Full Waveform Inversion Using the Wasserstein Metric

Brittany Froese Hamfeldt[1]
Björn Engquist[2]
Yunan Yang[3]

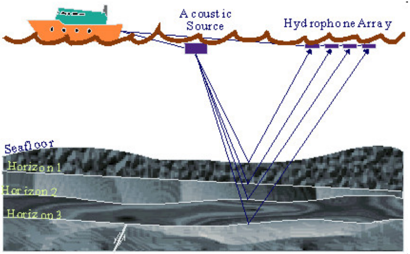
Optimization and Maching Learning Seminar
New Jersey Institute of Technology
September 30, 2021

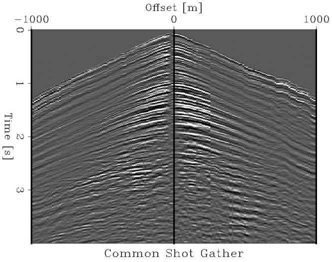
[1] New Jersey Institute of Technology
[2] University of Texas at Austin
[3] Cornell University

Outline

Outline

Seismic Full Waveform Inversion

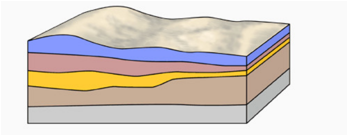




Forward Problem

Given velocity field $v(x)$:

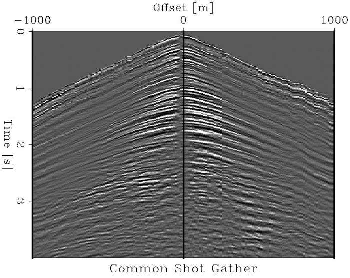
- Supply initial waved field $u_0(x, z)$ (eg Ricker wavelet)
- Wave equation $\Rightarrow u(x, z, t)$
- Obtain data $g = u(x, 0, t)$ from surface measurement



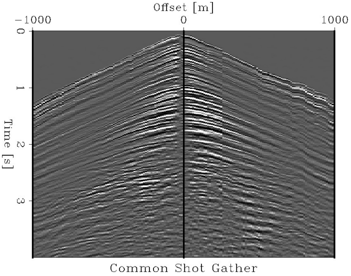
Velocity profile $v(x)$

wave equation

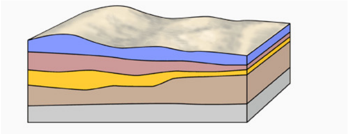
$\Rightarrow$



Inverse Problem



? $\Rightarrow$



Velocity profile $v(x)$

Estimate velocity field:

$$v^*(x) = \operatorname*{argmin}_v d(f(v), g)$$

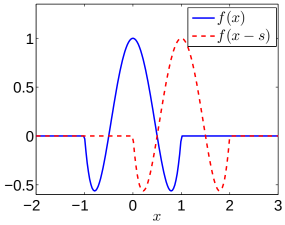
- g is observed data
- $f(v)$ is obtained from solving forward problem

Signal Skipping

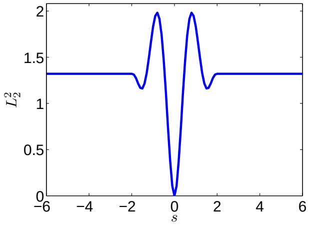
Example:

$$d(f, g) = \|f - g\|_2^2$$

Wavelet Profile

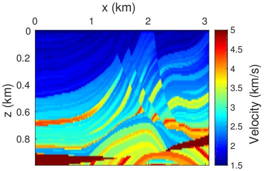


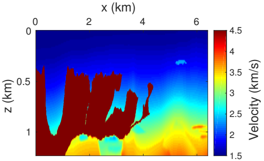
Misfit



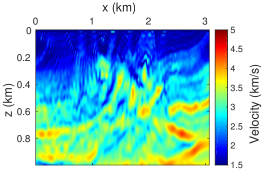
FWI with L_2 Misfit

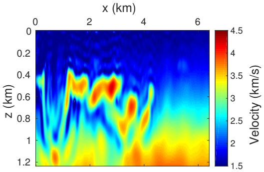
True Model





L^2 Inversion

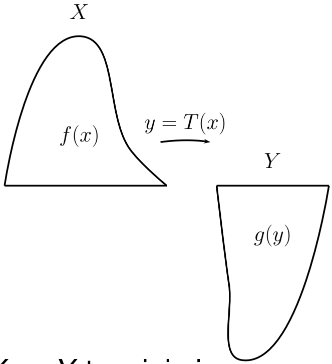




Outline

Optimal Transportation

Gaspard Monge



Find a mass-preserving map $T : X \to Y$ to minimise

$$\int_X |x - T(x)| \, f(x) \, dx.$$

Quadratic Cost

Find a mass-preserving map $T : X \to Y$ to minimise

$$\int_X |x - T(x)|^2 f(x)\, dx.$$

André-Marie Ampère

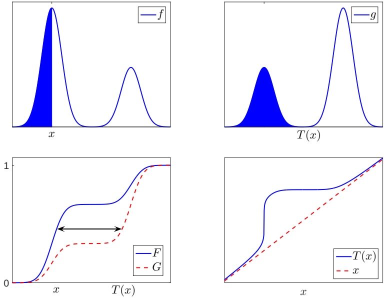
Defines *Wasserstein metric*

$$W_2(f, g) = \sqrt{\inf_{T \in \mathcal{M}} \int_X |x - T(x)|^2 f(x)\, dx}$$

where $\mathcal{M}$ is set of maps that rearrange f into g.

Optimal Map (1D)

- Use cumulative distributions $F(x) = \int_{-\infty}^{x} f(t)\,dt$
- Optimal map $T(x) = G^{-1}(F(x))$



$$W_2(f,g)^2 = \int_{\mathbb{R}} \left| x - G^{-1}(F(x)) \right|^2 f(x)\,dx$$

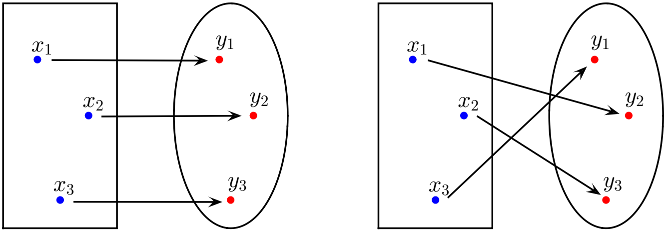
$$= \int_0^1 \left| F^{-1}(t) - G^{-1}(t) \right|^2 dt$$

Cyclical Monotonicity

For $c(x, y) = |x - y|^2$, the optimal map is *cyclically monotone*.

For any $m \in \mathbb{N}^+$, $x_i \in X$, $1 \leq i \leq m$, $x_0 \equiv x_m$,

$$\sum_{i=1}^{m} x_i \cdot T(x_i) \geq \sum_{i=1}^{m} x_i \cdot T(x_{i-1}).$$



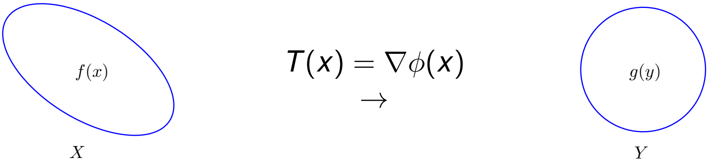
$T(x) = \nabla\phi(x)$ where ϕ is convex [Rockafellar, *Pac. J. Math.*, 1966]

The Monge-Ampère Equation

- Conservation of mass $\Rightarrow g(T(x))\det(\nabla T(x)) = f(x)$
- Quadratic cost $\Rightarrow T(x) = \nabla\phi(x)$

Obtain *Monge-Ampère equation*

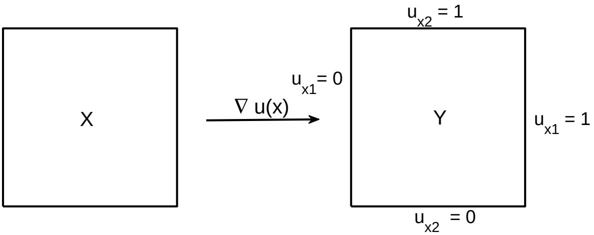
$$\begin{cases} \det(D^2\phi(x)) = f(x)/g(\nabla\phi(x)) \\ \phi \text{ is convex} \end{cases}$$



