Released to public domain under Creative Commons license type BY (https://creativecommons.org/licenses/by/4.0). Copyright (c) 2018 SINBAD consortium - SLIM group @ The University of British Columbia.

Efficient approach for quantifying uncertainty of wavefield reconstruction inversion

Zhilong Fang^{*}, Chia Ying Lee[†], Curt Da Silva^{*}, Tristan van Leeuwen^{**} and Felix J. Herrmann^{*} *Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia †Department of Mathematics, University of British Columbia **Mathematical Institute, Utrecht University

2016/10/26

SLM University of British Columbia



Motivation



Noisy acquired data



Motivation









Possible velocity model

Uncertainty in data





Risk in oil and gas volume



Motivation



Inversion result









Bayesian inference Prior probability density function (PDF): $\mathbf{m} \longrightarrow \rho_{\mathrm{prior}}(\mathbf{m})$ Likelihood PDF: given data d $\mathbf{m} \longrightarrow \rho_{\text{like}}(\mathbf{d}|\mathbf{m})$ Posterior PDF (Bayes' rule): $\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) = \rho_{\text{like}}(\mathbf{d}|\mathbf{m})\rho_{\text{prior}}(\mathbf{m})$ [A. Tarantola and B. Valette, 1982] [J. Kaipio and E. Somersalo, 2004]





Bayesian inference

Mean value of the model: $\mathbb{E}(\mathbf{m}) = \int \mathbf{m}\rho_{\text{post}}(\mathbf{m})d\mathbf{m},$ **Covariance matrix:**

 $C_{i,j} = \mathbb{E}(m_i m_j) - \mathbb{E}(m_i)\mathbb{E}(m_j),$



Bayes w/ FWI



7





Bayes w/ FWI

Reduced formulation in the frequency domain:

- $F(\mathbf{m}) = \mathbf{P}\mathbf{A}^{-1}\mathbf{q},$
 - $\mathbf{A} = \Delta + \omega^2 \mathbf{m}$
 - $\mathbf{m}: Squared-slowness$
 - **q** : Source
 - ω : Frequency
 - $\Delta:$ Laplacian operator
 - $\mathbf{P}:$ Projection operator of receiver

$+\omega^2 \mathbf{m}$



[J. Virieux and S. Operto, 2009]

Bayes w/ FWI Posterior PDF of FWI: $\rho_{\rm post}(\mathbf{m}|\mathbf{d}) \propto \exp\left(-\frac{1}{2}\|\mathbf{PA}(\mathbf{m})\|\mathbf{n}\|^2\right)$

9





[J. Virieux and S. Operto, 2009]

Bayes w/ FWI Posterior PDF of FWI: $\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto \exp\left(-\frac{1}{2}||\mathbf{PA}(\mathbf{m})||\mathbf{m}|\mathbf{d}|\right)$

Strong nonlinearity Many local minima

Wednesday, October 26, 2016

$$\mathbf{n})^{-1}\mathbf{q} - \mathbf{d}\|_{\boldsymbol{\Sigma}_{\text{noise}}^{-1}}^2 - \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{\text{prior}}\|_{\boldsymbol{\Sigma}_{\text{prior}}^{-1}}^2\right)$$



[J. Virieux and S. Operto, 2009]

Bayes w/ FWI Posterior PDF of FWI: $\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto \exp\left(-\frac{1}{2}\|\mathbf{PA}(\mathbf{m})\|\mathbf{m}\|^2\right)$

Strong nonlinearity $-\log \rho_{\text{post}}(\mathbf{m}|\mathbf{d})$ **Many local minima**

0

$$\mathbf{m}^{-1}\mathbf{q} - \mathbf{d}\|_{\boldsymbol{\Sigma}_{noise}^{-1}}^{2} - \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{prior}\|_{\boldsymbol{\Sigma}_{prior}^{-1}}^{2}\right)$$



