Time Domain Full Waveform Inversion involving Discontinuous Galerkin approximation

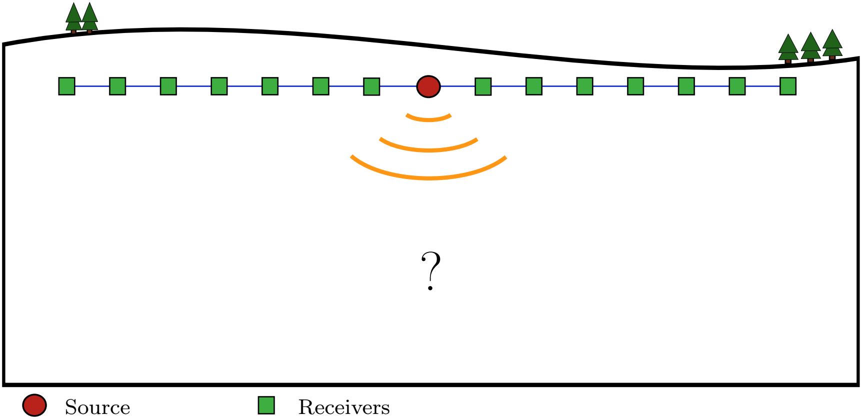
Waves 2019

Pierre Jacquet
pierre.jacquet@inria.fr

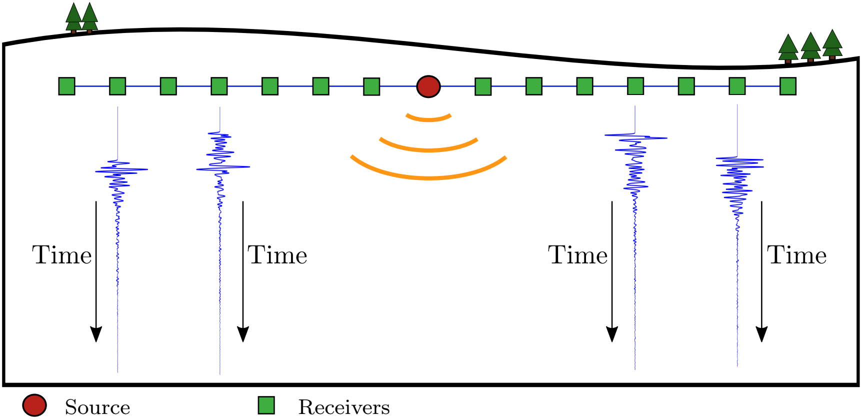
Barucq Hélène, Diaz Julien, Calandra Henri
First year PhD Student
Inria - Magique 3D - DIP

Pau, FRANCE

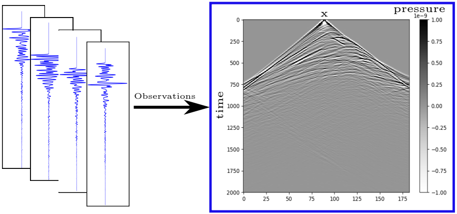
Seismic Acquisition



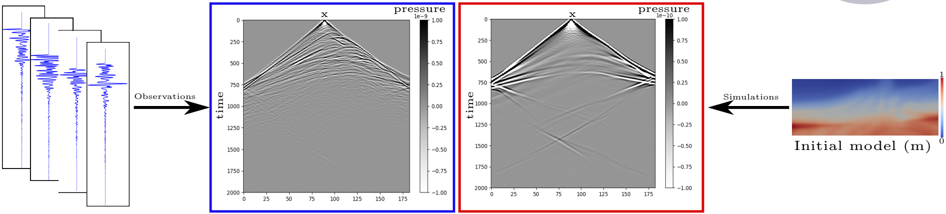
Seismic Acquisition



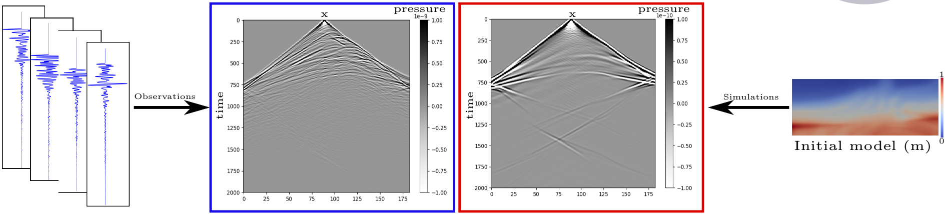
FWI Workflow



FWI Workflow



FWI Workflow

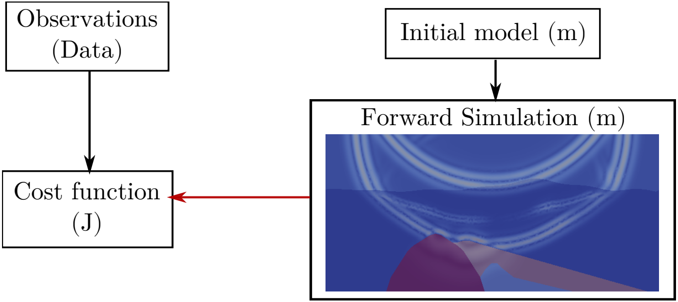


Cost function to minimize :

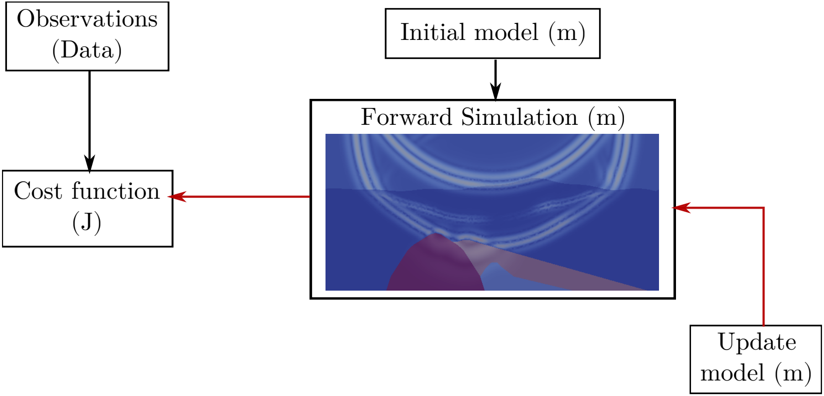
$$\mathcal{J}(m) = \frac{1}{2}||d_{obs} - \mathcal{F}(m)||^2 dt$$

- $\mathcal{F}(m)$ is the restriction on the receivers of the simulated waves in the media m. (With $m = \boldsymbol{c}, \boldsymbol{\rho}, \boldsymbol{\kappa}$...)
- FWI iterates until $\mathcal{J}(m) \longrightarrow 0$