Boundary Conditions

- Rectangle to rectangle conditions can be expressed as Neumann condition

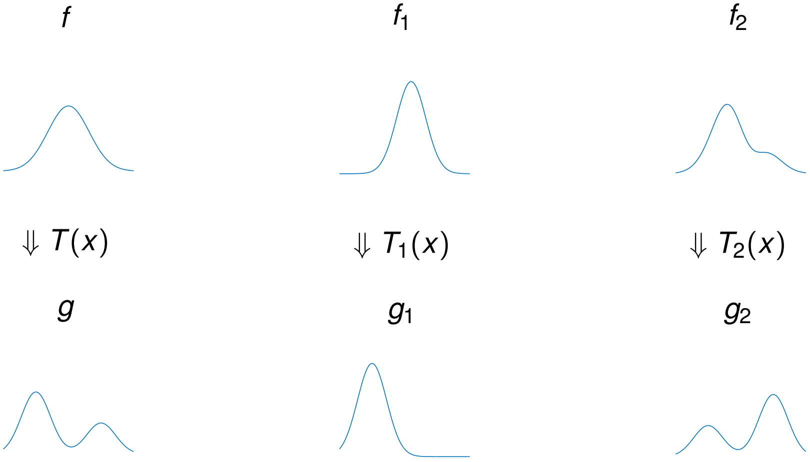
$$\nabla \phi \cdot n = x \cdot n, \quad x \in \partial X.$$



- Other convex geometries lead to a Hamilton-Jacobi boundary condition

$$H(\nabla \phi(x)) = 0, \quad x \in \partial X.$$

Convexity of W_2^2



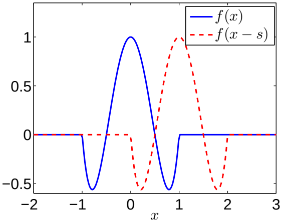
$$W_2^2(\alpha f_1 + (1-\alpha) f_2, \alpha g_1 + (1-\alpha) g_2) \leq \alpha W_2^2(f_1, g_1) + (1-\alpha) W_2^2(f_2, g_2)$$

W_2 as Misfit

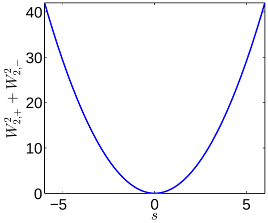
Propose:

$$d(f,g) = W_2(f/\langle f \rangle, g)^2$$

Wavelet Profile



Misfit



Remark: some pre-processing needed to turn f, g into densities

Forms of Error

$u(x,t;v) = u_0(t - x/v)$ solves 1D wave equation

$$\begin{cases} u_{tt} = v^2 u_{xx}, & x > 0, t > 0 \\ u = u_t = 0, & x > 0, t = 0 \\ u = u_0(t), & x = 0, t > 0 \end{cases}$$

Variations in v lead to

- Translations
- Dilations in x

Variations in v at discontinuities lead to

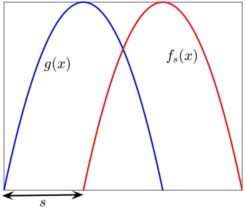
- Local changes in amplitude of reflected signal

Translations

Let $f_s(x) = g(x - s\eta)$, $\eta \in \mathbb{R}^n$

- Optimal map:

$$T_s(x) = x - s\eta = \nabla \left(\frac{|x|^2}{2} - s\eta \cdot x \right)$$



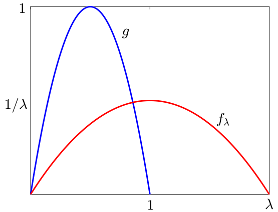
- Wasserstein metric:

$$W_2^2(f_s, g) = \int |x - T_s(x)|^2 f_s(x)\, dx$$
$$= s^2 |\eta|^2 \int f_s(x)\, dx$$

Dilations

Let $f_\Lambda(x) = g(\Lambda^{-1}x)/\det(\Lambda)$,

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$



- Optimal map:

$$T_s(x) = \Lambda^{-1}x = \nabla\left(\frac{1}{2}x^T\Lambda^{-1}x\right)$$

- Wasserstein metric:

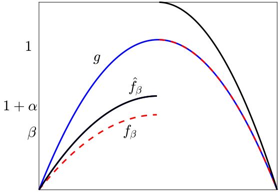
$$W_2^2(f_\Lambda, g) = \int |x - T_\Lambda(x)|^2 f_\Lambda(x)\,dx$$
$$= \int y^T(I-\Lambda)^2 y\,g(y)\,dy$$

Local Amplitude Change

Let $f_\beta(x) = \begin{cases} \beta g(x), & x \in E \\ g(x) & x \notin E \end{cases}$