Two layer example – FWI



10



FWI $-\log
ho_{\mathrm{post}}(\mathbf{m}|\mathbf{d})$



[James Martin et al, 2012]

Bayes w/ FWI

Local minima:

- difficult to find the maximum a posterior (MAP) estimate
- slow down McMC convergence
- render posterior PDF "less Gaussian"

n a posterior (MAP) estimate nce



[T. van Leeuwen and F. J. Herrmann, 2013]

[T. van Leeuwen and F. J. Herrmann , 2015]

Wavefield Reconstruction Inversion – WRI

Penalty formulation:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2} \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 -$$

Properties:

- bi-linear w/ respect tou and m
- larger searching space







Larger # of degrees of freedom



"more convex" less local minima





[T. van Leeuwen and F. J. Herrmann , 2013]

[T. van Leeuwen and F. J. Herrmann, 2015]

Solving WRI Variable projection: $\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{P}\overline{\mathbf{u}}(\mathbf{m}) - \mathbf{d}\|_{2}^{2} + \frac{\lambda^{2}}{2} \|\mathbf{A}(\mathbf{m})\overline{\mathbf{u}}(\mathbf{m}) - \mathbf{q}\|_{2}^{2}$

where

$$\overline{\mathbf{u}}(\mathbf{m}) = \arg\min_{\mathbf{u}} \frac{1}{2} \|\mathbf{P}\mathbf{u} - \mathbf{u}\|_{\mathbf{u}} = \frac{1}{2} \|\mathbf{P}\mathbf{u}\|_{\mathbf{u}} + \frac{1}{2} \|\mathbf{P}\mathbf{u}\|_{\mathbf{u}} \frac{1}{2} \|\mathbf{P}\mathbf{u}\|_{\mathbf{u$$

$-\mathbf{d}\|_{2}^{2} + rac{\lambda^{2}}{2} \|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{q}\|_{2}^{2}$



WRI – iterations

WRI method

for each source isolve $\begin{pmatrix} \mathbf{P}_i \\ \lambda \mathbf{A}_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{\lambda,i} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$ $\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \operatorname{diag}(\bar{\mathbf{u}}_{i,\lambda})^* (A(\mathbf{m}) \bar{\mathbf{u}}_{i,\lambda} - \mathbf{q}_i)$ end $\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$ correlation proxy wavefield & PDE residual

Conventional method





[Z. Fang *et al,* 2015]

Bayes w/ WRI Posterior PDF of WRI: $ho_{
m post}({f m}|{f d}) \propto$ $\exp\left(-\frac{1}{2}\|\mathbf{P}\overline{\mathbf{u}}(\mathbf{m})-\mathbf{d}\|_{\boldsymbol{\Sigma}_{noise}}^{2}-\right.$ Like where $\overline{\mathbf{u}}(\mathbf{m}) = \arg\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{P}\mathbf{u} - \mathbf{u}\|_{\mathbf{m}}$

16

$$-\frac{\lambda^2}{2} \|\mathbf{A}(\mathbf{m})\overline{\mathbf{u}}(\mathbf{m}) - \mathbf{q}\|^2 - \frac{1}{2} \|\mathbf{m} - \mathbf{m}_p\|_{\boldsymbol{\Sigma}_{\mathrm{pr}}^{-1}}^2$$
elihood Prior

$$\| \mathbf{d} \|_{\mathbf{\Sigma}_{\text{noise}}^{-1}}^2 + rac{\lambda^2}{2} \| \mathbf{A}(\mathbf{m}) \mathbf{u} - \mathbf{q} \|^2$$



Two layer example – WRI



17



WRI – $\log \rho_{\rm post}(\mathbf{m}|\mathbf{d})$



Two layer example – FWI



18



FWI $-\log
ho_{\mathrm{post}}(\mathbf{m}|\mathbf{d})$



Challenges for large-scale UQ

Evaluations of the posterior PDF

- many PDE solves to evaluate PDF
- expensive PDE solves

High-dimensional space to explore

- numerical integration too expensive
- McMC based methods are impractical
 - too many iterations
 - converge too slow