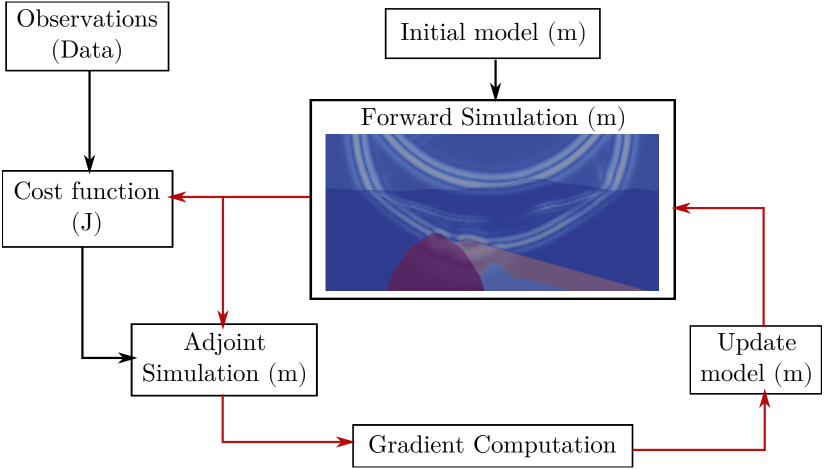
FWI Workflow



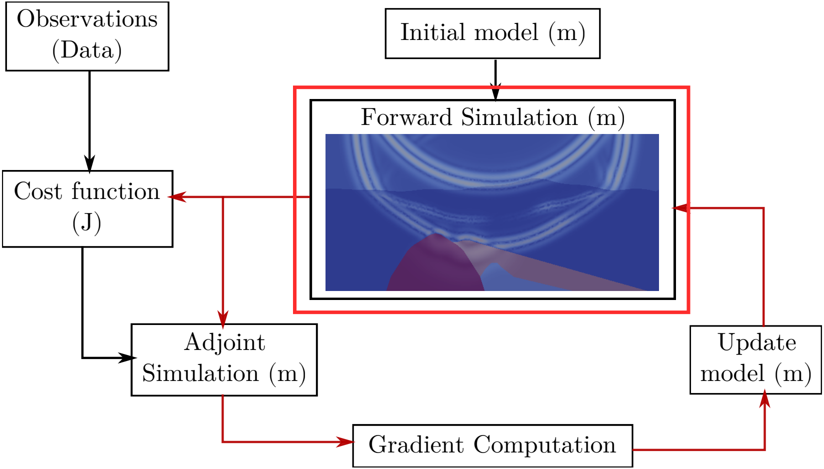
FWI Workflow



FWI Workflow



FWI Workflow



Continuous Forward Model

First order acoustic wave equation

$$\begin{cases} \dfrac{1}{\rho c^2}\dfrac{\partial \boldsymbol{p}}{\partial t} + \nabla \cdot \boldsymbol{v} = f_p \quad \text{on } \boldsymbol{\Omega} \\ \rho \dfrac{\partial \boldsymbol{v}}{\partial t} + \nabla \boldsymbol{p} = f_v \quad \text{on } \boldsymbol{\Omega} \\ \boldsymbol{p} = 0 \quad \text{on } \boldsymbol{\Gamma_1} \\ \dfrac{\partial \boldsymbol{p}}{\partial t} + \boldsymbol{c} \nabla \boldsymbol{p} \cdot \boldsymbol{n} = 0 \quad \text{on } \boldsymbol{\Gamma_2} \end{cases}$$

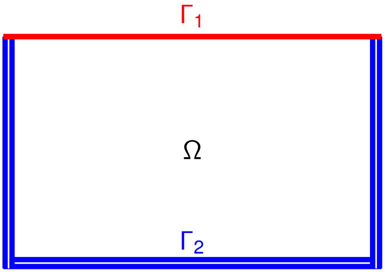


Figure: Domain with Absorbing Boundary Conditions

Discrete Forward Model

Space Discretization :
Discontinuous Galerkin Elements

- Nodal (Lagrangian / Jacobian)
- Modal (Bernstein-Bézier)

Discrete Forward Model

Space Discretization :
Discontinuous Galerkin Elements

- Nodal (Lagrangian / Jacobian)
- Modal (Bernstein-Bézier)

Semi-discretized model :

$$\frac{\partial}{\partial t}\bar{\boldsymbol{U}}(t) = A\bar{\boldsymbol{U}}(t) + \bar{\boldsymbol{F}}(t)$$

with :

$$\bar{\boldsymbol{U}}(t) = \begin{pmatrix} \bar{\boldsymbol{P}}(t) \\ \bar{\boldsymbol{V}}(t) \end{pmatrix}$$

Discrete Forward Model

Space Discretization :
Discontinuous Galerkin Elements

- Nodal (Lagrangian / Jacobian)
- Modal (Bernstein-Bézier)

Time schemes :

- Runge Kutta 2/4
- Adams Bashforth 3

Semi-discretized model :

$$\frac{\partial}{\partial t}\bar{\boldsymbol{U}}(t) = A\bar{\boldsymbol{U}}(t) + \bar{\boldsymbol{F}}(t)$$

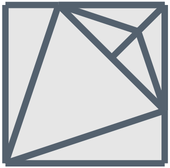
with :

$$\bar{\boldsymbol{U}}(t) = \begin{pmatrix} \bar{\boldsymbol{P}}(t) \\ \bar{\boldsymbol{V}}(t) \end{pmatrix}$$

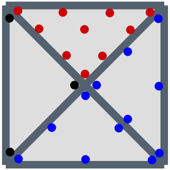
Forward time steps

Asset of Discontinuous Galerkin Methods :

- Unstructured grid (enable to match the topography and media irregularities)
- Robust to physical discontinuities
- hp-adaptivity
- Massively parallel performance properties



h-adaptivity



p-adaptivity with P1, P2, P3 elements

Outline

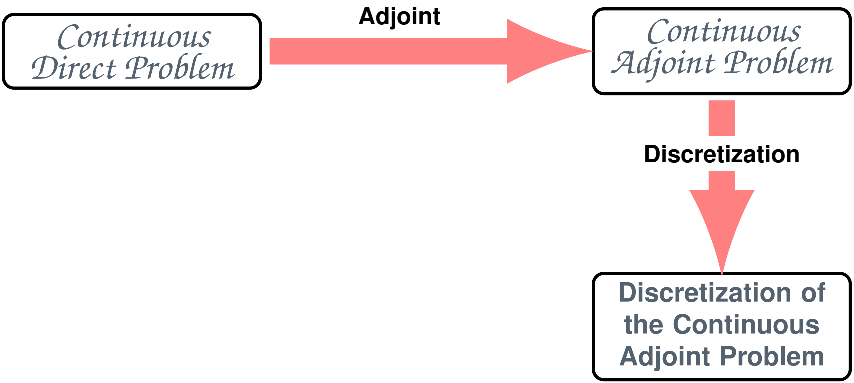
Adjoint Formulation

$$\boxed{\begin{matrix}\textit{Continuous} \\ \textit{Direct Problem}\end{matrix}}$$

