NTNU
Norwegian University of
Science and Technology

Computation of the Hessian for Full Waveform Inversion

Vegard Stenhjem Hagen

Iterative methods[1]

- Searching for a model **m** that describes the earth.
- Solve the wave equation

$$\frac{\partial^2}{\partial t^2}\mathbf{u} - \mathbf{c}^2\nabla^2\mathbf{u} = \mathbf{f}, \quad \mathbf{m} = [\mathbf{c}],$$

  where **u** is displacement and the model **m** only consists of velocities **c**, **f** is a source function.
- Compare with true data $\mathbf{u}_0$ using a misfit function

$$\Psi\left(\mathbf{u}(\mathbf{m},\mathbf{x}_r),\mathbf{u}_0\right) = \frac{1}{2}\left(\mathbf{u}_0(\mathbf{x}_r) - \mathbf{u}(\mathbf{m},\mathbf{x}_r)\right)^T\left(\mathbf{u}_0(\mathbf{x}_r) - \mathbf{u}(\mathbf{m},\mathbf{x}_r)\right).$$

- Iterative approach. Find a model update $\delta\mathbf{m}_k$ that decreases the misfit

$$\Psi(\mathbf{m}_{k+1} = \mathbf{m}_k + \delta\mathbf{m}_k) < \Psi(\mathbf{m}_k).$$

[1] Tarantola 1984; Mora 1987; Fichtner et al. 2006; Fichtner 2011.

Iterative methods

— To do this we can calculate the gradient of the misfit

$$\mathbf{J}(\mathbf{m}+\delta\mathbf{m}) = \nabla_m \Psi(\mathbf{m}+\delta\mathbf{m}).$$

— By linearising around the Jacobian we get

$$\mathbf{J}(\mathbf{m}+\delta\mathbf{m}) \simeq \mathbf{J}(\mathbf{m}) + \nabla_m \mathbf{J}(\mathbf{m})\,\delta\mathbf{m} = \mathbf{0}.$$

Iterative methods

— To do this we can calculate the gradient of the misfit

$$\mathbf{J}(\mathbf{m}+\delta\mathbf{m}) = \nabla_m \Psi(\mathbf{m}+\delta\mathbf{m}).$$

— By linearising around the Jacobian we get

$$\mathbf{J}(\mathbf{m}+\delta\mathbf{m}) \simeq \mathbf{J}(\mathbf{m}) + \underbrace{\nabla_m \mathbf{J}(\mathbf{m})}_{\mathbb{H}(\mathbf{m})}\delta\mathbf{m} = \mathbf{0}.$$

— The Hessian is then given by

$$\mathbb{H}(\mathbf{m}) = \nabla_m \mathbf{J}(\mathbf{m}) = \nabla_m \nabla_m \Psi(\mathbf{m}).$$

Newton method[2]

— By solving

$$\mathbb{H}(\mathbf{m})\delta\mathbf{m} = -\mathbf{J}(\mathbf{m})$$

for $\delta\mathbf{m}$ we find the next model update.

[2] Virieux and Operto 2009; Métivier et al. 2012; Epanomeritakis et al. 2008.

Newton method[2]

- By solving

$$\mathbb{H}(\mathbf{m})\delta\mathbf{m} = -\mathbf{J}(\mathbf{m})$$

  for $\delta\mathbf{m}$ we find the next model update.
- Iff $\mathbb{H}$ is invertible we can “simply” solve

$$\delta\mathbf{m} = -\mathbb{H}^{-1}\mathbf{J}.$$

[2] Virieux and Operto 2009; Métivier et al. 2012; Epanomeritakis et al. 2008.

Newton method[2]

— By solving

$$\mathbb{H}(\mathbf{m})\delta\mathbf{m} = -\mathbf{J}(\mathbf{m})$$

for $\delta\mathbf{m}$ we find the next model update.

— Iff $\mathbb{H}$ is invertible we can "simply" solve

$$\delta\mathbf{m} = -\mathbb{H}^{-1}\mathbf{J}.$$

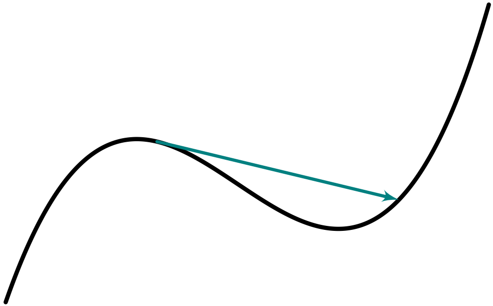
— A common approximation is

$$\delta\mathbf{m} \simeq \alpha\mathbf{J},$$

and a line search for the optimal $\alpha \in \mathbf{R}$.

