The Analysis and Application of Optimal Transport Related Misfit Functions in Seismic Imaging

Yunan Yang[1]

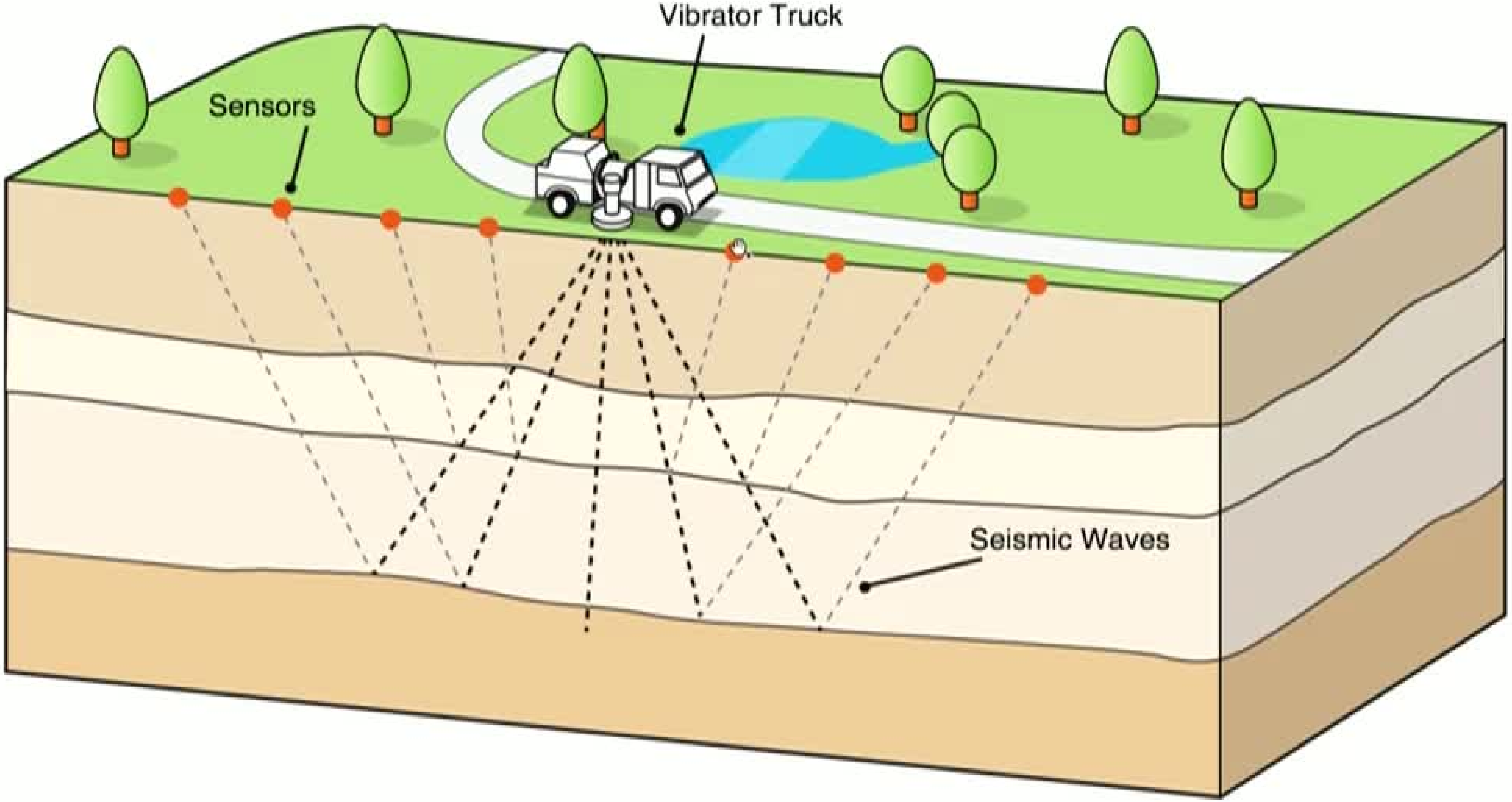
July 13, 2018

[1]The University of Texas at Austin, USA

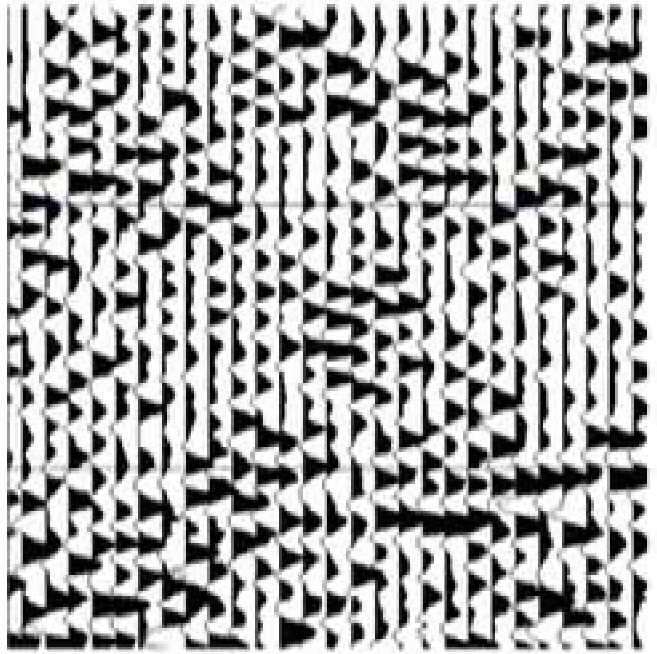
Special thanks to:
Dr. Björn Engquist (supervisor)
Junzhe Sun (ExxonMobil), Lingyun Qiu (Petroleum Geo-Services, Inc) and Brittany Froese (NJIT)
Texas Consortium for Computational Seismology (TCCS) colleagues and sponsors
Petroleum Geo-Services Modeling and Inversion Group

