Velocity Inversion Using the Quadratic Wasserstein Metric

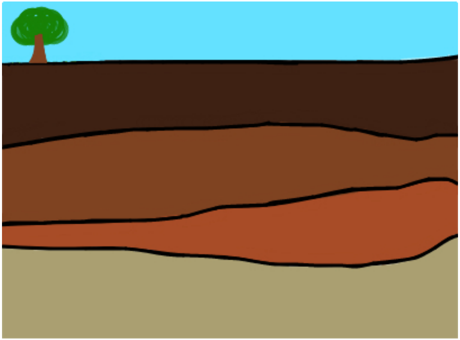
Srinath Mahankali
Mentor: Prof. Yunan Yang (NYU)

Stuyvesant High School

October 17, 2020
MIT PRIMES Conference

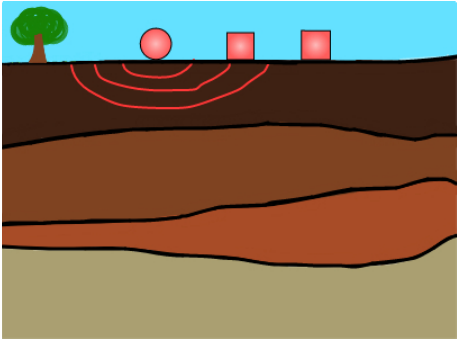
Motivation

- Goal: find materials lying underground and their positions



Motivation

- Goal: find materials lying underground and their positions



Approach

- $$\begin{cases} \frac{1}{C(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{\partial^2\psi}{\partial {x_1}^2} + \frac{\partial^2\psi}{\partial {x_2}^2} + \cdots + \frac{\partial^2\psi}{\partial {x_n}^2}\right) = f(\mathbf{x}, t), \\ \psi(\mathbf{x}, 0) = 0,\ \psi_t(\mathbf{x}, 0) = 0 \text{ on } \Omega, \\ \nabla\psi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \end{cases}$$

Approach

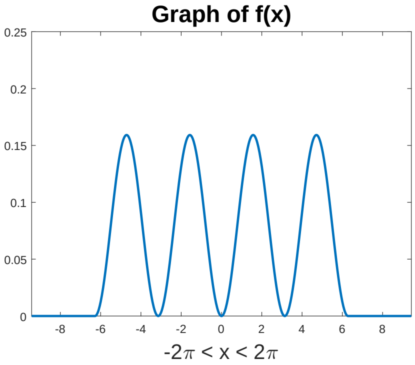
- $$\begin{cases} \frac{1}{\mathcal{C}(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{\partial^2\psi}{\partial {x_1}^2} + \frac{\partial^2\psi}{\partial {x_2}^2} + \cdots + \frac{\partial^2\psi}{\partial {x_n}^2}\right) = f(\mathbf{x}, t), \\ \psi(\mathbf{x}, 0) = 0,\ \psi_t(\mathbf{x}, 0) = 0 \text{ on } \Omega, \\ \nabla\psi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \end{cases}$$
- $\mathcal{C}^*(\mathbf{x}) = \operatorname{argmin} J(g(\mathcal{C}), h)$ where J is the objective function

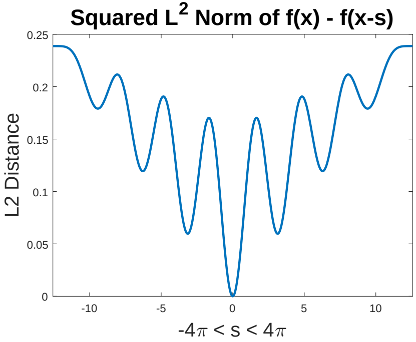
Approach

- $$\begin{cases} \frac{1}{\mathcal{C}(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{\partial^2\psi}{\partial {x_1}^2} + \frac{\partial^2\psi}{\partial {x_2}^2} + \cdots + \frac{\partial^2\psi}{\partial {x_n}^2}\right) = f(\mathbf{x}, t), \\ \psi(\mathbf{x}, 0) = 0,\ \psi_t(\mathbf{x}, 0) = 0 \text{ on } \Omega, \\ \nabla\psi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \end{cases}$$
- $\mathcal{C}^*(\mathbf{x}) = \mathrm{argmin}\ J(g(\mathcal{C}), h)$ where J is the objective function
- Convexity with respect to the velocity is beneficial