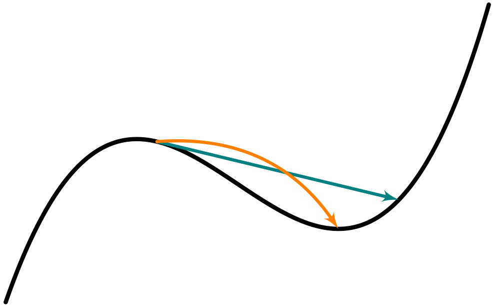
[2]Virieux and Operto 2009; Métivier et al. 2012; Epanomeritakis et al. 2008.

Advantages of using the full Hessian



Linear approximation

Advantages of using the full Hessian



Adjust for slope

Advantages of using the full Hessian[3]

— Fewer iterations are expected.

— Possible to perform resolution analysis.

— Parameter cross-talk analysis.

— Incorporate second order effects.

— Extra transmission-like information.

[3] Pratt et al. 1998; Fichtner and Trampert 2011b; Trampert et al. 2013; Biondi et al. 2015.

Fréchet derivative

— Use the adjoint approach by Tarantola 1984 and further work in Fichtner and Trampert 2011a.

— The Fréchet derivative is defined as

$$\nabla_m \Psi(\mathbf{m})\delta\mathbf{m} = \lim_{\nu \to 0} \frac{1}{\nu}\left[\Psi(\mathbf{m} + \nu\delta\mathbf{m}) - \Psi(\mathbf{m})\right],$$

i.e. the derivative of Ψ with respect to $\mathbf{m}$ in the $\delta\mathbf{m}$ direction.

Fréchet derivative

— Use the adjoint approach by Tarantola 1984 and further work in Fichtner and Trampert 2011a.

— The Fréchet derivative is defined as

$$\nabla_m \Psi(\mathbf{m})\delta\mathbf{m} = \lim_{\nu \to 0} \frac{1}{\nu}\Big[\Psi(\mathbf{m} + \nu\delta\mathbf{m}) - \Psi(\mathbf{m})\Big],$$

i.e. the derivative of Ψ with respect to $\mathbf{m}$ in the $\delta\mathbf{m}$ direction.

— The Jacobian acting on a model perturbation $\delta\mathbf{m}$ we can be written as

$$\mathbf{J}(\mathbf{m})\delta\mathbf{m} = \nabla_m \Psi(\mathbf{m})\delta\mathbf{m}.$$

Fréchet derivative

— Use the adjoint approach by Tarantola 1984 and further work in Fichtner and Trampert 2011a.

— The Fréchet derivative is defined as

$$\nabla_m \Psi(\mathbf{m})\delta\mathbf{m} = \lim_{\nu \to 0} \frac{1}{\nu}\Big[\Psi(\mathbf{m} + \nu\delta\mathbf{m}) - \Psi(\mathbf{m})\Big],$$

i.e. the derivative of Ψ with respect to $\mathbf{m}$ in the $\delta\mathbf{m}$ direction.

— The Jacobian acting on a model perturbation $\delta\mathbf{m}$ we can be written as

$$\mathbf{J}(\mathbf{m})\delta\mathbf{m} = \nabla_m \Psi(\mathbf{m})\delta\mathbf{m}.$$

— For the Hessian $\mathbb{H}$ we can calculate its action on a model perturbation $\delta\mathbf{m}$ as

$$\mathbb{H}(\mathbf{m})\delta\mathbf{m} = \nabla_m \nabla_m \Psi(\mathbf{m})\delta\mathbf{m}.$$

Fréchet kernel

— The Fréchet kernel is defined as the volumetric densities of the Fréchet derivative

$$\mathbf{F} = \frac{\mathrm{d}}{\mathrm{d}V}\nabla_m\Psi$$

giving us the relation

$$\mathbf{J}\delta\mathbf{m} = \nabla_m\Psi\delta\mathbf{m} = \int_G \mathbf{F}\delta\mathbf{m}\mathrm{d}^3\mathbf{x}, \quad G \in \mathbb{R}^3.$$

— By studying **F** we can design more efficient inversion schemes and interpret the results in a meaningful way.