Background

Seismic inversion



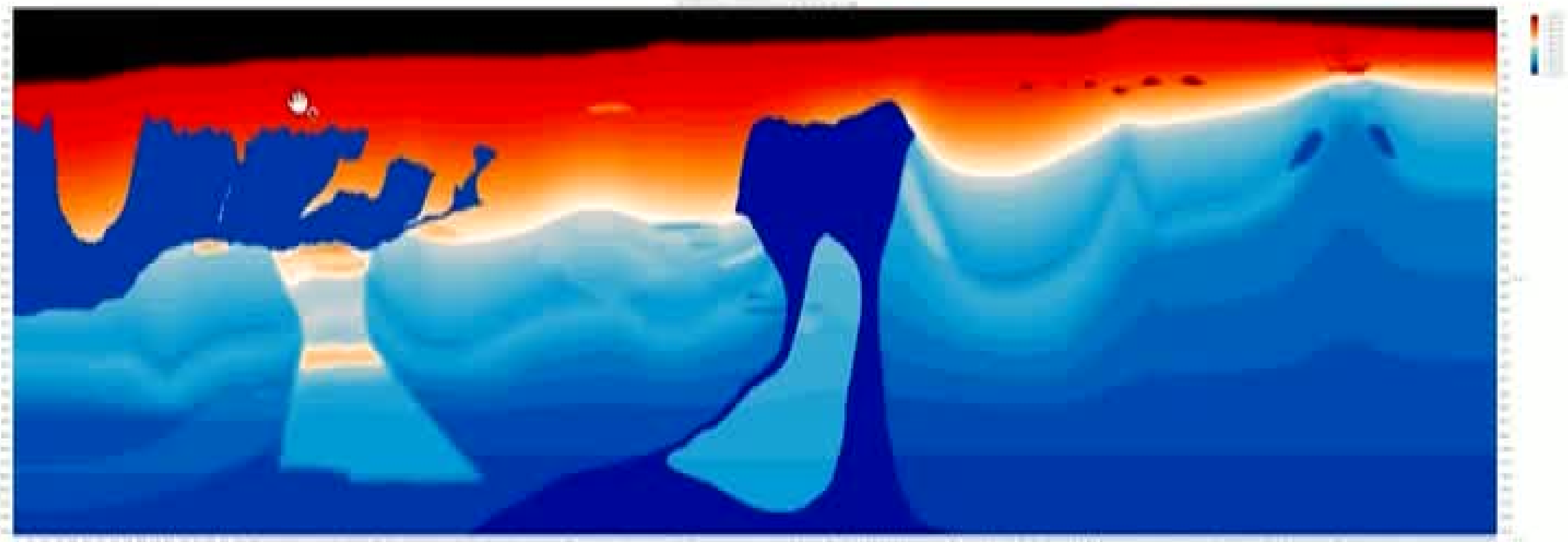
Seismic inversion



Waveforms from receivers (i.e. wave equation solution on the boundary)