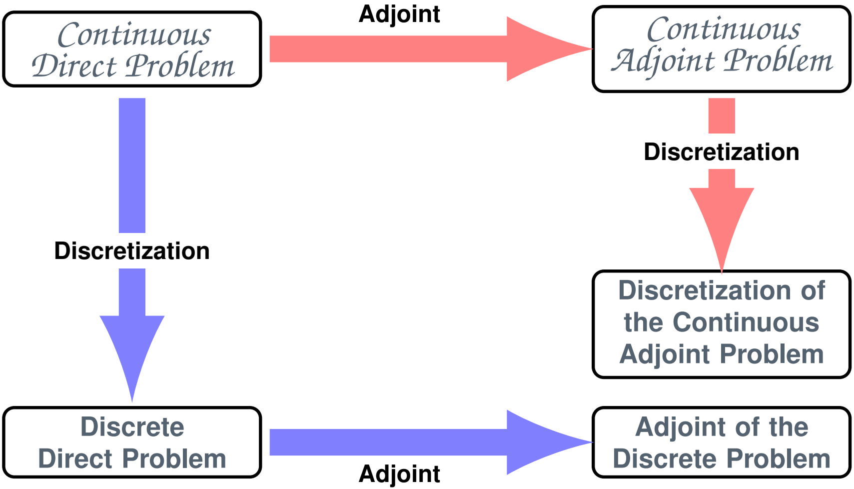
Adjoint Formulation

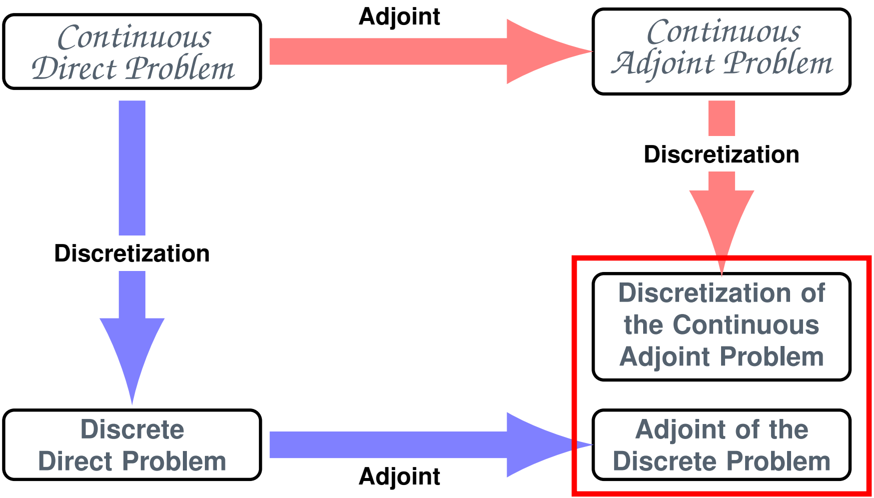
Continuous Direct Problem $\xrightarrow{\textbf{Adjoint}}$ Continuous Adjoint Problem

Adjoint Formulation



Adjoint Formulation




Continuous Direct Problem
Adjoint
Continuous Adjoint Problem
Discretization
Discretization
Discretization of the Continuous Adjoint Problem
Discrete Direct Problem
Adjoint
Adjoint of the Discrete Problem

AtD : Adjoint then Discretized Strategy

$$\mathcal{J}(\boldsymbol{p}) = \frac{1}{2}||d_{obs} - R\boldsymbol{p}||^2$$

$$\begin{cases} \dfrac{1}{\rho \boldsymbol{c}^2}\dfrac{\partial \boldsymbol{p}}{\partial t} + \nabla \cdot \boldsymbol{v} = f_p \quad \text{on } \boldsymbol{\Omega} \\ \rho \dfrac{\partial \boldsymbol{v}}{\partial t} + \nabla \boldsymbol{p} = 0 \quad \text{on } \boldsymbol{\Omega} \\ \boldsymbol{p} = 0 \quad \text{on } \boldsymbol{\Gamma_1} \\ \dfrac{\partial \boldsymbol{p}}{\partial t} + \boldsymbol{c}\nabla \boldsymbol{p} \cdot \boldsymbol{n} = 0 \quad \text{on } \boldsymbol{\Gamma_2} \\ \boldsymbol{p}(0) = 0, \quad \boldsymbol{v}(0) = 0 \end{cases}$$

$$\begin{cases} \dfrac{1}{\rho \boldsymbol{c}^2}\dfrac{\partial \boldsymbol{\lambda_1}}{\partial t} + \nabla \cdot \boldsymbol{\lambda_2} = \dfrac{\partial \mathcal{J}}{\partial \boldsymbol{p}} \quad \text{on } \boldsymbol{\Omega} \\ \rho \dfrac{\partial \boldsymbol{\lambda_2}}{\partial t} + \nabla \boldsymbol{\lambda_1} = 0 \quad \text{on } \boldsymbol{\Omega} \\ \boldsymbol{\lambda_1} = 0 \quad \text{on } \boldsymbol{\Gamma_1} \\ \dfrac{\partial \boldsymbol{\lambda_1}}{\partial t} - \boldsymbol{c}\nabla \boldsymbol{\lambda_1} \cdot \boldsymbol{n} = 0 \quad \text{on } \boldsymbol{\Gamma_2} \\ \boldsymbol{\lambda_1}(T) = 0, \quad \boldsymbol{\lambda_2}(T) = 0 \end{cases}$$

$t \in [0, T]$ $\qquad$ $t \in [T, 0]$

AtD : Adjoint then Discretized Strategy

$$\mathcal{J}(\boldsymbol{p}) = \frac{1}{2}||d_{obs} - R\boldsymbol{p}||^2$$

$$\begin{cases} \dfrac{\partial \bar{\boldsymbol{U}}^n}{\partial t} = A\bar{\boldsymbol{U}}^n + \bar{\boldsymbol{F}}^n \\ \text{With :} \quad \bar{\boldsymbol{U}}^n = \begin{pmatrix} \bar{\boldsymbol{P}}^n \\ \bar{\boldsymbol{V}}^n \end{pmatrix} \end{cases}$$

$$\begin{cases} \dfrac{\partial \bar{\boldsymbol{\Lambda}}^n}{\partial t} = A\bar{\boldsymbol{\Lambda}}^n + R^*(R\bar{\boldsymbol{U}}^n - d_{obs}) \\ \text{With :} \quad \bar{\boldsymbol{\Lambda}}^n = \begin{pmatrix} \bar{\boldsymbol{\Lambda}}_1{}^n \\ \bar{\boldsymbol{\Lambda}}_2{}^n \end{pmatrix} \end{cases}$$

