Efficient Hessian computation in deterministic and Bayesian inverse problems

Daniel I. Gendin, Paul E. Barbone

Boston University
Acknowledgment: NIH R01 CA195527

March 26, 2021

Introduction

- Forward problem

$$\boxed{\text{Parameter(m)}} \rightarrow \boxed{\text{Model}} \rightarrow \boxed{\text{State(u)}}$$

- Inverse problem

$$\boxed{\text{Parameter(m)}} \leftarrow \boxed{\text{Model}} \leftarrow \boxed{\text{Measurement}(\tilde{u})}$$

- Bayes's Theorem

$$P_{post}(m|\tilde{u}) \propto P_{like}(\tilde{u}|m)P_{prior}(m) \tag{1}$$

- Posterior probability is easy to evaluate but difficult to interpret.
- How do we characterize the posterior?

Background

- Sampling full posterior (e.g. MCMC)
  - [Petra et al., 2014, Bardsley et al., 2020, Chen and Ghattas, 2020, Vigliotti et al., 2018, Zou et al., 2019]
- Laplace approximation: $\mathrm{P}_{\mathrm{post}}(\mathrm{m}) \approx N(\bar{\mathrm{m}}, H^{-1}[\bar{\mathrm{m}}])$
  - [Bui-Thanh et al., 2013, Saibaba et al., 2020, Chang et al., 2014, Fatehiboroujeni et al., 2020, Cui et al., 2016]
- Approximating the Hessian
  - [Saibaba et al., 2020, Ambartsumyan et al., 2020, Flath et al., 2011]
- In this work we find the MAP through Newton-CG and approximate the Hessian by using the Krylov basis found in computing the MAP.
  - This gives us the Hessian *"for free."*
  - Finding the MAP is a constrained optimization problem.

Optimization formulation

- Cost = -Log posterior

$$\underbrace{C(\mathrm{m})}_{-\log(\mathrm{P_{post}}(\mathrm{m}|\tilde{\mathrm{u}}))} = \underbrace{\frac{1}{2}||\tilde{\mathrm{u}} - \mathrm{u}(\mathrm{m})||^2_{noise}}_{-\log(\mathrm{P_{like}}(\tilde{\mathrm{u}}|\mathrm{m}))} + \underbrace{\frac{1}{2}R(\mathrm{m},\mathrm{m})}_{-\log(\mathrm{P_{prior}}(\mathrm{m}))}\ . \tag{2}$$

- Constraint equation (weak form).

$$a(\hat{\mathrm{w}}, \mathrm{u}; \mathrm{m}) = l(\hat{\mathrm{w}}) \qquad \forall \hat{\mathrm{w}} \in \mathcal{W}. \tag{3}$$

- Laplace approximation
  - Close to the MAP point

$$C(\mathrm{m}) = C(\bar{\mathrm{m}}) + \cancel{(G[\bar{\mathrm{m}}], \mathrm{m} - \bar{\mathrm{m}})_m} + \frac{1}{2}H[\bar{\mathrm{m}}](\mathrm{m}-\bar{\mathrm{m}}, \mathrm{m}-\bar{\mathrm{m}}) + O(||\mathrm{m}-\bar{\mathrm{m}}||^3). \tag{4}$$

$$P_{\mathrm{post}}(\mathrm{m}) \underset{\sim}{\propto} \exp(\frac{1}{2}H[\bar{\mathrm{m}}](\mathrm{m} - \bar{\mathrm{m}}, \mathrm{m} - \bar{\mathrm{m}})) \sim N(\bar{\mathrm{m}}, H^{-1}[\bar{\mathrm{m}}]) \tag{5}$$

MAP evaluation

- We use a Newton-CG method to find the MAP point.
- Newton (outer) iterations
  - Tend to converge in few iterations
  - Consistent with Laplace approximation.
  - Explicit construction of full Hessian is prohibitive
- Preconditioned-CG (inner) iterations
  - Requires only the action of the Hessian in the search directions.
  - Constructs a Krylov space of H-conjugate search directions $\{p\}$ and R-orthogonal gradients $\{r\}$.
  - Algorithm theoretically converges in $K_d + 1$ steps, where K_d is the rank of the data part of the Hessian.