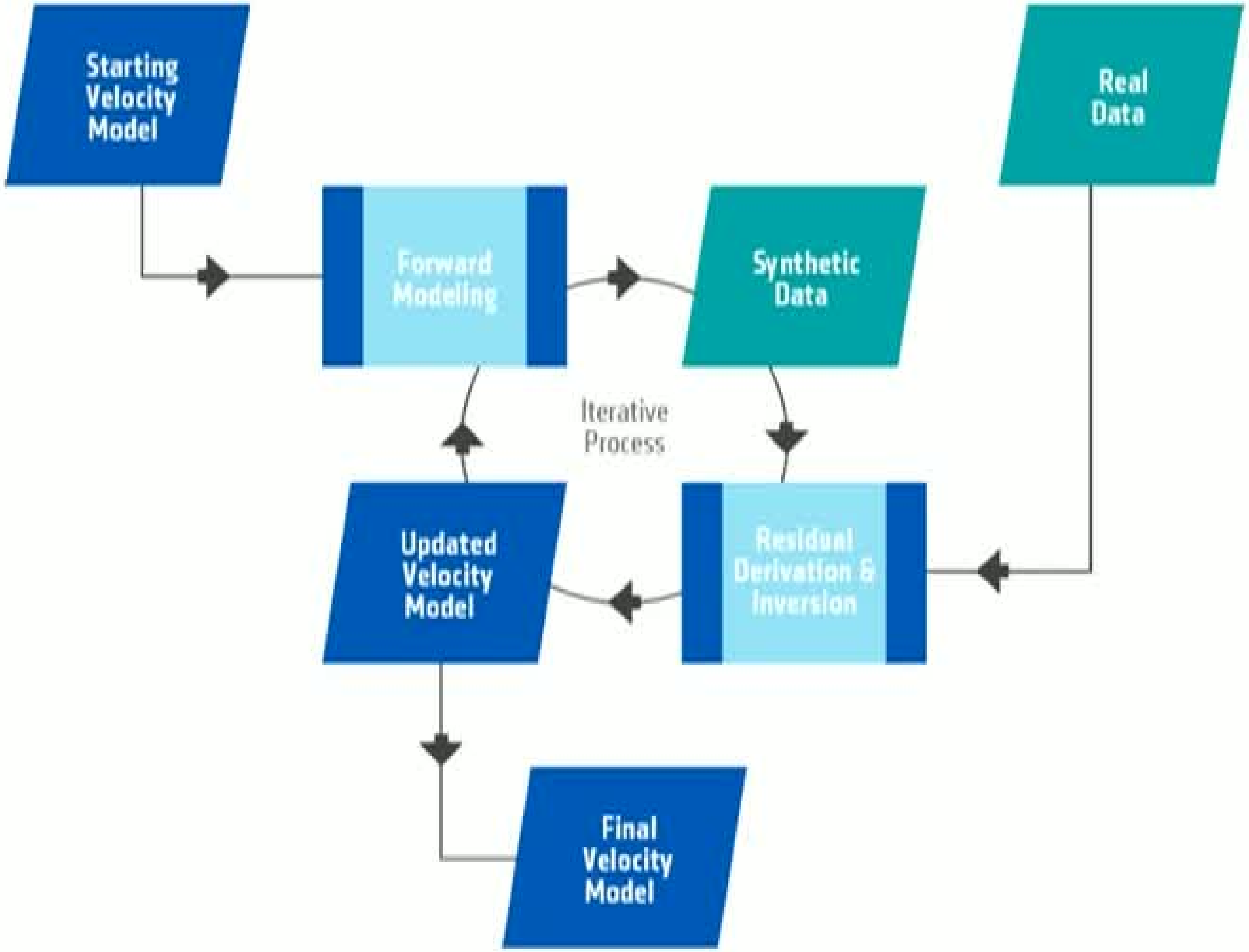
$\Longrightarrow$

Invert



The velocity under the ground/sea surface (i.e. velocity/bulk modulus/impedance in the wave equation)