Problems With the Squared L^2 Norm

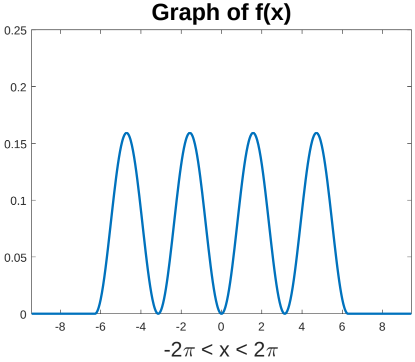
- Beydoun-Tarantola, 1988
  - Nonconvexity of squared L^2 norm and local minima
- Brossier et al, 2010
  - Sensitivity of L^2 norm to noise

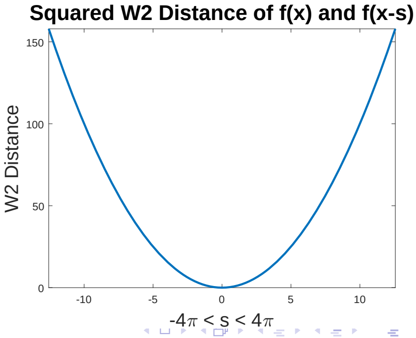




An Alternative Objective Function

- Engquist-Froese, 2013 introduced the Wasserstein metric for FWI
- Yang, 2019
  - Convexity in translations and dilations of the data
  - Insensitivity to noise
- Engquist et al, 2020
  - Low-frequency bias





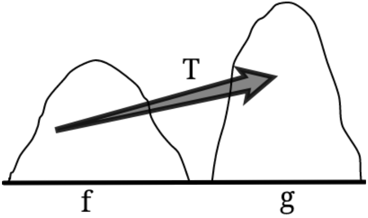
What is the Wasserstein Metric?

- Optimal transport introduced by Monge

Definition

The p^{th} Wasserstein metric is defined as

$$W_p(f, g) = \left(\inf_{T \in \mathbb{M}(\mu,\nu)} \int_\Omega |x - T(x)|^p \mathrm{d}\mu \right)^{\frac{1}{p}}.$$



What is the Wasserstein Metric?

- Explicit formula when data is in one dimension

Theorem

Let $\widetilde{g}$ and $\widetilde{h}$ be two probability distributions defined on $\mathbb{R}$, and let $G(t) = \int_{-\infty}^{t} \widetilde{g}(s)\,\mathrm{d}s$ and $H(t) = \int_{-\infty}^{t} \widetilde{h}(s)\,\mathrm{d}s$. Then

$$W_2^2(\widetilde{g}, \widetilde{h}) = \int_0^1 (G^{-1}(s) - H^{-1}(s))^2\,\mathrm{d}s.$$

The Constrained Optimization Problem

- $$\begin{cases} \frac{1}{C(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{\partial^2\psi}{\partial {x_1}^2} + \frac{\partial^2\psi}{\partial {x_2}^2} + \cdots + \frac{\partial^2\psi}{\partial {x_n}^2}\right) = f(\mathbf{x}, t), \\ \psi(\mathbf{x}, 0) = 0,\ \psi_t(\mathbf{x}, 0) = 0 \text{ on } \Omega, \\ \nabla\psi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \end{cases}$$

The Constrained Optimization Problem

- $$\begin{cases} \frac{1}{\mathcal{C}(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{\partial^2\psi}{\partial {x_1}^2} + \frac{\partial^2\psi}{\partial {x_2}^2} + \cdots + \frac{\partial^2\psi}{\partial {x_n}^2}\right) = f(\mathbf{x}, t), \\ \psi(\mathbf{x}, 0) = 0,\ \psi_t(\mathbf{x}, 0) = 0 \text{ on } \Omega, \\ \nabla\psi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \end{cases}$$
- $\mathcal{C}^*(\mathbf{x}) = \operatorname{argmin} W_2^2(g(\mathcal{C}), h)$