Fréchet kernel

— Introducing the adjoint wavefield $\mathbf{u}^\dagger$ defined as the solution to

$$\frac{\partial^2}{\partial t^2}\mathbf{u}^\dagger - \mathbf{c}^2\nabla^2\mathbf{u}^\dagger = -\big(\mathbf{u}_0(\mathbf{x}_r) - \mathbf{u}(\mathbf{x}_r)\big),$$

which is referred to as backpropagating the residuals.

— The Fréchet kernel can now be calculated as

$$\mathbf{F}(\mathbf{u}^\dagger, \mathbf{u}) = 2\mathbf{c}\int_T \nabla\mathbf{u}^\dagger \nabla\mathbf{u}\,\mathrm{d}t,$$

i.e. by cross correlating the divergence of the forward and adjoint displacement fields and scaling by the background velocity.

Perturbed fields

- In order to more easily calculate the Hessian we need to introduce two more fields.
- The perturbed forward field

$$\begin{aligned} \delta \mathbf{u} &= \nabla_m \mathbf{u} \delta \mathbf{m} \\ &= \lim_{\nu \to 0} \frac{1}{\nu} [\mathbf{u}^\dagger(\mathbf{m} + \nu \delta \mathbf{m}) - \mathbf{u}^\dagger(\mathbf{m})] \end{aligned}$$

- The perturbed adjoint field

$$\begin{aligned} \delta \mathbf{u}^\dagger &= \nabla_m \mathbf{u}^\dagger \delta \mathbf{m} \\ &= \lim_{\nu \to 0} \frac{1}{\nu} [\mathbf{u}(\mathbf{m} + \nu \delta \mathbf{m}) - \mathbf{u}(\mathbf{m})] \end{aligned}$$

Hessian kernels

- The Hessian kernel **H** can be broken down into three parts

$$\mathbf{H} = \mathbf{H}^{\mathbf{u}^\dagger,\delta\mathbf{u}} + \mathbf{H}^{\delta\mathbf{u}^\dagger,\mathbf{u}} + \mathbf{H}^{\mathbf{u}^\dagger,\mathbf{u}}.$$

- Written out these are

$$\mathbf{H}^{\mathbf{u}^\dagger,\delta\mathbf{u}} = 2\mathbf{c}\int_T \nabla(\mathbf{u}^\dagger)\cdot\nabla(\delta\mathbf{u})\mathrm{d}t$$
$$\mathbf{H}^{\delta\mathbf{u}^\dagger,\mathbf{u}} = 2\mathbf{c}\int_T \nabla(\delta\mathbf{u}^\dagger)\cdot\nabla(\mathbf{u})\mathrm{d}t$$
$$\mathbf{H}^{\mathbf{u}^\dagger,\mathbf{u}} = 2\delta\mathbf{c}\int_T \nabla(\mathbf{u}^\dagger)\cdot\nabla(\mathbf{u})\mathrm{d}t$$

Hessian kernels

— The Hessian kernel $\mathbf{H}$ can be broken down into three parts

$$\mathbf{H} = \mathbf{H}^{\mathbf{u}^\dagger,\delta\mathbf{u}} + \mathbf{H}^{\delta\mathbf{u}^\dagger,\mathbf{u}} + \mathbf{H}^{\mathbf{u}^\dagger,\mathbf{u}}.$$