UQ for large-scale problems

Approximate posterior PDF by Gaussians

Sample the Gaussians w/ Randomize Then Optimize (RTO) method







m

Approximate posterior PDF

Gaussian approximation:

where



$\rho_{\text{post}}(\mathbf{m}) \approx \rho_{\text{Gauss}}(\mathbf{m}) = \mathcal{N}(\mathbf{m}_*, \mathbf{H}^{-1})$



Approximate posterior PDF





[T. van Leeuwen and F. J. Herrmann, 2015]

Form Hessian

Gauss-Newton Hessian:

$\mathbf{H}_{\mathrm{l}} = \mathbf{G}^{\top} \mathbf{A}^{-\top} \mathbf{P}^{\top} (\mathbf{I} + \frac{1}{\lambda^2 \sigma^2} \mathbf{P} \mathbf{A}^{-1} \mathbf{A}^{-\top} \mathbf{P}^{\top})^{-1} \mathbf{P} \mathbf{A}^{-1} \mathbf{G},$

where $\Sigma_{\text{noise}} = \sigma^2 \mathbf{I}$ and $\mathbf{G} = \frac{\partial \mathbf{A} \overline{\mathbf{u}}}{\partial \mathbf{m}}$.



[T. van Leeuwen and F. J. Herrmann, 2015]

Form Hessian

Gauss-Newton Hessian:





GN Hessian of WRI





 \blacksquare

S

Computational cost:

 $n_{\rm freq} \times (n_{\rm src} + n_{\rm rcv})$

storage cost:

 $n_{\rm freq} \times n_{\rm grid} \times (n_{\rm src} + n_{\rm rcv})$ in parallel !!



[T. van Leeuwen and F. J. Herrmann , 2015]

Selection of λ

Compute wavefields:

$$\overline{\mathbf{u}}(\mathbf{m}) = \left(\frac{1}{\sigma^2} \mathbf{P}^\top \mathbf{P} + \lambda^2 \mathbf{A}\right)$$

$$\lambda^2 >> \mu_1 \rightarrow \overline{\mathbf{u}}(\mathbf{n})$$
$$\lambda^2 << \mu_1 \rightarrow \frac{1}{\sigma^2} \mathbf{I}$$

Select
$$\lambda^2 = lpha \mu_1$$

$(\mathbf{m})^{ op} \mathbf{A}(\mathbf{m}))^{-1} (\lambda^2 \mathbf{A}(\mathbf{m})^{ op} \mathbf{q} + rac{1}{\sigma^2} \mathbf{P}^{ op} \mathbf{d})$ μ_1 : maximum eigenvalue of the matrix $\frac{1}{\sigma^2} \mathbf{A}^{-\top} \mathbf{P}^{\top} \mathbf{P} \mathbf{A}^{-1}$, \mathbf{m}) $\approx \mathbf{A}(\mathbf{m})^{-1}\mathbf{q}$ $\mathbf{P}^{\top}\mathbf{P} + \lambda^2 \mathbf{A}(\mathbf{m})^{\top}\mathbf{A}(\mathbf{m})$ is ill conditioned



UQ for large-scale problems

Approximate posterior PDF by Gaussians

Sample the Gaussians w/ Randomize Then Optimize (RTO) method



Conventional method

Sample Gaussian distribution:

Cholesky factorization:

 $\mathbf{m}_{s} = \mathbf{m}_{*} + \mathbf{L}^{-1} \mathbf{r}, \mathbf{r} \sim \mathcal{N}(0, \mathcal{I}_{n_{\text{grid}} \times n_{\text{grid}}})$

H should be an explicit matrix, computational cost is $\mathcal{O}(n_{\text{grid}}^3)$.