- Reparameterise and normalise as

$$h_\alpha(x) = \begin{cases} (1+\alpha)g(x), & x \in E \\ (1-\gamma_\alpha)g(x), & x \notin E \end{cases}$$



- Exploit convexity of W_2^2

$$W_2^2(sh_{\alpha_1} + (1-s)h_{\alpha_2}, g) \leq sW_2^2(h_{\alpha_1}, g) + (1-s)W_2^2(h_{\alpha_2}, g)$$

- Obtain

$$W_2^2(\hat{f}_{s\beta_1+(1-s)\beta_2}, g) \leq sW_2^2(\hat{f}_{\beta_1}, g) + (1-s)W_2^2(\hat{f}_{\beta_2}, g)$$

Convexity

$$f(x;s) = g(x + s\eta), \quad \eta \in \mathbb{R}^n, \tag{1}$$

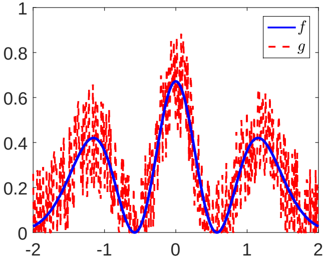
$$f(x;A) = g(Ax), \quad A^T = A,\ A > 0, \tag{2}$$

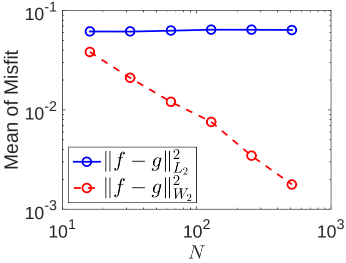
$$f(x;\beta) = \begin{cases} \beta g(x), & x \in E \\ g(x), & x \in \mathbb{R}^n \setminus E. \end{cases} \tag{3}$$

Theorem (Enguist, F, and Yang, *Comm. Math. Sci.*, 2016)

The squared Wasserstein metric $W_2^2(f(m), g)$ is convex with respect to the model parameters m corresponding to a shift s in (1), the eigenvalues of the dilation matrix A in (2), or the local rescaling parameter β in (3).

Noise

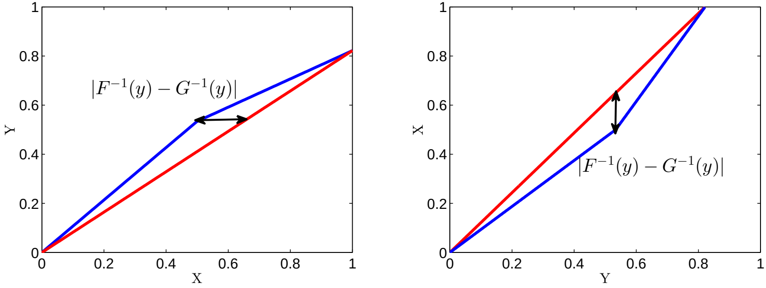

1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
f
g


Mean of Misfit
10^{-1}
10^{-2}
10^{-3}
10^{1}
10^{2}
10^{3}
N
$\|f-g\|_{L_2}^2$
$\|f-g\|_{W_2}^2$

Noise (1D)

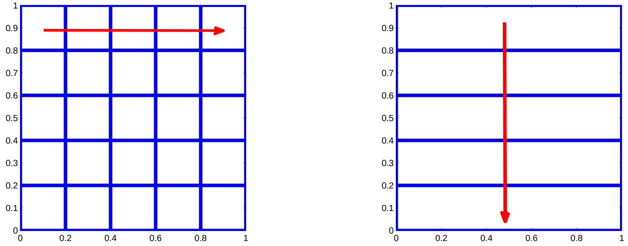
Theorem (Enguist, F, and Yang, *Comm. Math. Sci.*, 2016)

Let g be a positive probability density function on $[0,1]$ *and choose* $0 < c < \min g$. *Let* $f_N(x) = g(x) + r^N(x)$, *which contains piecewise constant additive noise* r^N *drawn from the uniform distribution* $U[-c,c]$. *Then* $\mathbb{E} W_2^2(f_N/\langle f_N\rangle, g) = \mathcal{O}(\frac{1}{N})$.