— Written out these are

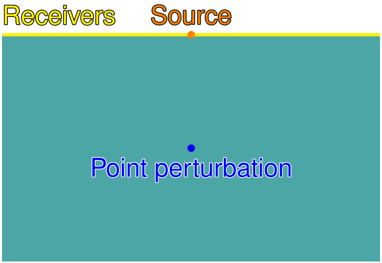
$$\begin{aligned}
\mathbf{H}^{\mathbf{u}^\dagger,\delta\mathbf{u}} &= 2\mathbf{c}\int_T \nabla(\mathbf{u}^\dagger)\cdot\nabla(\delta\mathbf{u})\mathrm{d}t &&= \mathbf{F}(\mathbf{u}^\dagger,\delta\mathbf{u}),\\
\mathbf{H}^{\delta\mathbf{u}^\dagger,\mathbf{u}} &= 2\mathbf{c}\int_T \nabla(\delta\mathbf{u}^\dagger)\cdot\nabla(\mathbf{u})\mathrm{d}t &&= \mathbf{F}(\delta\mathbf{u}^\dagger,\mathbf{u}),\\
\mathbf{H}^{\mathbf{u}^\dagger,\mathbf{u}} &= 2\delta\mathbf{c}\int_T \nabla(\mathbf{u}^\dagger)\cdot\nabla(\mathbf{u})\mathrm{d}t &&= \frac{\delta\mathbf{c}}{\mathbf{c}}\mathbf{F}(\mathbf{u}^\dagger,\mathbf{u}).
\end{aligned}$$

Fields

— We now have four fields we need to calculate.

$\mathbf{u}$ — Forward field.
$\mathbf{u}^{\dagger}$ — Adjoint field.
$\delta\mathbf{u}$ — Perturbed forward field.
$\delta\mathbf{u}^{\dagger}$ — Perturbed adjoint field.

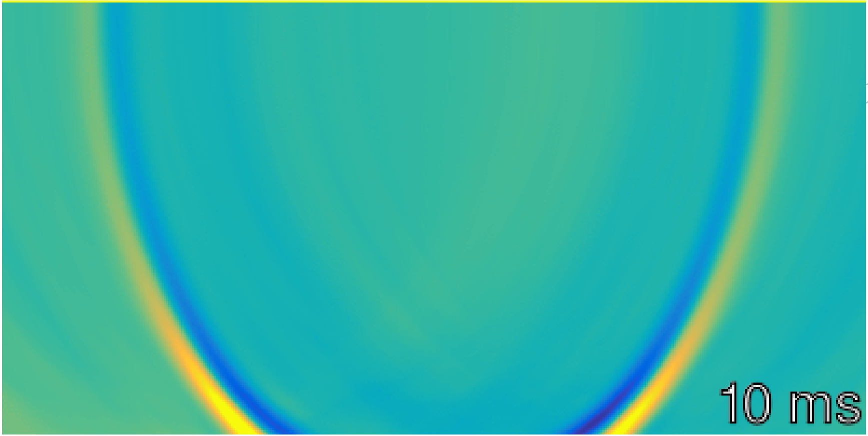
Model



- Acoustic 2-D
- Background 2.0 km/s
- Perturbation 2.2 km/s
- Width 1500 m, Depth 750 m
- 3.75 m × 3.75 m grid cells
- 650 ms in 0.1 ms time steps
- 20 Hz Ricker wavelet