Full Waveform Inversion (FWI): a PDE-constrained optimization

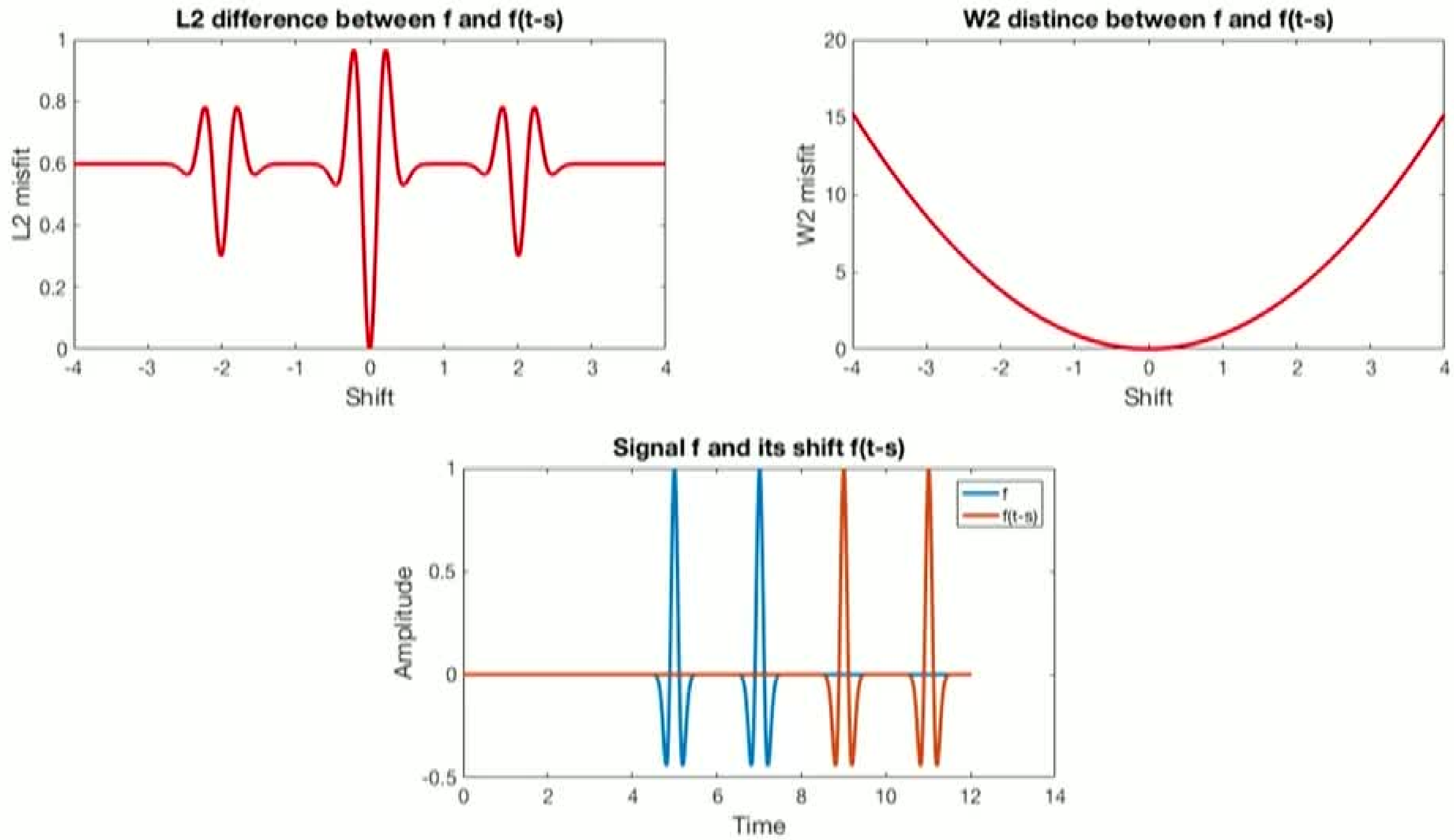


$$\begin{cases} m(\mathbf{x})\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \triangle