Noise (2D)

- Compute optimal maps dimension by dimension
- Obtain non-optimal composition $T_{x_1} \circ T_{x_2}$



$$\mathbb{E} W_2^2(f_N, g) \leq \mathbb{E} \int_X f_N(x) \left| x - T_{x_1} \circ T_{x_2}(x) \right|^2 dx = \mathcal{O}\left(\frac{1}{N}\right)$$

Outline

The Optimisation Problem

Goal:

$$v^* = \underset{v}{\operatorname{argmin}}\, J(v)$$

where

$$J(v) \equiv W_2(f(v), g)^2$$

Challenge:

- Need
  $$\nabla J(v) = \nabla_f W_2^2(f(v), g) \nabla_v f(v)$$
  for efficient optimisation.
- $J(v)$ depends on intermediate variables u, ϕ
- Parameter space (v) may be high-dimensional.

Example: Least Squares

Eg: $g, \phi, v \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$

$$\begin{aligned} \text{Minimise} \quad & J(v) \equiv \frac{1}{2}\|\phi - g\|_2^2 \\ \text{s. t.} \quad & A\phi = v \end{aligned}$$

- Introduce perturbations

$$\delta v = |\delta v|\, e_j, \quad j = 1, \ldots, n$$

- Intermediate perturbation

$$\delta\phi = A^{-1}\delta v$$

- Obtain final perturbation δJ

Example: Least Squares

- Compute perturbation:

$$
\begin{aligned}
J + \delta J &= \frac{1}{2}\|\phi + \delta\phi - g\|^2 \\
&= J + \langle \phi - g, \delta\phi \rangle + h.o.t. \\
&= J + \langle \phi - g, A^{-1}\delta v \rangle + h.o.t.
\end{aligned}
$$

- Problem: Computing all $A^{-1}\delta v$ requires n linear solves.
- Solution: Introduce adjoint state equation to obtain

$$\delta J = \langle (A^{-1})^*(\phi - g), \delta v \rangle$$

Example: Least Squares

- Compute perturbation:

$$\begin{aligned} J + \delta J &= \frac{1}{2}\|\phi + \delta\phi - g\|^2 \\ &= J + \langle \phi - g, \delta\phi \rangle + h.o.t. \\ &= J + \langle \phi - g, A^{-1}\delta v \rangle + h.o.t. \end{aligned}$$

- Problem: Computing all $A^{-1}\delta v$ requires n linear solves.
- Solution: Introduce adjoint state equation to obtain

$$\delta J = \langle (A^{-1})^*(\phi - g), \delta v \rangle$$

Gradient Approximations

$$W_2^2 \xrightarrow{\text{Linearise}} \nabla W_2^2 \xrightarrow{\text{Discretise}} (\nabla W_2^2)^h$$

$$W_2^2 \xrightarrow{\text{Discretise}} W_2^{2,h} \xrightarrow{\text{Linearise}} \nabla (W_2^{2,h})$$

Linearisation of W_2^2

- Introduce functional

$$J(f) = \int_X f(x)\,|x - \nabla u_f|^2 \,dx.$$

- Perturb f,

$$J + \delta J = \int_X (f + \delta f)\,|x - \nabla(u_f + \delta u)|^2 \,dx.$$

- To first order,

$$\delta J = \int_X \Big(|x - \nabla u_f|^2\,\delta f - 2f(x - \nabla u_f)\cdot\nabla(\delta u)\Big)\,dx.$$

Linearisation of Monge-Ampère

- Let $\nabla(u_f + \delta u)$ be optimal map from $f + \delta f$ to g.
- Perturbed Monge-Ampère equation is

$$f + \delta f = g(\nabla(u_f + \delta u)) \det(D^2(u_f + \delta u)).$$

- Linearise:

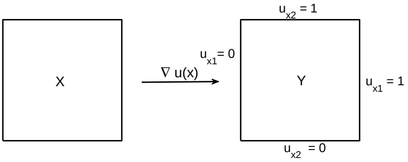
$$\begin{aligned}\mathcal{L}[\delta u] \equiv\ & g(\nabla u_f)\mathrm{tr}((D^2 u_f)_{adj} D^2(\delta u)) \\ & + \det(D^2 u_f)\nabla g(\nabla u_f) \cdot \nabla(\delta u) = \delta f.\end{aligned}$$

- A linear elliptic PDE for δu.