Time-steps going Forward

Time-steps going Backward

DtA : Discretize then Adjoint Strategy

RK4 example

All time scheme can be summed-up such as :

$$\mathbf{L}\bar{\mathbf{U}} = \mathbf{E}\bar{\mathbf{F}}$$

RK4 time-scheme leads to :

$$\bar{\mathbf{U}}^{n+1} = B\bar{\mathbf{U}}^{n} + \mathbf{C_0}\bar{\mathbf{F}}^{n} + \mathbf{C_{\frac{1}{2}}}\bar{\mathbf{F}}^{n+\frac{1}{2}} + \mathbf{C_1}\bar{\mathbf{F}}^{n+1}$$

$$\mathbf{L}\bar{\mathbf{U}} = \mathbf{E}\bar{\mathbf{F}} = \bar{\mathbf{G}}$$

$$\begin{pmatrix} I & & & & \\ -B & I & & & \\ & -B & I & & \\ & & \ddots & \ddots & \\ & & & -B & I \end{pmatrix} \begin{pmatrix} \bar{\mathbf{U}}^0 \\ \bar{\mathbf{U}}^1 \\ \bar{\mathbf{U}}^2 \\ \vdots \\ \bar{\mathbf{U}}^n \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{G}}^0 \\ \bar{\mathbf{G}}^1 \\ \bar{\mathbf{G}}^2 \\ \vdots \\ \bar{\mathbf{G}}^n \end{pmatrix}$$

DtA : Discretize then Adjoint Strategy

All time scheme can be summed-up such as :

$$\boldsymbol{L}\bar{\boldsymbol{U}} = \boldsymbol{E}\bar{\boldsymbol{F}}$$

We are looking for a Discrete Adjoint state satisfying :

$$\boldsymbol{L}^{*}\bar{\boldsymbol{\Lambda}} = -R^{*}(d_{obs} - R\bar{\boldsymbol{U}})$$

With the adjoint operator $\boldsymbol{L}^{*}$ satisfying :

$$< \boldsymbol{L}\bar{\boldsymbol{U}}, \bar{\boldsymbol{\Lambda}} > = < \bar{\boldsymbol{U}}, \boldsymbol{L}^{*}\bar{\boldsymbol{\Lambda}} >$$

DtA : Discretize then Adjoint Strategy

All time scheme can be summed-up such as :

$$\boldsymbol{L}\bar{\boldsymbol{U}} = \boldsymbol{E}\bar{\boldsymbol{F}} = \bar{\boldsymbol{G}}$$

We are looking for a Discrete Adjoint state satisfying :

$$\boldsymbol{L}^*\bar{\boldsymbol{\Lambda}} = -R^*(d_{obs} - R\bar{\boldsymbol{U}}) = \bar{\boldsymbol{D}}$$

With the adjoint operator $\boldsymbol{L}^*$ satisfying :

$$< \boldsymbol{L}\bar{\boldsymbol{U}}, \bar{\boldsymbol{\Lambda}} > = < \bar{\boldsymbol{U}}, \boldsymbol{L}^*\bar{\boldsymbol{\Lambda}} >$$

$$< \bar{\boldsymbol{G}}, \bar{\boldsymbol{\Lambda}} > = < \bar{\boldsymbol{U}}, \bar{\boldsymbol{D}} > \quad \text{(Adjoint Test)}$$

Adjoint test succeeds $\Longleftrightarrow$ operator $\boldsymbol{L}^*$ well established

Adjoint Strategies Comparison

Adjoint Then Discretize

- + Physical approach
- + Same discrete operators for Forward and Backward
- - - Inexact gradient [?]

Discretize then Adjoint

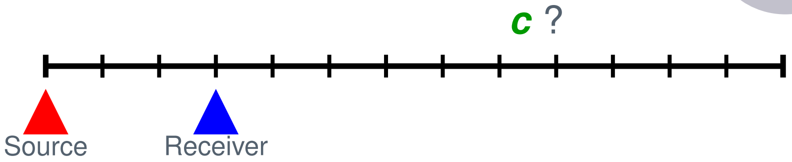
- + Numerical approach
- + Has an Adjoint Test
- - Tremendous work to develop the adjoint operators
- ? Non-consistency of the adjoint state [?]

[1] Sirkes, Ziv and Tziperman, Eli
Finite Difference of Adjoint or Adjoint of Finite Difference ?
1997

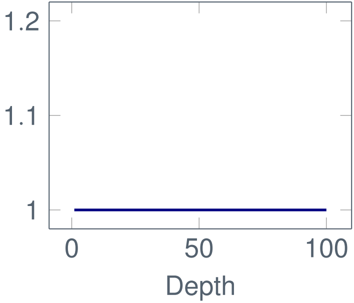
[2] Sei Alain and Symes William
A Note on Consistency and Adjointness for Numerical Schemes
1997

Outline

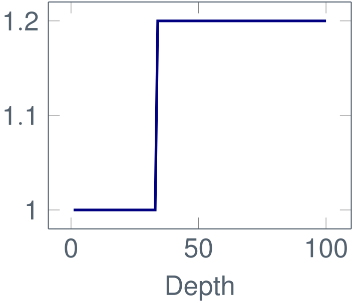
1D Preliminary tests



Initial *c* Model



Target *c* Model



1D Preliminary tests :

1D FWI :

- Lagrange / B-Bézier Operators
- RK4 / AB3 time-schemes

Gradient expression :

$$\nabla_{\boldsymbol{c}} \mathcal{J} = - \int_0^T \int_\Omega \frac{2}{\rho \boldsymbol{c}^3} \frac{\partial \boldsymbol{p}}{\partial t} \boldsymbol{\lambda_1} d\Omega dt$$

1D Preliminary tests :

1D FWI :

- Lagrange / B-Bézier Operators
- RK4 / AB3 time-schemes

Adjoint test passed with :

- With a canonical space inner-product ($< u, v >_X = \sum_i u_i v_i$)
- With a M-space inner product ($< u, v >_X^M = < Mu, v >_X$)

Gradient expression :

$$\nabla_{\boldsymbol{c}} \mathcal{J} = - \int_0^T \int_\Omega \frac{2}{\rho \boldsymbol{c}^3} \frac{\partial \boldsymbol{p}}{\partial t} \boldsymbol{\lambda_1} d\Omega dt$$

```
./run
--- Adjoint test ----
 inner product U/D 553123.57586755091
 inner product G/Q 553123.57586756046
```

1D FWI :

- Lagrange / B-Bézier Operators
- RK4 / AB3 time-schemes

Adjoint test passed with :