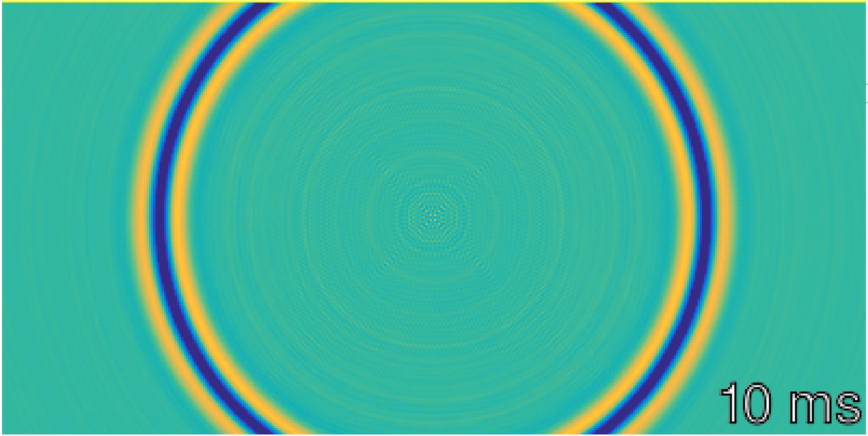
Forward field

Adjoint field



Perturbed forward field

Perturbed adjoint field

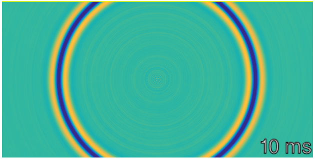


Fields

Forward field

Perturbed forward field

Adjoint field

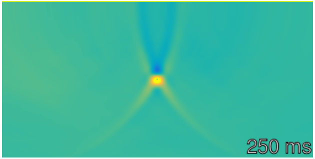


Perturbed adjoint field

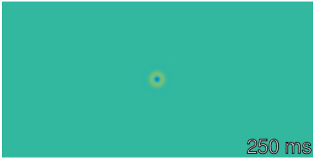
Fields

Forward field

Perturbed forward field



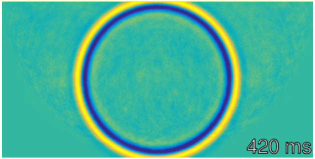
Adjoint field



Perturbed adjoint field

Fields

Forward field



Perturbed forward field

Adjoint field

Perturbed adjoint field

Reflection - Fréchet kernel — $\mathbf{F}(\mathbf{u}^{\dagger}, \mathbf{u})$

Divergence Correlation

Background field

Adjoint field

Reflection — $\mathbf{H}^{\mathbf{u}^\dagger,\delta\mathbf{u}} = \mathbf{F}(\mathbf{u}^\dagger,\delta\mathbf{u})$

Divergence Correlation

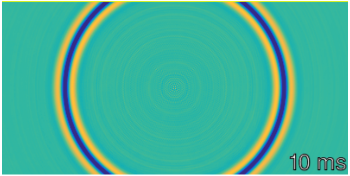
Perturbed forward field

Adjoint Field

Reflection — $\mathbf{H}^{\delta \mathbf{u}^{\dagger},\mathbf{u}} = \mathbf{F}(\delta \mathbf{u}^{\dagger}, \mathbf{u})$

Divergence Correlation

Background field



Perturbed adjoint field

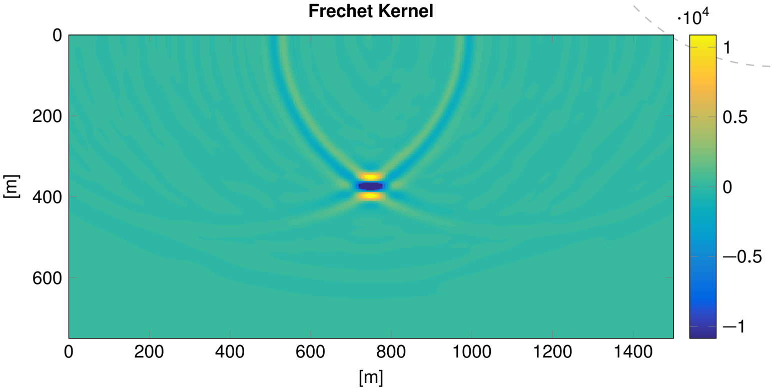
Reflection — $\mathbf{H}^{\mathbf{u}^{\dagger},\mathbf{u}} = \frac{\delta \mathbf{c}}{\mathbf{c}} \mathbf{F}(\mathbf{u}^{\dagger}, \mathbf{u})$

Divergence Correlation

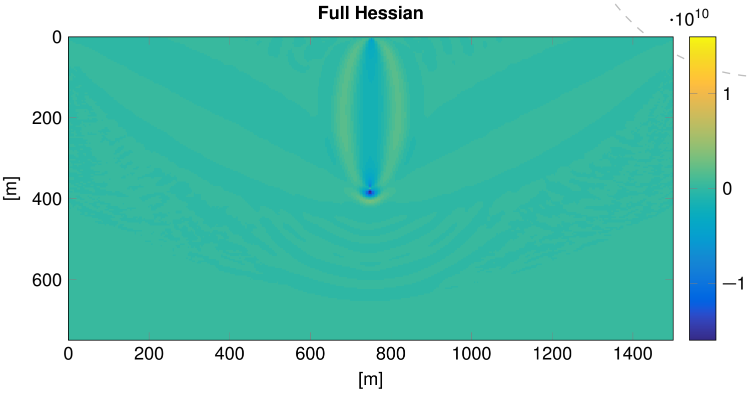
Background field

Adjoint field

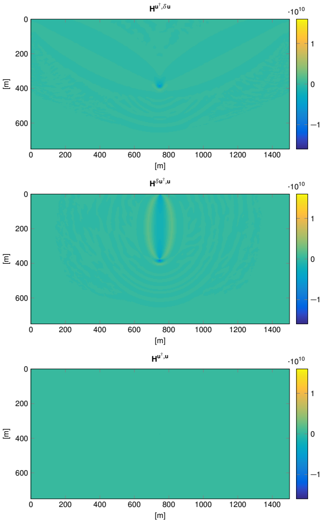
Reflection - Fréchet kernel — $F(u^{\dagger}, u)$