Boundary Conditions

- Rectangle to rectangle case,

$$\nabla u_f \cdot n = x \cdot n, \quad \nabla(u_f + \delta u) \cdot n = x \cdot n, \quad x \in \partial X.$$



- Linearised problem requires homogeneous Neumann conditions

$$\nabla(\delta u) \cdot n = 0, \quad x \in \partial X.$$

Linearisation of W_2^2

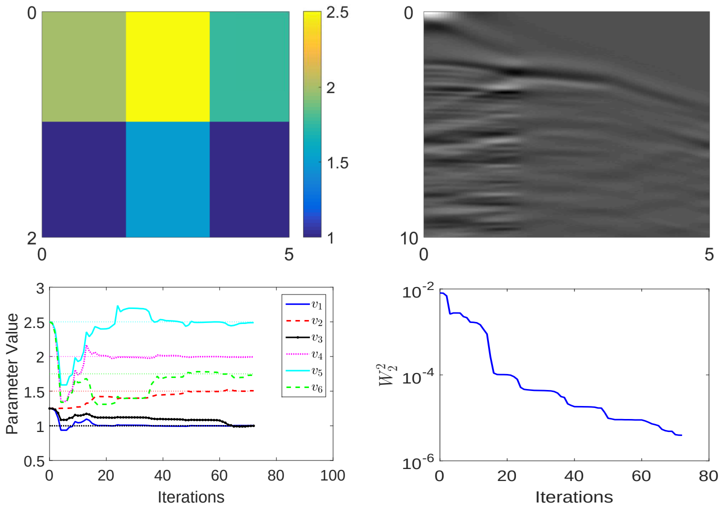
- The first variation is

$$\begin{aligned}\delta J &= \int_X \left[|x - \nabla u_f|^2 - 2f(x - \nabla u_f) \cdot \nabla \mathcal{L}^{-1} \right] \delta f \\ &= \int_X \left[|x - \nabla u_f|^2 + 2(\mathcal{L}^{-1})^* (\nabla \cdot (f(x - \nabla u_f))) \right] \delta f \\ &= \langle |x - \nabla u_f|^2 + 2(\mathcal{L}^{-1})^* (\nabla \cdot (f(x - \nabla u_f))), \delta f \rangle\end{aligned}$$

- Requires solution of a single linear elliptic PDE!

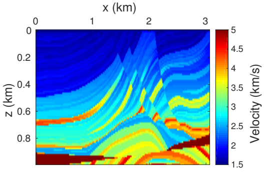
Outline

6 Parameter Model

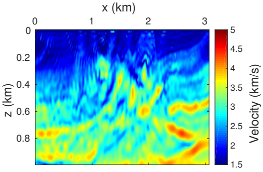


Marmousi Model

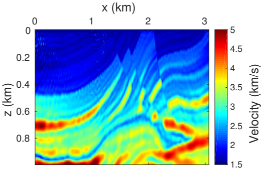
True Model

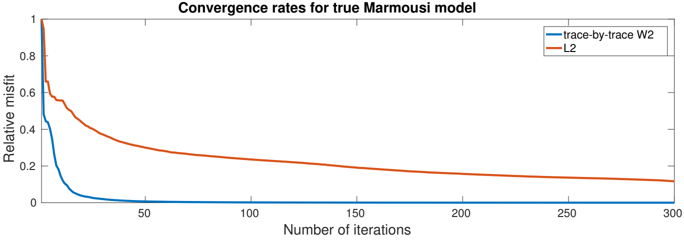


L^2 Inversion



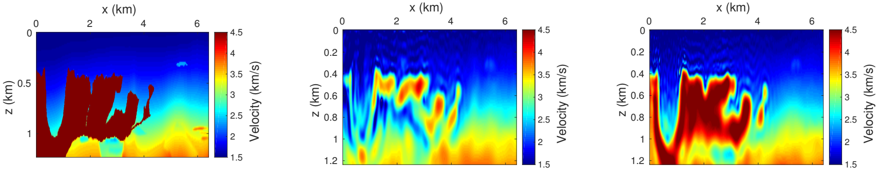
W_2 Inversion

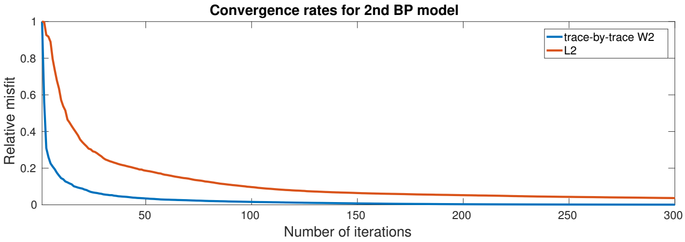




BP Model

True Model | L^2 Inversion | W_2 Inversion





Thanks!