Efficient Hessian evaluation

- Given preconditioned-CG products p_a, r_a and q_a and s_a:

Main result

$$H[\mathrm{m}^n](\delta\mathrm{m}_a, \delta\mathrm{m}_b) = \sum_{j=1}^{k} \frac{1}{D_{jj}} (\delta\mathrm{m}_a, \mathrm{q}_j)(\mathrm{q}_j, \delta\mathrm{m}_b) - \sum_{j=0}^{k-1} \frac{1}{C_{jj}} (\delta\mathrm{m}_a, \mathrm{s}_j)(\mathrm{s}_j, \delta\mathrm{m}_b) + R(\delta\mathrm{m}_a, \delta\mathrm{m}_b) \tag{6}$$

- Where:

$$(\mathrm{v}, \mathrm{q}_a) = H[\mathrm{m}^n](\mathrm{v}, \mathrm{p}_a) \qquad \forall \mathrm{v} \in \mathcal{M} \tag{7}$$

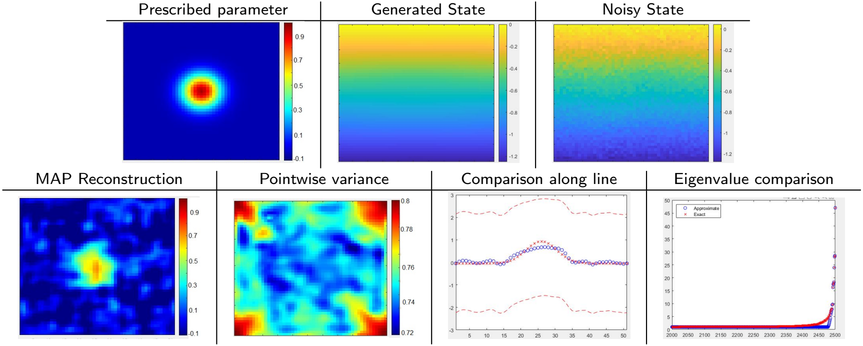
$$(\mathrm{v}, \mathrm{s}_a) = R(\mathrm{v}, \mathrm{r}_a) \qquad \forall \mathrm{v} \in \mathcal{M} \tag{8}$$

$$D_{aa} = H[\mathrm{m}^n](\mathrm{p}_a, \mathrm{p}_a) \quad \text{(no sum)} \qquad \forall a \in \{1, \ldots, k\} \tag{9}$$

$$C_{aa} = R(\mathrm{r}_a, \mathrm{r}_a) \quad \text{(no sum)} \qquad \forall a \in \{0, \ldots, k-1\}. \tag{10}$$

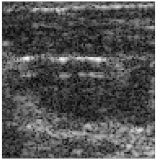
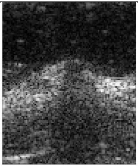
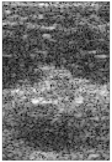
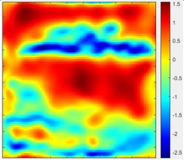
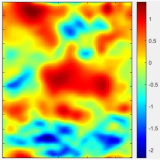
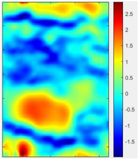
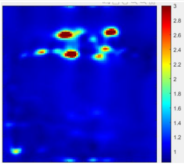
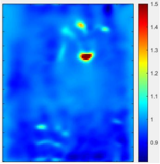
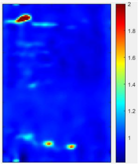
Simulated data

- We consider the inverse elasticity problem where the state u is displacements and parameter m = log(Shear Modulus).