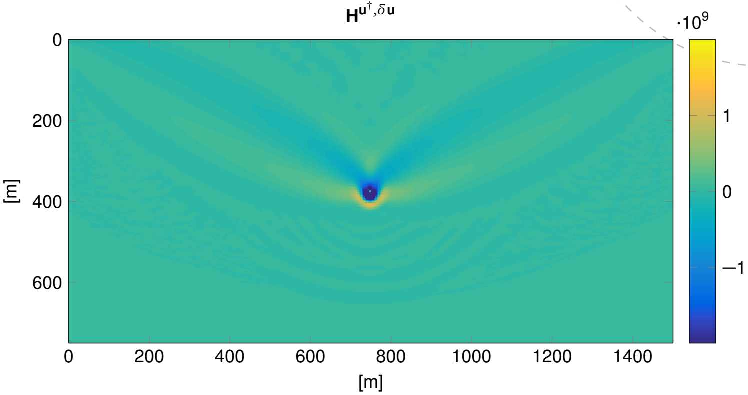
Reflection - Full Hessian Kernel — H



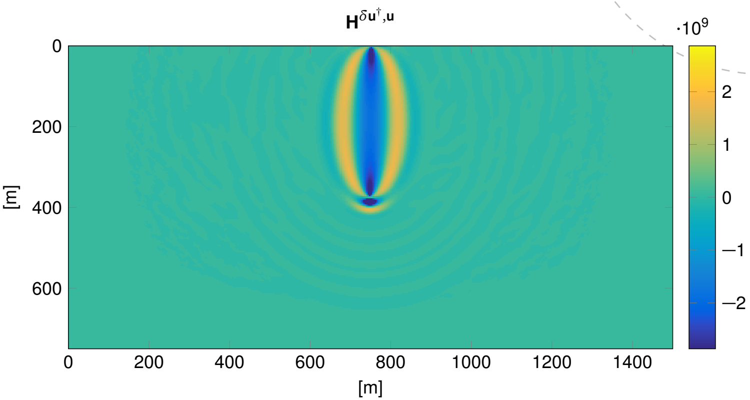
Reflection


$\mathbf{H}^{\mathbf{u}^\dagger,\delta\mathbf{u}}$
$\mathbf{H}^{\delta\mathbf{u}^\dagger,\mathbf{u}}$
$\mathbf{H}^{\mathbf{u}^\dagger,\mathbf{u}}$
$\cdot 10^{10}$
[m]
0
200
400
600
800
1000
1200
1400
1
0
−1

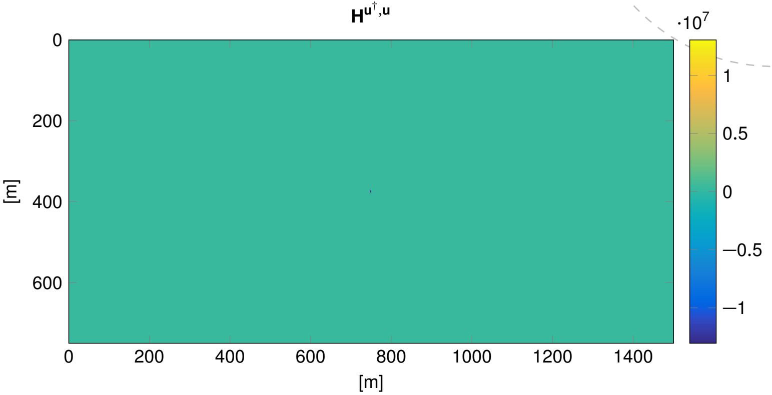
Reflection — $\mathbf{H}^{\mathbf{u}^\dagger, \delta\mathbf{u}} = \mathbf{F}(\mathbf{u}^\dagger, \delta\mathbf{u})$



