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
Goal: find materials lying underground and their positions
Approach

\[ \frac{1}{C(x)^2} \frac{\partial^2 \psi}{\partial t^2} - \left( \frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \cdots + \frac{\partial^2 \psi}{\partial x_n^2} \right) = f(x, t), \]

\[
\begin{align*}
\psi(x, 0) &= 0, \quad \psi_t(x, 0) = 0 \text{ on } \Omega, \\
\nabla \psi \cdot \mathbf{n} &= 0 \text{ on } \partial \Omega
\end{align*}
\]
\[
\begin{align*}
\frac{1}{C(x)^2} \frac{\partial^2 \psi}{\partial t^2} - \left( \frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \cdots + \frac{\partial^2 \psi}{\partial x_n^2} \right) &= f(x, t), \\
\psi(x, 0) = 0, \quad \psi_t(x, 0) = 0 \text{ on } \Omega, \\
\nabla \psi \cdot \mathbf{n} &= 0 \text{ on } \partial \Omega \\
C^*(x) &= \arg\min J(g(C), h) \text{ where } J \text{ is the objective function}
\end{align*}
\]
Approach

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

- \( C^*(x) = \arg\min J(g(C), h) \) 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

**Graph of $f(x)$**

- $-2\pi < x < 2\pi$
- $0 < y < 0.25$

**Squared $L^2$ Norm of $f(x) - f(x-s)$**

- $-4\pi < s < 4\pi$
- $0 < L^2 \text{ Distance} < 0.25$
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

Graph of $f(x)$

Squared W2 Distance of $f(x)$ and $f(x-s)$
What is the Wasserstein Metric?

- Optimal transport introduced by Monge

**Definition**

The $p$th Wasserstein metric is defined as

$$W_p(f, g) = \left( \inf_{T \in \mathcal{M}(\mu, \nu)} \int_{\Omega} |x - T(x)|^p d\mu \right)^{\frac{1}{p}}.$$
What is the Wasserstein Metric?

- Explicit formula when data is in one dimension

**Theorem**

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

$$W_2^2(\tilde{g}, \tilde{h}) = \int_{0}^{1} (G^{-1}(s) - H^{-1}(s))^2 \, ds.$$
The Constrained Optimization Problem

\[
\begin{aligned}
\frac{1}{C(x)^2} \frac{\partial^2 \psi}{\partial t^2} - \left( \frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \cdots + \frac{\partial^2 \psi}{\partial x_n^2} \right) &= f(x, t), \\
\psi(x, 0) &= 0, \quad \psi_t(x, 0) = 0 \text{ on } \Omega, \\
\nabla \psi \cdot n &= 0 \text{ on } \partial \Omega
\end{aligned}
\]
The Constrained Optimization Problem

\[
\begin{aligned}
\frac{1}{C(x)^2} \frac{\partial^2 \psi}{\partial t^2} - \left( \frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \cdots + \frac{\partial^2 \psi}{\partial x_n^2} \right) &= f(x, t), \\
\psi(x, 0) &= 0, \quad \psi_t(x, 0) = 0 \text{ on } \Omega, \\
\nabla \psi \cdot \mathbf{n} &= 0 \text{ on } \partial \Omega \\
C^*(\mathbf{x}) &= \text{argmin } W_2^2(g(C), h)
\end{aligned}
\]
\begin{align*}
&\frac{1}{C(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(x, t), \\
&\psi(x, 0) = 0, \quad \psi_t(x, 0) = 0 \text{ on } \Omega, \\
&\nabla \psi \cdot n = 0 \text{ on } \partial \Omega \\
&C^*(x) = \arg\min W_2^2(g(C), h) \\
&\text{Previous results involve convexity in changes of the data}
\end{align*}
Wave data is generally not a probability distribution.
Wave data is generally not a probability distribution

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

- Normalize wave data \( k(t) \) by \( \tilde{k}(t) = \frac{k(t)+\gamma}{\int_0^T k(t)+\gamma \, dt} \)

- Computing the \( \mathcal{W}_2 \) distance is difficult for higher dimensions
Wave data is generally not a probability distribution

- Normalize wave data $k(t)$ by
  $$\tilde{k}(t) = \frac{k(t) + \gamma}{\int_0^T k(t) + \gamma \, dt}$$

- 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 $\tilde{k}(t) = \frac{k(t) + \gamma}{\int_0^T k(t) + \gamma \, dt}$

- Computing the $W_2$ distance is difficult for higher dimensions
  - Data is only a function of time $\rightarrow$ apply one dimensional formula

- No explicit solution for wave equation in general
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 $C(X, z) = a + bz$
Ray Tracing

Distance and traveltime formulas:

\[ X = 2 \int z_{p0} \sqrt{u^2 (z) - p^2} \, dz \]

\[ T = 2 \int z_{p0} u^2 (z) \sqrt{u^2 (z) - p^2} \, dz \]

For source wave data \( f(t) \), the predicted wave data is approximated by

\[ A_{\text{pred}} = f(t - T_{\text{pred}}) \]
Ray Tracing

- **Distance and traveltime formulas:**

\[
X = 2p \int_0^{z_p} \frac{dz}{\sqrt{u^2(z) - p^2}}
\]

\[
T = 2 \int_0^{z_p} \frac{u^2(z)}{\sqrt{u^2(z) - p^2}} \, dz
\]
Distance and traveltime formulas:

\[ X = 2p \int_0^{z_p} \frac{dz}{\sqrt{u^2(z) - p^2}} \]

\[ T = 2 \int_0^{z_p} \frac{u^2(z)}{\sqrt{u^2(z) - p^2}} \, dz \]

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

Assume $f(t)$ is nonnegative and compactly supported. Then, $W^2_2(\tilde{g}, \tilde{h})$ 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 $S_0$. 
Theorem

Assume \( f(t) \) is nonnegative and compactly supported. Then, \( W_2^2(\tilde{g}, \tilde{h}) \) 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 \((a^*, b^*)\)
Theorem

Assume $f(t)$ is nonnegative and compactly supported. Then, $W^2_2(\tilde{g}, \tilde{h})$ 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

![Graph 1](image1)

![Graph 2](image2)

![Graph 3](image3)
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
I am extremely thankful to:

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


References