The Constrained Optimization Problem

- $$\begin{cases} \frac{1}{\mathcal{C}(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{\partial^2\psi}{\partial {x_1}^2} + \frac{\partial^2\psi}{\partial {x_2}^2} + \cdots + \frac{\partial^2\psi}{\partial {x_n}^2}\right) = f(\mathbf{x}, t), \\ \psi(\mathbf{x}, 0) = 0,\ \psi_t(\mathbf{x}, 0) = 0 \text{ on } \Omega, \\ \nabla\psi \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \end{cases}$$
- $\mathcal{C}^*(\mathbf{x}) = \text{argmin } W_2^2(g(\mathcal{C}), h)$
- Previous results involve convexity in changes of the data

Challenges and Solutions

- Wave data is generally not a probability distribution

Challenges and Solutions

- Wave data is generally not a probability distribution
  - Normalize wave data $k(t)$ by $\widetilde{k}(t) = \frac{k(t)+\gamma}{\int_0^T k(t)+\gamma \, dt}$

Challenges and Solutions

- Wave data is generally not a probability distribution
  - Normalize wave data $k(t)$ by $\widetilde{k}(t) = \frac{k(t)+\gamma}{\int_0^T k(t)+\gamma \, \mathrm{d}t}$
- Computing the W_2 distance is difficult for higher dimensions

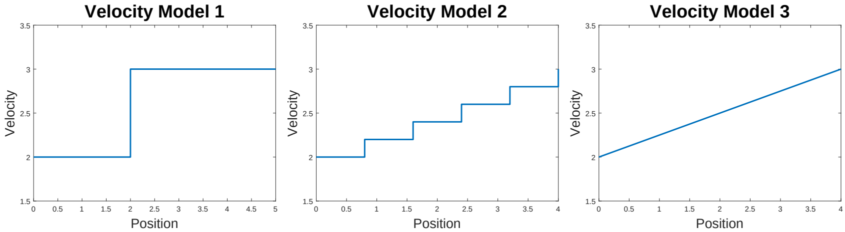
Challenges and Solutions

- Wave data is generally not a probability distribution
  - Normalize wave data $k(t)$ by $\widetilde{k}(t) = \frac{k(t)+\gamma}{\int_0^T k(t)+\gamma \ dt}$
- Computing the W_2 distance is difficult for higher dimensions
  - Data is only a function of time $\rightarrow$ apply one dimensional formula

Challenges and Solutions

- Wave data is generally not a probability distribution
  - Normalize wave data $k(t)$ by $\widetilde{k}(t) = \frac{k(t)+\gamma}{\int_0^T k(t)+\gamma \, \mathrm{d}t}$
- Computing the W_2 distance is difficult for higher dimensions
  - Data is only a function of time $\rightarrow$ apply one dimensional formula
- No explicit solution for wave equation in general

Velocity Models in One Dimension



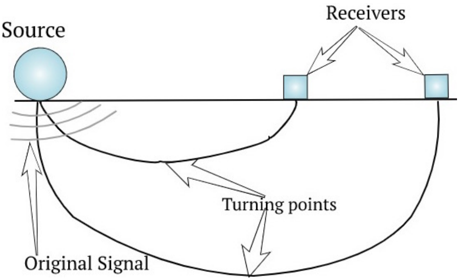
Theorem

Suppose $f(t)$ is nonnegative and compactly supported, and let k be the velocity parameter in these three cases. Then $W_2^2(g(k), h)$ is convex in k over the interval $(0, k^*]$.