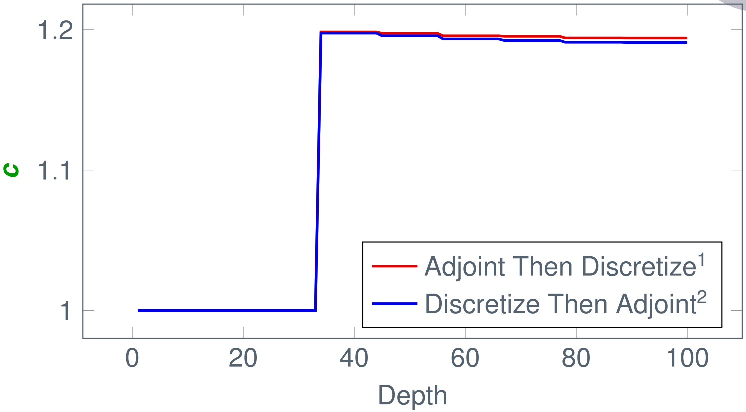
- With a canonical space inner-product ($< u, v >_X = \sum_i u_i v_i$)
- With a M-space inner product ($< u, v >_X^M = < Mu, v >_X$)

Gradient expression :

$$\nabla_{\boldsymbol{c}} \mathcal{J} = - \int_0^T \int_\Omega \frac{2}{\rho \boldsymbol{c}^3} \frac{\partial \boldsymbol{p}}{\partial t} \boldsymbol{\lambda_1} d\Omega dt$$

```
./run
--- Adjoint test ----
 inner product U/D 553123.57586755091
 inner product G/Q 553123.57586756046
./run
--- Adjoint test ----
 inner product U/D -75077.332007383695
 inner product G/Q -75077.332007386358
./run
--- Adjoint test ----
 inner product U/D 125669.89223600870
 inner product G/Q 125669.89223600952
```

1D Velocity Model Reconstructions

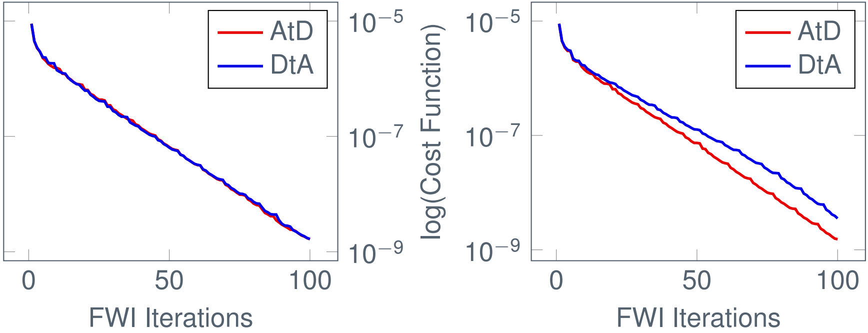


c Model at the 100th FWI iteration

[1] With Bernstein-Bézier elements
[2] With canonical scalar product

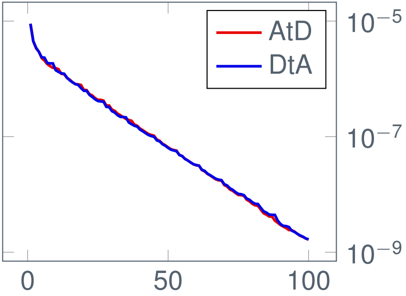
1D Velocity Model Reconstructions

With RK4 : With AB3 :

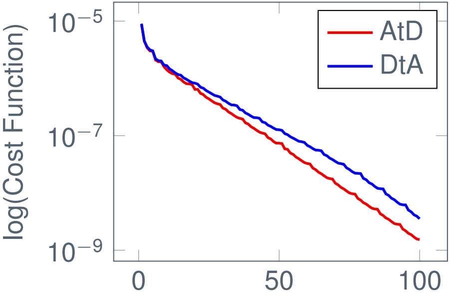


1D Velocity Model Reconstructions

With RK4 :



With AB3 :



- For RK4 scheme : Similar convergency
- For AB3 scheme : AtD is slighly better than DtA
- The slope strongly depends on the optimizer -> Impossibilty to conclude

2D Time Domain Reconstruction

2D FWI :

- Developped in Total environnement (DIP[3])
- Nodal Space Operators (Lagrangian/Jacobian)
- Modal Space Operators (Bernstein-Bézier)
- Runge Kutta 2/4 and Adams Bashforth time-schemes

Discretize Then Adjoint strategy not implemented :

- Tremendous task in a complex industrial code

[3] http://dip.inria.fr/

2D Time Domain Reconstruction

2D FWI :

- Developped in Total environnement (DIP[4])
- Nodal Operators (Lagrangian/Jacobian)
- Modal Operators (Bernstein-Bézier)
- Runge Kutta 2/4 and Adams Bashforth time-schemes

Gradient expression :

$$\nabla_{\frac{1}{\kappa}} \mathcal{J} = \int_0^T \int_\Omega \frac{\partial \boldsymbol{p}}{\partial t} \boldsymbol{\lambda_1} \, d\Omega dt \qquad \text{with : } \boldsymbol{\kappa} = \boldsymbol{\rho} \boldsymbol{c}^2$$

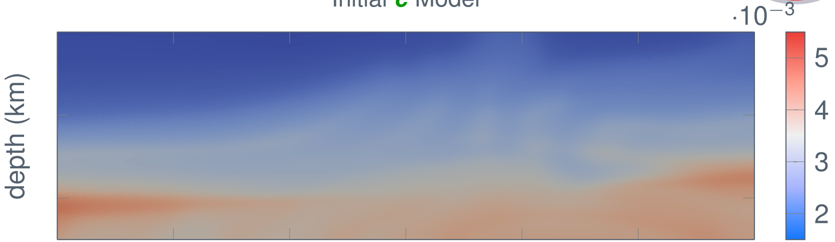
$\boldsymbol{c}$, $\boldsymbol{\rho}$ and $\boldsymbol{\kappa}$ Constant per elements

