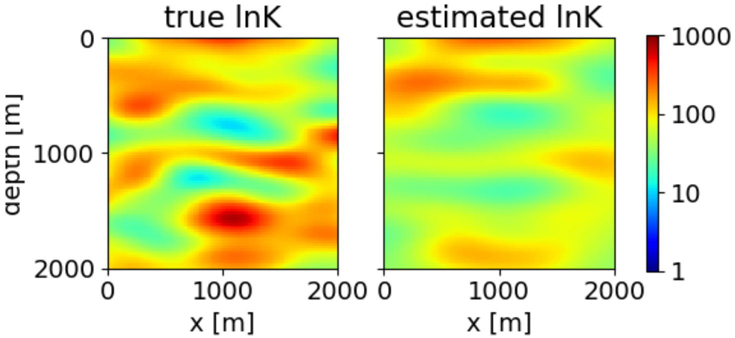
- $n_{pc}+1$ simulation runs in each iteration
- $n_{pc} \sim O(100)$ or less for many problems in earth science
- Can handle large measurements (e.g, 10^7 measurements)
- Easy to implement; treat multi-physics models as a “blackbox” like Ensemble-based methods
- Parallel executions

Numerical Example - Real-time EM imaging

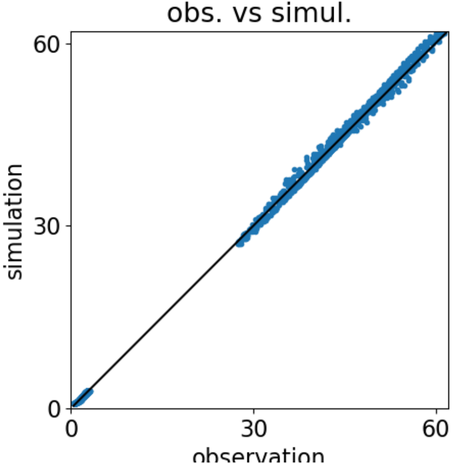


- $n_{unknowns} = 10,000$, $n_{obs} = 5,248$ (21 receivers x 32 frequency bands, EM amplitude and phase)
- measurement error assumed to be Gaussian with 0.1
- $n_{pc} = 50$
- Took 2 iterations in 10 mins with $\sim$ 100 model runs on a 36 CPU core workstation.

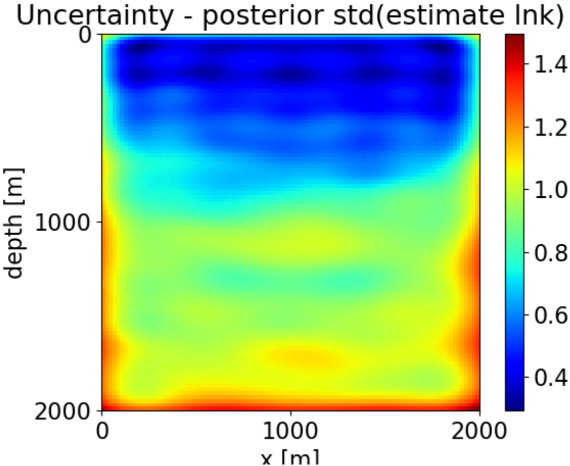
Results - the best estimate



Results - Data Fitting



Results - Estimation Uncertainty



Concluding Remarks

- Dimension reduction technique for prior covariance allows fast inverse modeling and data assimilation without much loss of accuracy.
- Work in-progress : sharp boundary object reconstruction using total-variation (compressed sensing) and level set priors

References

- `https://github.com/jonghyunharrylee/pyPCGA`
- Code example in this presentation `https://github.com/jonghyunharrylee/pyPCGA/tree/master/examples/mare2dem_MT`
- Lee and Kitanidis, Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (PCGA), Water Resources Research, 50(7), 2014
- Lee, Yoon, Kitanidis, Werth, and Valocchi, Scalable subsurface inverse modeling of huge data sets with an application to tracer concentration breakthrough data from magnetic resonance imaging, Water Resources Research, 52(7), 5213-5231, 2016
- Saibaba, Lee, and Kitanidis, Randomized algorithms for generalized Hermitian eigenvalue problems with application to computing Karhunen-Loeve expansion, Numerical Linear Algebra with Applications, 23(2), 314-339, 2016