u(\mathbf{x},t) = s(\mathbf{x},t) \\ u(\mathbf{x},0) = 0 \\ \frac{\partial u}{\partial t}(\mathbf{x},0) = 0 \end{cases}$$

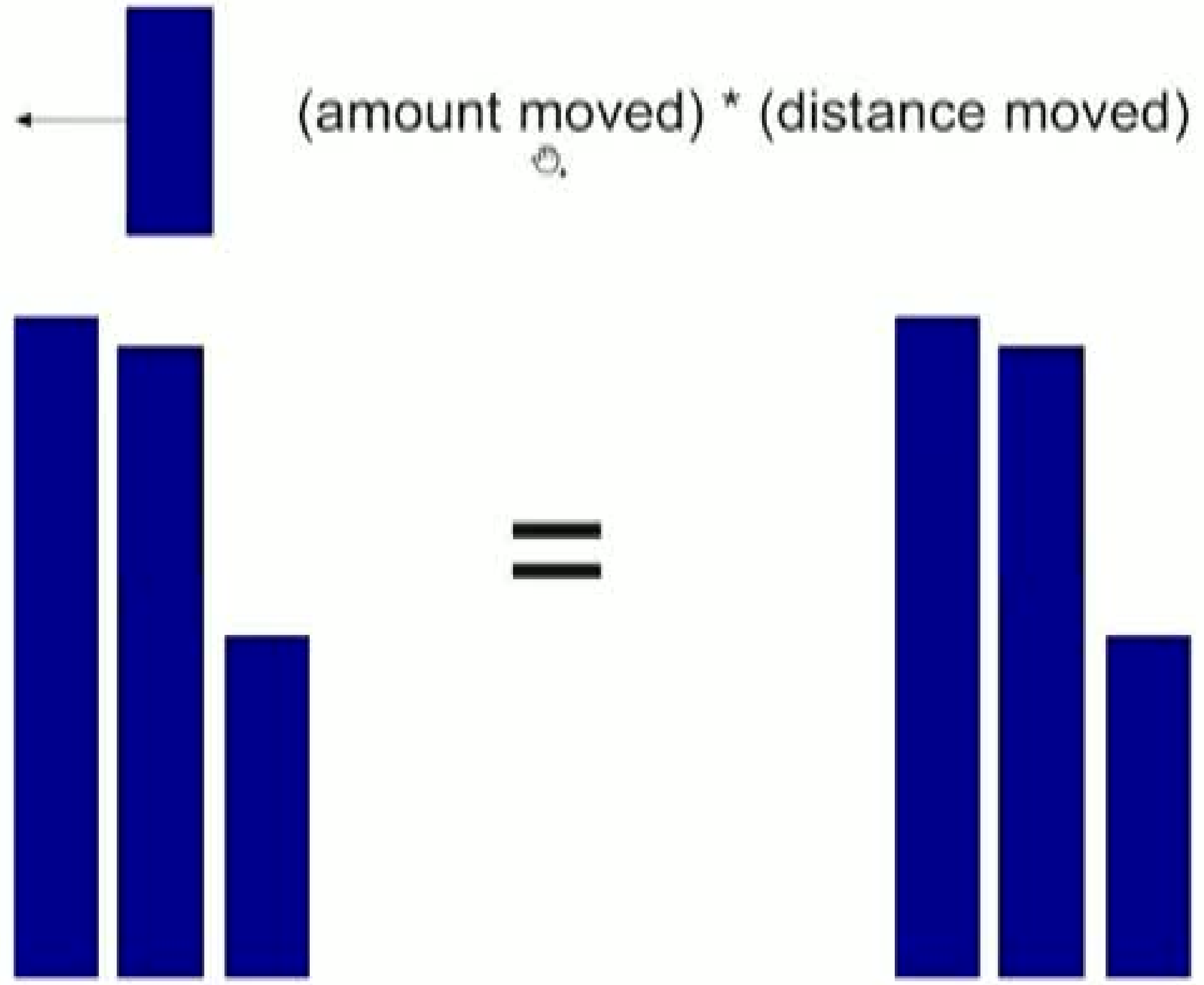
$$m^\star = \underset{m}{\operatorname{argmin}}\, \chi(f(m), g),$$