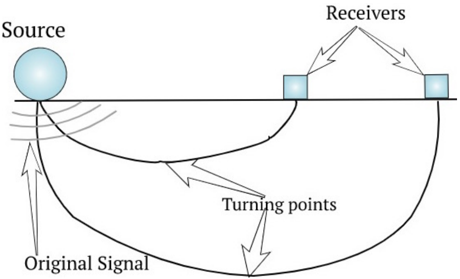
Ray Tracing

- Velocity function of the form $\mathcal{C}(X, z) = a + bz$

Ray Tracing



Ray Tracing

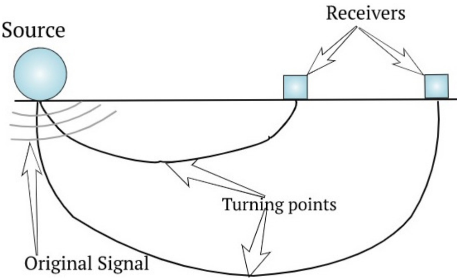


- Distance and traveltime formulas:

$$X = 2p \int_0^{z_p} \frac{\mathrm{d}z}{\sqrt{u^2(z) - p^2}}$$

$$T = 2 \int_0^{z_p} \frac{u^2(z)}{\sqrt{u^2(z) - p^2}} \, \mathrm{d}z$$

Ray Tracing



- Distance and traveltime formulas:

$$X = 2p \int_0^{z_p} \frac{\mathrm{d}z}{\sqrt{u^2(z) - p^2}}$$

$$T = 2 \int_0^{z_p} \frac{u^2(z)}{\sqrt{u^2(z) - p^2}} \, \mathrm{d}z$$

- For source wave data $f(t)$, the predicted wave data is approximated by $A_{\text{pred}} f(t - T_{\text{pred}})$

Velocity Model in Two Dimensions

Theorem

Assume $f(t)$ is nonnegative and compactly supported. Then, $W_2^2(\widetilde{g}, \widetilde{h})$ is jointly convex in a, b over the following region U:

$$U := \{(a,b) \in \mathbb{R}^2 : a, b > 0, \quad \frac{bX_r}{2a} \geq S_0, \quad T(X_r, a, b) \geq T(X_r, a^*, b^*)\}$$

for some positive constant S_0.

Velocity Model in Two Dimensions

Theorem

Assume $f(t)$ is nonnegative and compactly supported. Then, $W_2^2(\widetilde{g}, \widetilde{h})$ is jointly convex in a, b over the following region U:

$$U := \{(a, b) \in \mathbb{R}^2 : a, b > 0, \quad \frac{bX_r}{2a} \geq S_0, \quad T(X_r, a, b) \geq T(X_r, a^*, b^*)\}$$

for some positive constant S_0.

- For large X_r, the convex region contains (a^*, b^*)

Velocity Model in Two Dimensions

Theorem

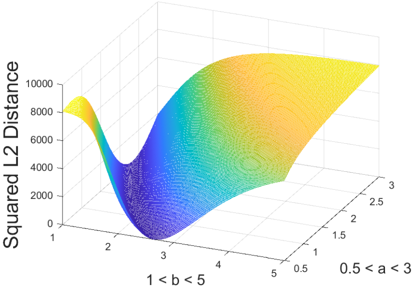
Assume $f(t)$ is nonnegative and compactly supported. Then, $W_2^2(\widetilde{g}, \widetilde{h})$ is jointly convex in a, b over the following region U:

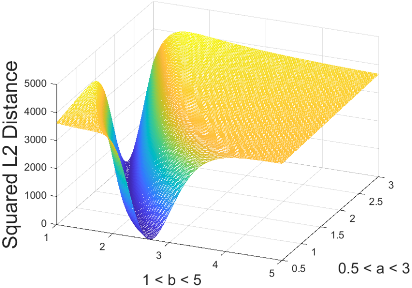
$$U := \{(a, b) \in \mathbb{R}^2 : a, b > 0, \quad \frac{bX_r}{2a} \geq S_0, \quad T(X_r, a, b) \geq T(X_r, a^*, b^*)\}$$

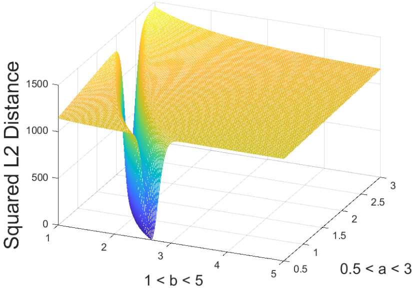
for some positive constant S_0.