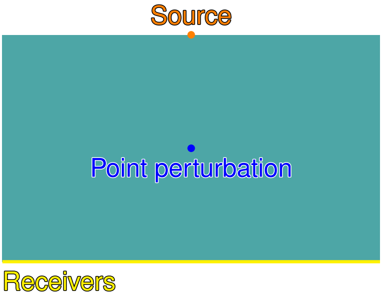
Reflection — $\mathbf{H}^{\delta \mathbf{u}^{\dagger},\mathbf{u}} = \mathbf{F}(\delta \mathbf{u}^{\dagger}, \mathbf{u})$



Reflection — $\mathbf{H}^{\mathbf{u}^\dagger,\mathbf{u}} = \frac{\delta \mathbf{c}}{\mathbf{c}} \mathbf{F}(\mathbf{u}^\dagger, \mathbf{u})$

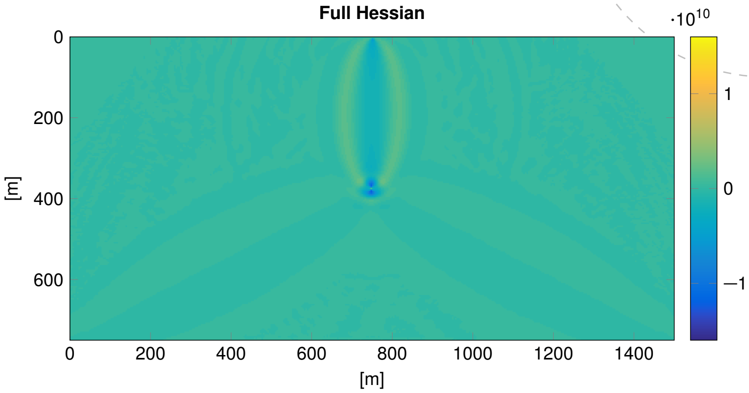


Model

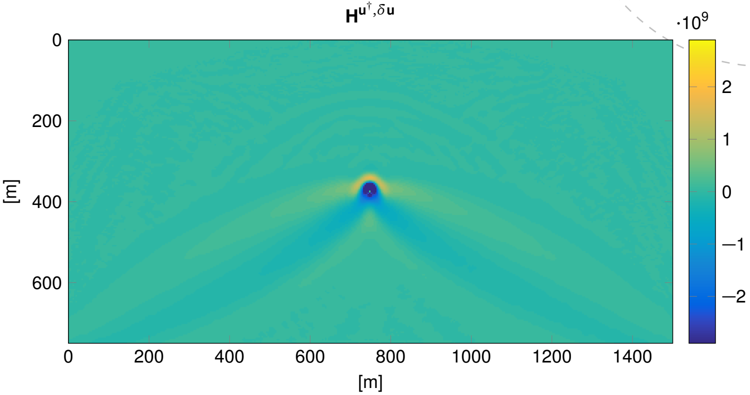


- Acoustic 2-D
- Background 2.0 km/s
- Perturbation 2.2 km/s
- Width 1500 m, Depth 750 m
- 3.75 m × 3.75 m grid cells
- 650 ms in 0.1 ms time steps
- 20 Hz Ricker wavelet

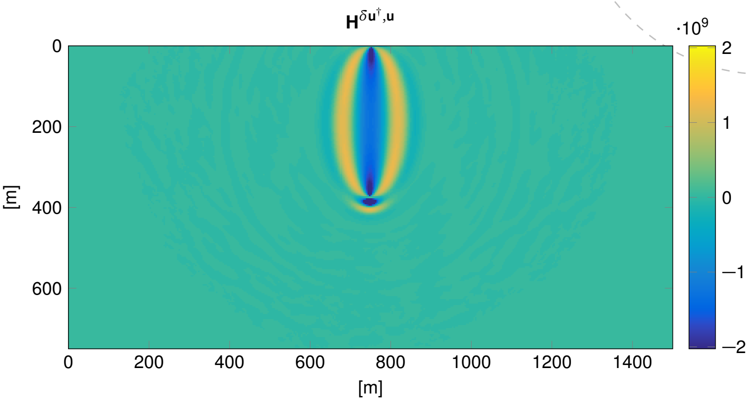
Transmission - Full Hessian Kernel — H



Transmission — $\mathbf{H}^{\mathbf{u}^{\dagger},\delta\mathbf{u}} = \mathbf{F}(\mathbf{u}^{\dagger},\delta\mathbf{u})$



Transmission — $\mathbf{H}^{\delta\mathbf{u}^\dagger,\mathbf{u}} = \mathbf{F}(\delta\mathbf{u}^\dagger, \mathbf{u})$



Conclusions

— Need twice the amount of fields. More computational expensive than linear methods.