χ is the objective function;
$f = Ru$ is the simulated data;
g is the reference/true data.

Limitation of L^2 norm — Many local minima



Optimal transport

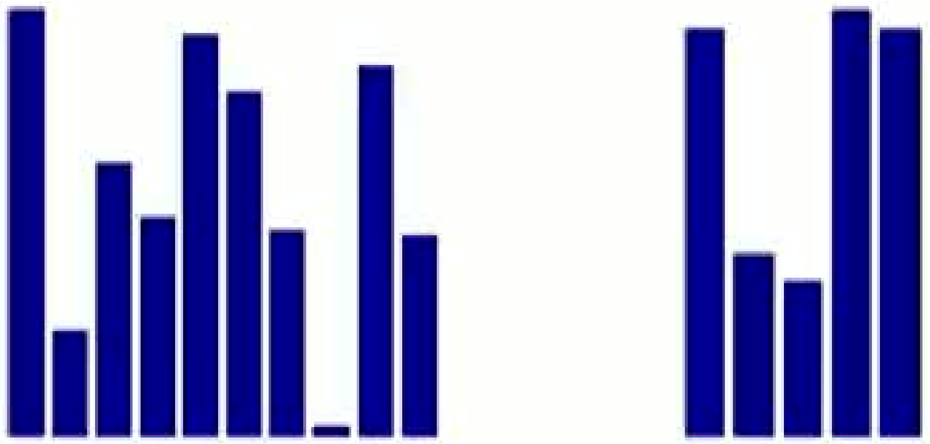


Synthetic data f (left) and observed data g (right)

Optimal transport: the Wasserstein distance

Finally, for general functions f and g, the Wasserstein distance is

$$\min_{\text{All the map } T} \left(\sum_{\text{All movements of } T} \text{distance moved} \times \text{amount moved} \right)$$



Function f and g sharing the same mass by normalization

Different choice of distance: W_1 ($|x - y|$) and W_2 ($|x - y|^2$)

Quadratic Wasserstein Distance (Earth Mover's Distance)

Definition of the Wasserstein distance

For $f : X \to \mathbb{R}^+$, $g : Y \to \mathbb{R}^+$, the distance can be formulated as

$$W_p(f, g) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p f(x) dx \right)^{\frac{1}{p}} \tag{1}$$

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

Quadratic Wasserstein distance: $p = 2$

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

Properties of W_2

Convexity: motivation

The shift and dilation are typical effects from variations in velocity c. For example:

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, & x > 0, t > 0, \\ u = 0, \quad \frac{\partial u}{\partial t} = 0, & x > 0, t = 0, \\ u = f(t), & x = 0, t > 0. \end{cases}$$

The solution to the equation is $u(c; x, t) = f(t - x/c)$.

For fixed x, variation in c relates **shifts** in the signal.

For fixed t, variation in c generates the **dilation** in f as a function of x.

The change in amplitude may originate from measurement errors and variations in strength of reflecting surfaces.