$ho_{\mathrm{Gauss}}(\mathbf{m}) \propto \exp(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_*)^\top \mathbf{H}(\mathbf{m} - \mathbf{m}_*))$

$\mathbf{H} = \mathbf{L}^{\top} \mathbf{L},$



RTO method

Re-formulate the posterior distribution with: $\mathbf{H} = \mathbf{H}_{l} + \mathbf{H}_{p}, \ \mathbf{H}_{l} = \mathbf{L}_{l}^{\top} \mathbf{L}_{l}, \ \text{and} \ \mathbf{H}_{p} = \mathbf{L}_{p}^{\top} \mathbf{L}_{p},$

then

 $ho_{\mathrm{Gauss}}(\mathbf{m}) \propto \exp(-\frac{1}{2}(\mathbf{L}_{\mathrm{l}}\mathbf{m} - \mathbf{L}_{\mathrm{l}}\mathbf{m}_{*})^{\top}(\mathbf{L}_{\mathrm{l}}\mathbf{m} - \mathbf{L}_{\mathrm{l}}\mathbf{m}_{*})$

$-rac{1}{2}(\mathbf{L}_{\mathrm{p}}\mathbf{m}-\mathbf{L}_{\mathrm{p}}\mathbf{m}_{*})^{ op}(\mathbf{L}_{\mathrm{p}}\mathbf{m}-\mathbf{L}_{\mathrm{p}}\mathbf{m}_{*}))$



[J. M. Bardsley *et al*, 2014]

RTO method

Generate a sample by solving the optimization problem:

$\mathbf{m}_{s} = rgmin \|\mathbf{L}_{l}\mathbf{m}\|$ m

where

 $\mathbf{r}_{\mathrm{l}} \sim \mathcal{N}(0, \mathcal{I}_{n_{\mathrm{rev}} \times n_{\mathrm{rev}}})$ a

$$egin{aligned} & \|\mathbf{L}_{\mathrm{l}}\mathbf{m} - (\mathbf{L}_{\mathrm{l}}\mathbf{m}_{*} + \mathbf{r}_{\mathrm{l}})\|^{2} + \ & \|\mathbf{L}_{\mathrm{p}}\mathbf{m} - (\mathbf{L}_{\mathrm{p}}\mathbf{m}_{*} + \mathbf{r}_{\mathrm{p}})\|^{2}, \end{aligned}$$

and
$$\mathbf{r}_{\mathrm{p}} \sim \mathcal{N}(0, \mathcal{I}_{n_{\mathrm{grid}} \times n_{\mathrm{grid}}}).$$



RTO method

Factorization of H_1 :

$\mathbf{H}_{l} = \mathbf{L}_{l}^{\top} \mathbf{L}_{l},$ $\mathbf{L}_{\mathrm{l}} = (\mathbf{I} + \frac{1}{\lambda^2 \sigma^2} \mathbf{P} \mathbf{A}^{-1} \mathbf{A}^{-\top} \mathbf{P}^{\top})^{-\frac{1}{2}} \mathbf{P} \mathbf{A}^{-1} \mathbf{G}.$

 $n_{\rm rcv} \times n_{\rm rcv}$







Covariance matrix

RTO method (100000 realizations) vs analytical solution:







Marginal distribution comparison

RTO method with 100000 realizations (red) vs Analytical solution (blue)





Numerical Experiment – layer model

Depth of sources and receivers: 50 m Number of sources and receivers: 61 Frequency: 2,3 and 4 Hz Lambda: 4e4 sigma: 10