[4] http://dip.inria.fr/

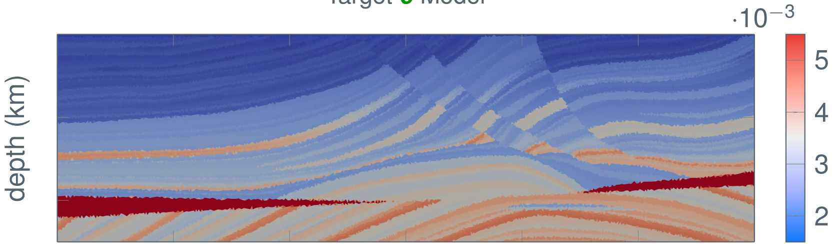
2D Time Domain FWI Reconstructions

Time-schemes comparison

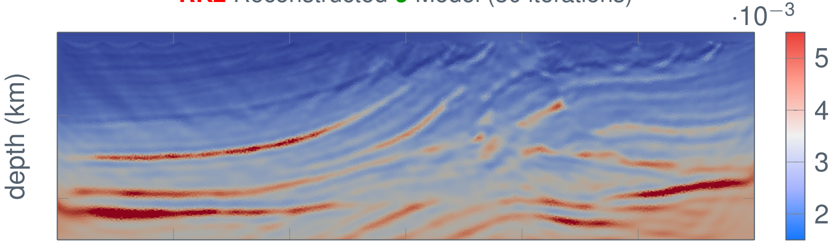
Initial *c* Model



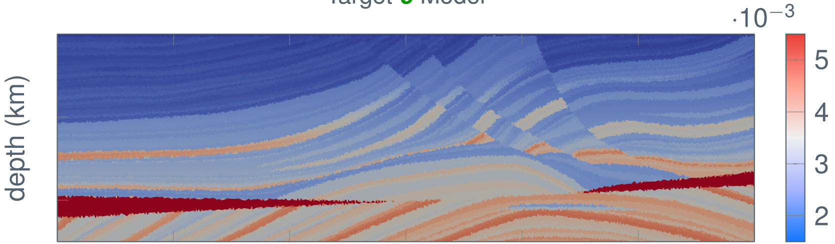
Target *c* Model



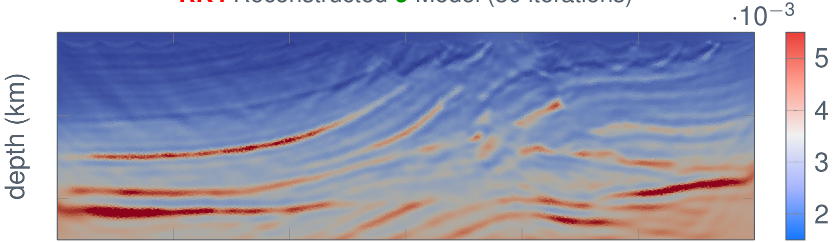
RK2 Reconstructed *c* Model (30 iterations)



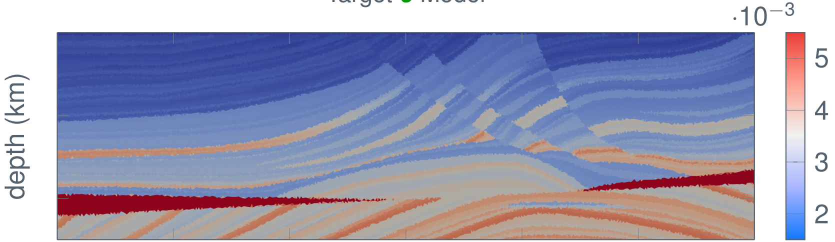
Target *c* Model



RK4 Reconstructed ***c*** Model (30 iterations)

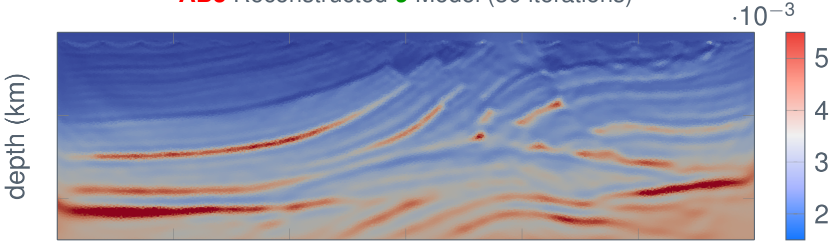


Target ***c*** Model

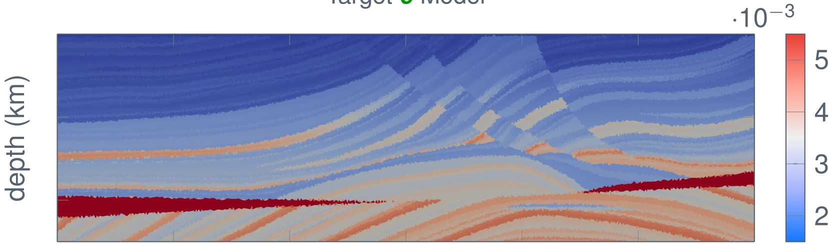


2D Time Domain FWI Reconstructions
Time-schemes comparison

AB3 Reconstructed ***c*** Model (30 iterations)



Target ***c*** Model

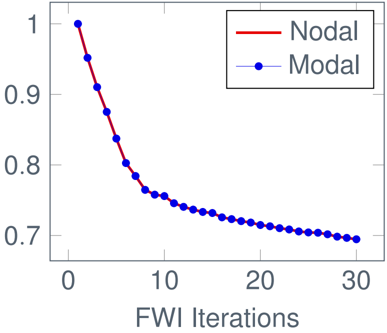


2D Time Domain FWI Reconstructions

Nodal/Modal Comparison

- 47k P1 elements
- Time Scheme : AB3
- Constant ρ model ($\rho = 1$)
- 19 sources / 181 Receivers
- 30 iterations
- 120 cores
- Nodal computation time : 5h10
- Modal computation time : 7h10[1]

Cost function evolution :

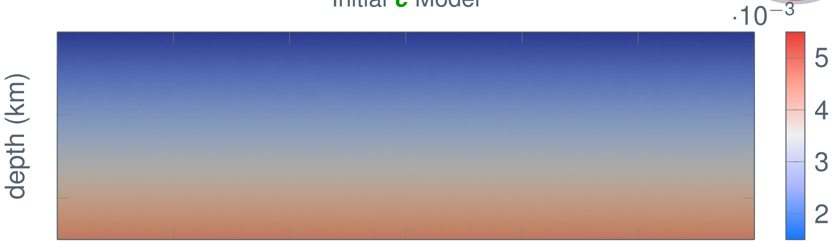


[1] Chan J. and Warburton T.
GPU-Accelerated Bernstein Bézier Discontinuous Galerkin Methods for Wave Problems
SIAM Journal on Scientific Computing 2017