Elastic modulus maps of breast masses: UQ

Data courtesy of M. Fatemi at Mayo Clinic and T.J. Hall at University of Wisconsin

	Subject 1	Subject 2	Subject 3
Ultrasound B-mode			
MAP estimate			
Pointwise variance estimate			
$\frac{\lvert\lvert u-\tilde{u}\rvert\rvert}{\lvert\lvert\tilde{u}\rvert\rvert}$	0.0498	0.0461	0.0401
Information gain (nats)	264.0	161.2	220.7

Conclusions

- We develop a method to construct an approximation for the Hessian "for free" using components obtained during the process of optimization.
- Our method takes advantage of the conjugacies of the directions that comprise the Krylov space used to build the solution.
- The UFL interface in FEniCS facilitates easy implementation of our method.
- Thank you!

Sources I

[Ambartsumyan et al., 2020] Ambartsumyan, I., Boukaram, W., Bui-Thanh, T., Ghattas, O., Keyes, D., Stadler, G., Turkiyyah, G., and Zampini, S. (2020).
Hierarchical matrix approximations of hessians arising in inverse problems governed by pdes.
SIAM Journal on Scientific Computing, 42(5):A3397–A3426.

[Bardsley et al., 2020] Bardsley, J. M., Cui, T., Marzouk, Y. M., and Wang, Z. (2020).
Scalable optimization-based sampling on function space.
SIAM Journal on Scientific Computing, 42(2):A1317–A1347.

[Bui-Thanh et al., 2013] Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G. (2013).
A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion.
SIAM Journal on Scientific Computing, 35(6):A2494–A2523.

[Chang et al., 2014] Chang, J. C., Savage, V. M., and Chou, T. (2014).
A path-integral approach to bayesian inference for inverse problems using the semiclassical approximation.
Journal of Statistical Physics, 157(3):582–602.

[Chen and Ghattas, 2020] Chen, P. and Ghattas, O. (2020).
Projected stein variational gradient descent.
arXiv preprint arXiv:2002.03469.

[Cui et al., 2016] Cui, T., Marzouk, Y., and Willcox, K. (2016).
Scalable posterior approximations for large-scale bayesian inverse problems via likelihood-informed parameter and state reduction.
Journal of Computational Physics, 315:363–387.

[Fatehiboroujeni et al., 2020] Fatehiboroujeni, S., Petra, N., and Goyal, S. (2020).
Linearized bayesian inference for young's modulus parameter field in an elastic model of slender structures.
Proceedings of the Royal Society A, 476(2238):20190476.

Sources II

[Flath et al., 2011] Flath, H. P., Wilcox, L. C., Akçelik, V., Hill, J., van Bloemen Waanders, B., and Ghattas, O. (2011).
Fast algorithms for bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations.
SIAM Journal on Scientific Computing, 33(1):407–432.

[Petra et al., 2014] Petra, N., Martin, J., Stadler, G., and Ghattas, O. (2014).
A computational framework for infinite-dimensional bayesian inverse problems, part ii: Stochastic newton mcmc with application to ice sheet flow inverse problems.
SIAM Journal on Scientific Computing, 36(4):A1525–A1555.

[Saibaba et al., 2020] Saibaba, A. K., Chung, J., and Petroske, K. (2020).
Efficient krylov subspace methods for uncertainty quantification in large bayesian linear inverse problems.
Numerical Linear Algebra with Applications, 27(5):e2325.

[Spantini et al., 2015] Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L., and Marzouk, Y. (2015).
Optimal low-rank approximations of bayesian linear inverse problems.
SIAM Journal on Scientific Computing, 37(6):A2451–A2487.

[Vigliotti et al., 2018] Vigliotti, A., Csányi, G., and Deshpande, V. (2018).
Bayesian inference of the spatial distributions of material properties.
Journal of the Mechanics and Physics of Solids, 118:74–97.

[Villa et al., 2019] Villa, U., Petra, N., and Ghattas, O. (2019).
hippylib: an extensible software framework for large-scale inverse problems governed by pdes; part i: deterministic inversion and linearized bayesian inference.
arXiv preprint arXiv:1909.03948.

[Zou et al., 2019] Zou, Z., Mukherjee, S., Antil, H., and Aquino, W. (2019).
Adaptive particle-based approximations of the gibbs posterior for inverse problems.
arXiv preprint arXiv:1907.01551.