Randomized Maximum Likelihood - RML

Generate independent samples from $\rho_{\text{post}}(\mathbf{m})$ by solving: $\min_{\mathbf{m}} \frac{1}{2} (\sigma^{-2} \| \mathbf{P} \overline{\mathbf{u}}(\mathbf{m}) - \mathbf{d} - \mathbf{r}_{d} \|^{2} + \lambda^{2} \| \mathbf{A}(\mathbf{m}) \overline{\mathbf{u}}(\mathbf{m}) - \mathbf{q} - \mathbf{r}_{s} \|^{2})$ $+\frac{1}{2}\|\mathbf{m}-\mathbf{m}_{\mathrm{p}}-\mathbf{r}_{\mathrm{p}}\|_{\Sigma_{\mathrm{prior}}^{-1}}^{2},$ where

$$\begin{split} \mathbf{r}_{\mathrm{d}} &\sim \mathcal{N}(0, \sigma^{2} \mathcal{I}_{n_{\mathrm{rev}} \times n_{\mathrm{rev}}}), \\ \mathbf{r}_{\mathrm{s}} &\sim \mathcal{N}(0, \lambda^{-2} \mathcal{I}_{n_{\mathrm{grid}} \times n_{\mathrm{grid}}}), \end{split}$$

$$\mathbf{r}_{\mathrm{p}} \sim \mathcal{N}(0)$$

 $, \Sigma_{\mathrm{prior}}).$



STD result comparison





Cross section comparison





BG Compass model



(a) True model



Wednesday, October 26, 2016

Depth of sources and receivers: 50 Number of sources and receivers: 91 / 451 Central frequency: 15 Hz Frequency: 2-31 Hz Lambda: 1e4



Data $\sigma = 34$







Prior and initial model





MAP estimate



Lateral [m]



Posterior STD









Cross section comparison – prior

Cross section comparison – posterior vs prior

Cross section comparison - 95% confidence interval vs 10 realizations by RML

Conclusions

Penalty formulation of posterior PDF

- is a bi-Gaussian PDF
- has a better Gaussian approximation compared to reduced formulation

Efficient sampling method

- with evaluating posterior PDF iteratively
- matrix-vector product
- expensive Cholesky factorization

Gaussian approximation avoids large computational cost associated

• the implicit GN Hessian operator provides a fast way to compute

• RTO method does not require an explicit Hessian matrix and

Future work

Application to 3D problems.

Bayesian with constraint prior information.

Effects of different acquisition scenarios to the UQ analysis.

Reference

- 1. Johnathan M Bardsley, Antti Solonen, Heikki Haario, and Marko Laine. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. SIAM Journal on Scientific Computing, 36(4):A1895–A1910, 2014.
- 2.Yan Chen and Dean S Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.
- data using multiple-point geostatistical a priori information. *Geophysics*, 77(2):H19–H31, 2012.
- 4.Zhilong Fang, C Lee, C Silva, F Herrmann, and Rachel Kuske. Uncertainty quantification for wavefield reconstruction inversion. In 77th EAGE Conference and Exhibition 2015, 2015.
- 5. James Martin, Lucas C. Wilcox, Carsten Burstedde, and Omar Ghattas. A Stochastic Newton MCMC Method for Largescale Statistical Inverse Problems with Application to Seismic Inversion. SIAM Journal on Scientific Computing, 34(3):A1460–A1487, 2012.
- space. *Geophysical Journal International*, 195:661–667, 10 2013b.
- 7. Tristan van Leeuwen and Felix J. Herrmann. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):015007, 12 2015.
- 8.Håvard Rue. Fast sampling of gaussian markov random fields. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2):325–338, 2001.
- 9.Albert Tarantola and Bernard Valette. Inverse problems = quest for information. Journal of Geophysics, 50: 159–170, 1982

3.Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard. Monte carlo full-waveform inversion of crosshole gpr

6.Tristan van Leeuwen and Felix J. Herrmann. Mitigating local minima in full-waveform inversion by expanding the search

Acknowledgements

support of the member organizations of the SINBAD Consortium.

This research was carried out as part of the SINBAD project with the

Acknowledgements

The authors wish to acknowledge the SENAI CIMATEC Supercomputing Center for Industrial Innovation, with support from BG Brasil, Shell, and the Brazilian Authority for Oil, Gas and Biofuels (ANP), for the provision and operation of computational facilities and the commitment to invest in Research & Development.