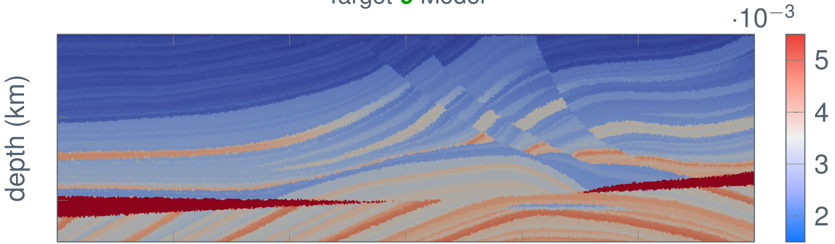
2D Multiscale Reconstructions

Reconstruction with an initial smooth model

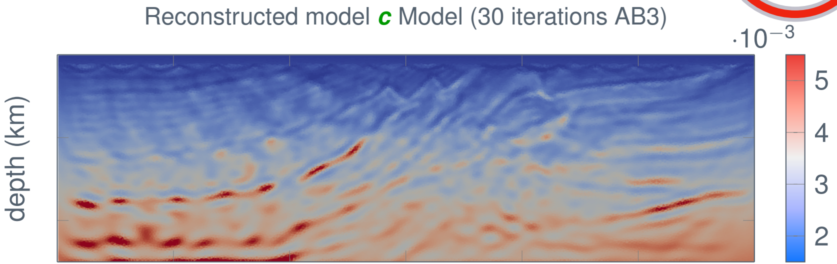
Initial c Model

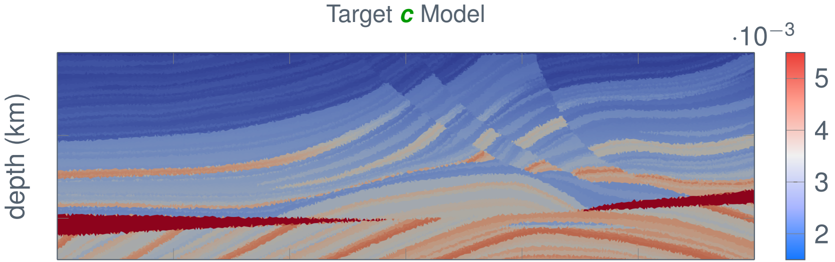


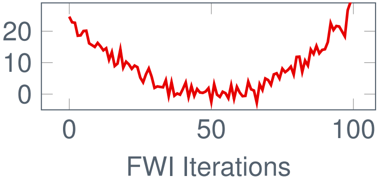
Target c Model



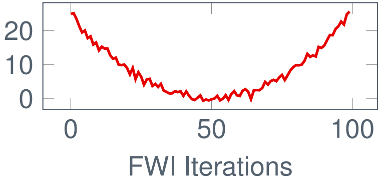
2D Multiscale Reconstructions
Reconstruction with an initial smooth model