Conclusions

- Need twice the amount of fields. More computational expensive than linear methods.
- We can compute the action of the Hessian on a model perturbation without the need to calculate the whole Hessian.

Conclusions

— Need twice the amount of fields. More computational expensive than linear methods.

— We can compute the action of the Hessian on a model perturbation without the need to calculate the whole Hessian.

— Much of the code can be recycled.

Conclusions

- Need twice the amount of fields. More computational expensive than linear methods.
- We can compute the action of the Hessian on a model perturbation without the need to calculate the whole Hessian.
- Much of the code can be recycled.
- Expected to converge in fewer iterations.

Conclusions

- Need twice the amount of fields. More computational expensive than linear methods.
- We can compute the action of the Hessian on a model perturbation without the need to calculate the whole Hessian.
- Much of the code can be recycled.
- Expected to converge in fewer iterations.
- Possible to perform accuracy and resolution analysis on the results.

Conclusions

- Need twice the amount of fields. More computational expensive than linear methods.
- We can compute the action of the Hessian on a model perturbation without the need to calculate the whole Hessian.
- Much of the code can be recycled.
- Expected to converge in fewer iterations.
- Possible to perform accuracy and resolution analysis on the results.
- Parameter cross-talk analysis.

Future work

- Implement in 3-D for Madagascar.
- Calculate the Hessian for an elastic medium.
- Write a full Newton inversion algorithm.
- Investigate cross-talk in different parametrisations.

References

- Tarantola, A. (1984) Inversion of seismic reflection data in the acoustic approximation in: *Geophysics*, **49**:8, 1259 doi: 10.1190/1.1441754
- Mora, P. (1987) Nonlinear two-dimensional elastic inversion of multioffset seismic data in: *Geophysics*, **52**:9, 1211 doi: 10.1190/1.1442384
- Fichtner, A., H. P. Bunge and H. Igel (2006) The adjoint method in seismology. I. Theory in: *Physics of the Earth and Planetary Interiors*, **157**:1-2, 86–104 doi: 10.1016/j.pepi.2006.03.016
- Fichtner, A. (2011) Full Seismic Waveform Modelling and Inversion in: *Advances*, 1981, 83–88 doi: 10.1007/978-3-642-15807-0
- Virieux, J. and S. Operto (2009) An overview of full-waveform inversion in exploration geophysics in: *Geophysics*, **74**:6, WCC1 doi: 10.1190/1.3238367
- Métivier, L, R Brossier, J Virieux and S Operto (2012) The truncated Newton method for Full Waveform Inversion in: *Journal of Physics: Conference Series*, **386**:2, 012013 doi: 10.1088/1742-6596/386/1/012013
- Epanomeritakis, I, V Akçelik, O Ghattas and J Bielak (2008) A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion in: *Inverse Problems*, **24**:3, 034015 doi: 10.1088/0266-5611/24/3/034015
- Pratt, R. G., C Shin and G. J. Hicks (1998) Gauss-Newton and full Newton methods in frequency domain seismic waveform inversion. in: *Geophysical Journal International*, **133**: 341–362
- Fichtner, A. and J. Trampert (2011b) Resolution analysis in full waveform inversion in: *Geophysical Journal International*, **187**:3, 1604–1624 doi: 10.1111/j.1365-246x.2011.05218.x
- Trampert, J., A. Fichtner and J. Ritsema (2013) Resolution tests revisited: The power of random numbers in: *Geophysical Journal International*, **192**: 676–680 doi: 10.1093/gji/ggs057
- Biondi, B., E. Biondi, M. Maharramov and Y. Ma (2015) Dissection of the full-waveform inversion Hessian in: *Unpublished*,
- Fichtner, A. and J. Trampert (2011a) Hessian kernels of seismic data functionals based upon adjoint techniques in: *Geophysical Journal International*, **185**: 775–798 doi: 10.1111/j.1365-246X.2011.04966.x

Acknowledgements

ROSE Consortium

NTNU