- For large X_r, the convex region contains (a^*, b^*)
- Nonuniqueness of solution fixed by adding multiple receiver locations

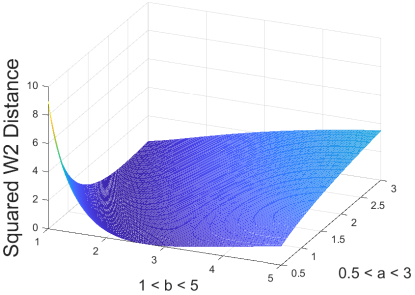
Velocity Model in Two Dimensions

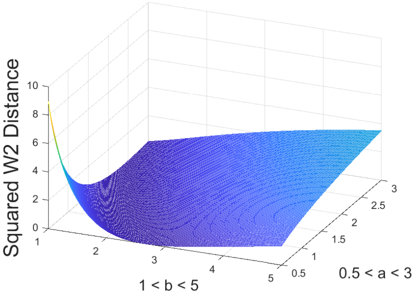


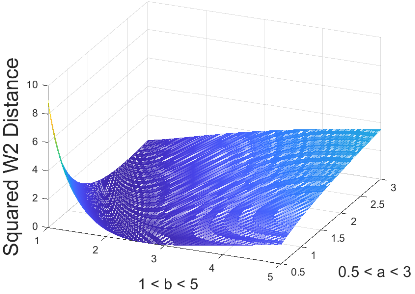




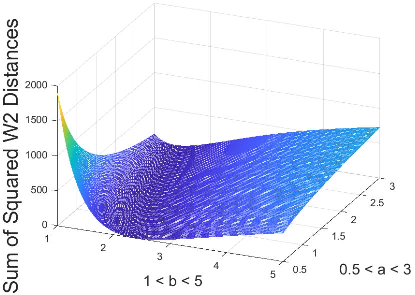
Velocity Model in Two Dimensions

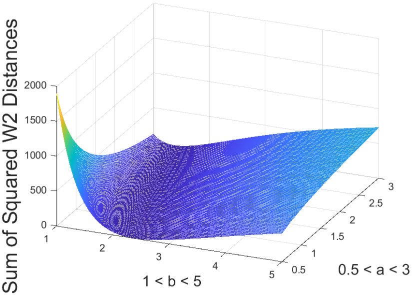


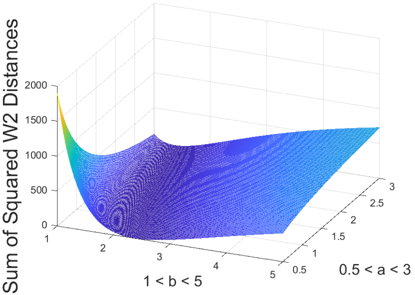




Velocity Model in Two Dimensions







Future Work

- Study models with a larger number of parameters

Future Work

- Study models with a larger number of parameters
- Possible to study convexity using frequency instead of time domain

Acknowledgements

I am extremely thankful to:

- My mentor, Prof. Yunan Yang
- The PRIMES program
- My parents and my brother

Acknowledgements

I am extremely thankful to:

- My mentor, Prof. Yunan Yang
- The PRIMES program
- My parents and my brother

Link to my paper (posted on PRIMES website):
https://arxiv.org/abs/2009.00708

References

- Wafik B. Beydoun and Albert Tarantola. First Born and Rytov approximations: Modeling and inversion conditions in a canonical example. The Journal of the Acoustical Society of America, 83(3):1045–1055, 1988.
- Romain Brossier, Stéphane Operto, and Jean Virieux. Which residual norm for robust elastic frequency-domain full waveform inversion. Geophysics, 75, 05 2010.
- Björn Engquist and Brittany D. Froese. Application of the Wasserstein metric to seismic signals. arXiv preprint arXiv:1311.4581, 2013.
- Björn Engquist, Kui Ren, and Yunan Yang. The quadratic Wasserstein metric for inverse data matching. Inverse Problems, 36(5):055001, May 2020.

References

- Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie Royale des Sciences de Paris, 1781.
- Yunan Yang. Analysis and Application of Optimal Transport For Challenging Seismic Inverse Problems, 2019.