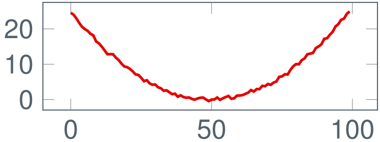




20
10
0
0
50
100
FWI Iterations

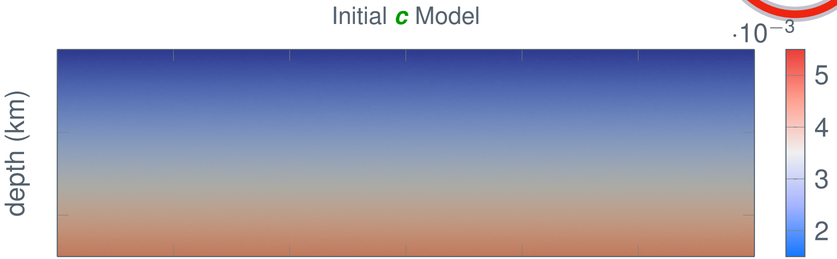


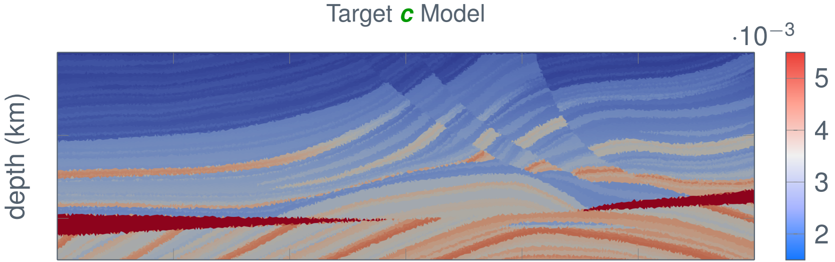
20
10
0
0
50
100
FWI Iterations



20
10
0
0
50
100

2D Multiscale Reconstructions
Reconstruction with an initial smooth model

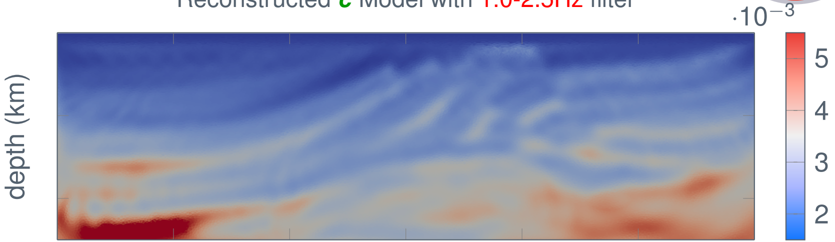




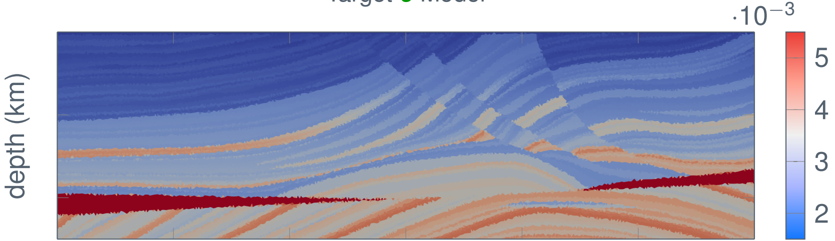
2D Multiscale Reconstructions

Reconstruction with an initial smooth model

Reconstructed *c* Model with 1.0-2.5Hz filter



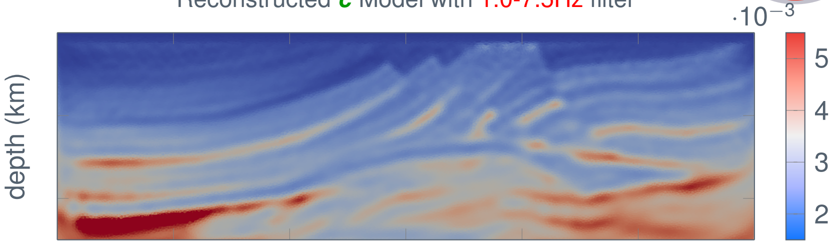
Target *c* Model



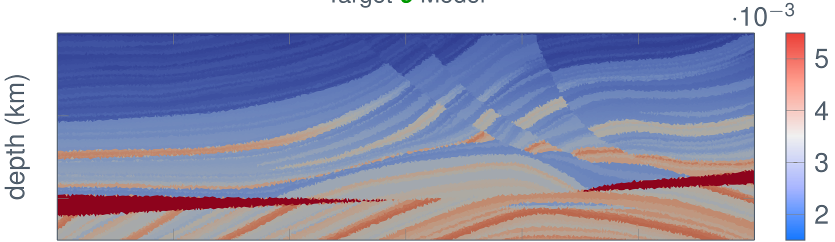
2D Multiscale Reconstructions

Reconstruction with an initial smooth model

Reconstructed ***c*** Model with 1.0-7.5Hz filter



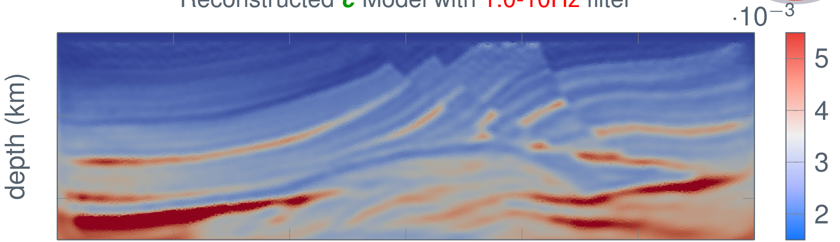
Target ***c*** Model



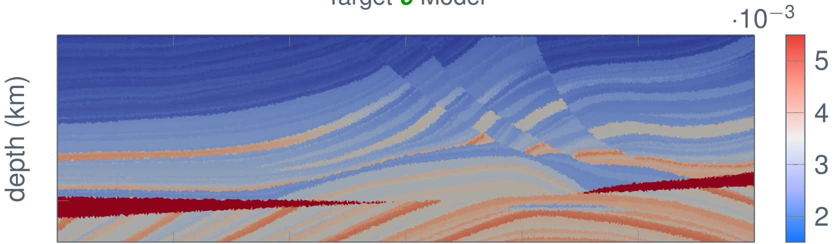
2D Multiscale Reconstructions

Reconstruction with an initial smooth model

Reconstructed *c* Model with 1.0-10Hz filter



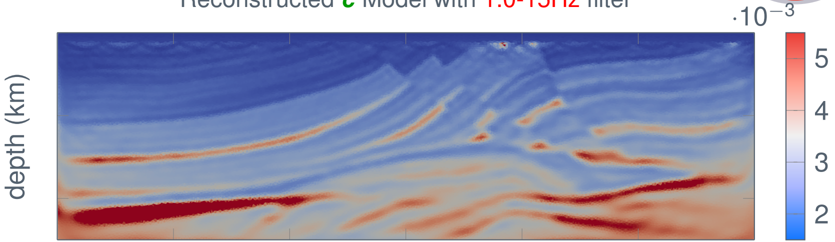
Target *c* Model



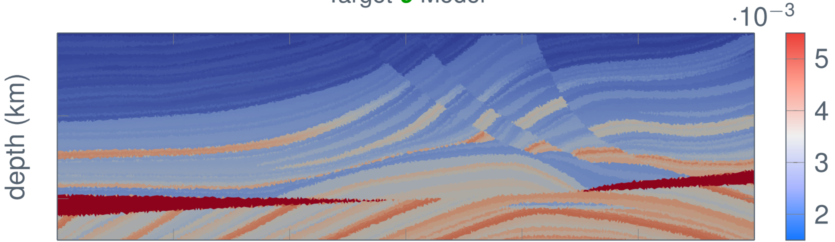
2D Multiscale Reconstructions

Reconstruction with an initial smooth model

Reconstructed *c* Model with 1.0-15Hz filter



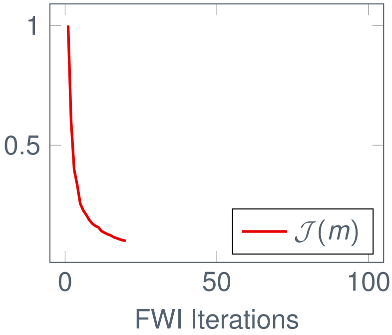
Target *c* Model



2D Multiscale Reconstructions

- 47k P1 elements
- Time Scheme : AB3
- Constant ρ model ($\rho = 1$)
- 19 sources / 181 Receivers
- 120 cores
- Computation time : 17h
- Frequencies : 1-2.5Hz

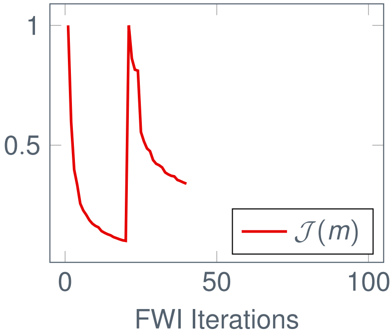
Cost function evolution :



2D Multiscale Reconstructions

- 47k P1 elements
- Time Scheme : AB3
- Constant ρ model ($\rho = 1$)
- 19 sources / 181 Receivers
- 120 cores
- Computation time : 17h
- Frequencies : 1-2.5Hz / 1-5Hz

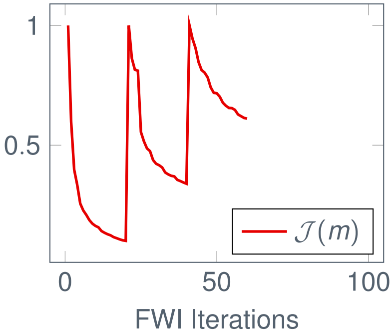
Cost function evolution :



2D Multiscale Reconstructions

- 47k P1 elements
- Time Scheme : AB3
- Constant ρ model ($\rho = 1$)
- 19 sources / 181 Receivers
- 120 cores
- Computation time : 17h
- Frequencies : 1-2.5Hz / 1-5Hz / 1-7.5Hz

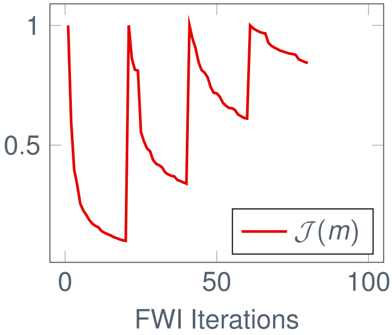
Cost function evolution :



2D Multiscale Reconstructions

- 47k P1 elements
- Time Scheme : AB3
- Constant ρ model ($\rho = 1$)
- 19 sources / 181 Receivers
- 120 cores
- Computation time : 17h
- Frequencies : 1-2.5Hz / 1-5Hz / 1-7.5Hz / 1-10Hz

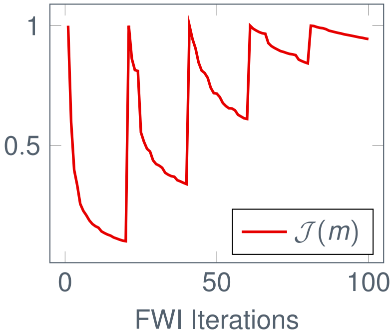
Cost function evolution :



2D Multiscale Reconstructions

- 47k P1 elements
- Time Scheme : AB3
- Constant ρ model ($\rho = 1$)
- 19 sources / 181 Receivers
- 120 cores
- Computation time : 17h
- Frequencies : 1-2.5Hz / 1-5Hz / 1-7.5Hz / 1-10Hz / 1-15Hz

Cost function evolution :



Conclusion

conclusion