Convexity: translation and dilation (for any dimension)

Theorem (Convexity of translation and dilation)

$W_2^2(f,g)$ is convex with respect to translation, s and dilation, λ:

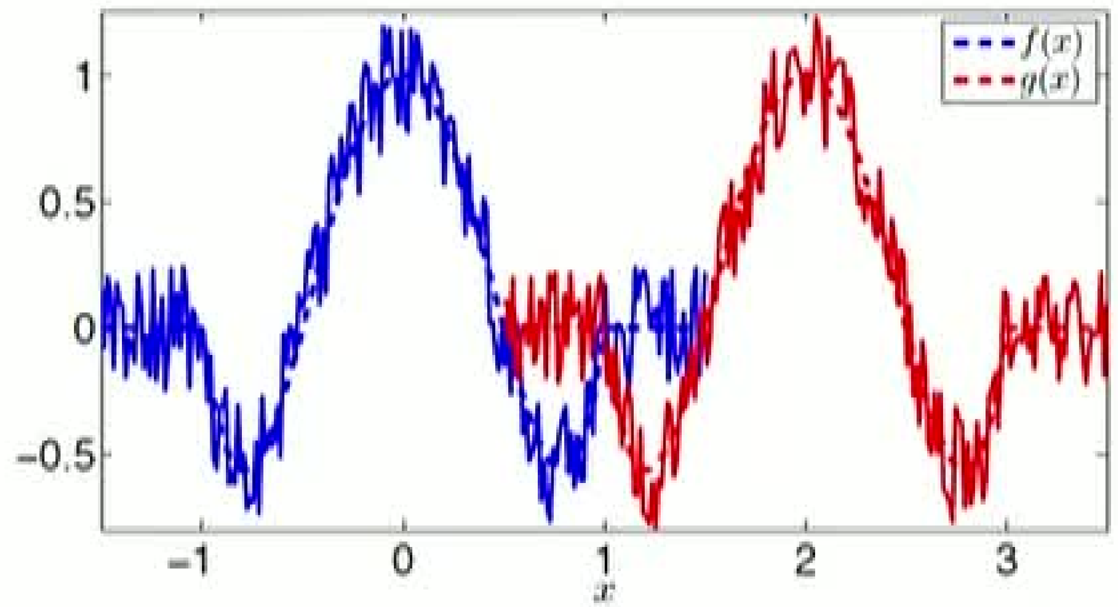
$$W_2^2(f,g)[s,\lambda], \quad f(x) = \lambda^d g(\lambda x - s), \quad \lambda > 0, \quad s, x \in \mathbb{R}^d$$

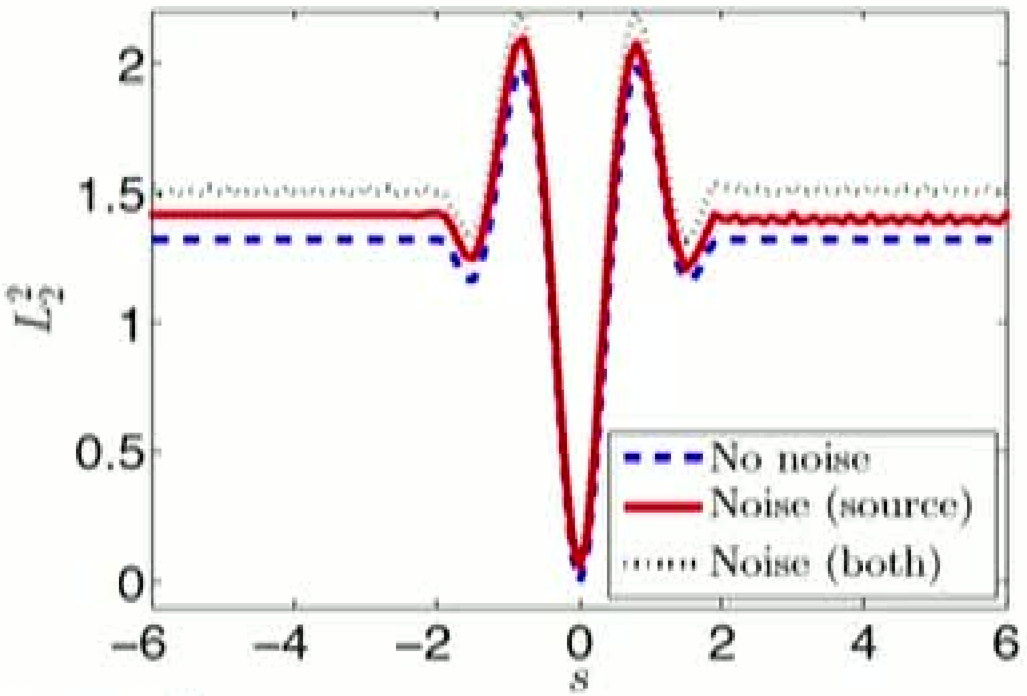
- The dilation λx can be generalized to Ax, where A is a symmetric positive definite matrix. Then the convexity is in terms of the eigenvalues.
- The proof is based on c-cyclic monotonicity of the transference plan Γ: For any $m \in \mathbb{N}^+$, $(x_i, y_i) \in \Gamma, 1 \leq i \leq m$, and any permutation σ:

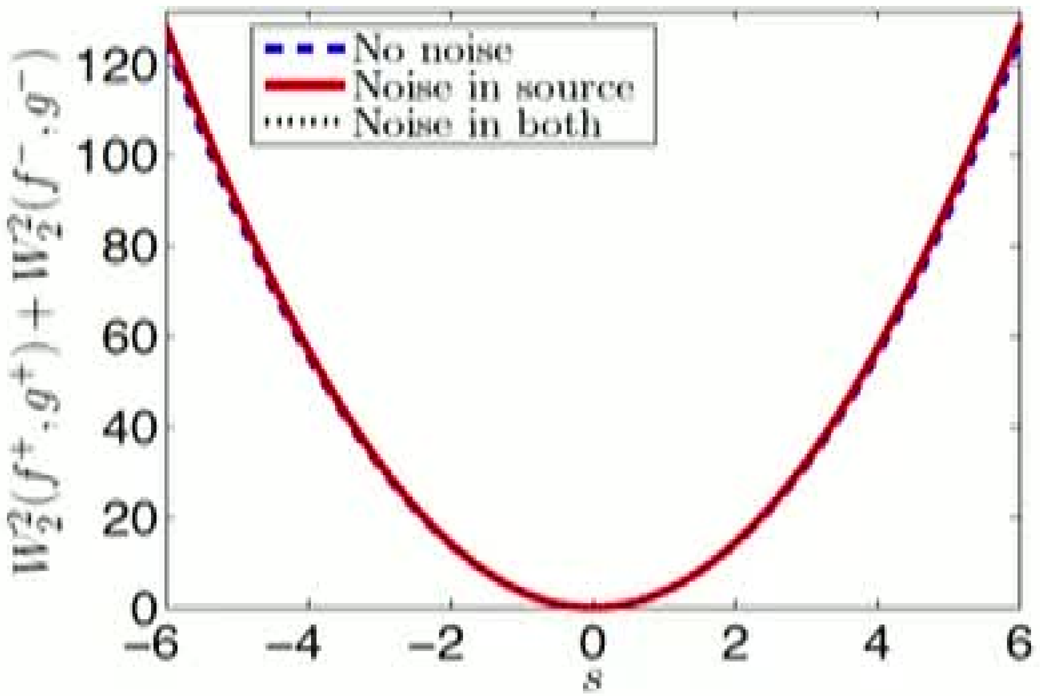
$$\sum_{i=1}^{m} c(x_i - y_i) \leq \sum_{i=1}^{m} c(x_i - y_{\sigma(i)}) \tag{3}$$

where $x_0 \equiv x_m$ and $y_0 \equiv y_m$.

W_2: Insensitivity to noise







[Engquist and Froese, 2014]

1D Optimal Transport approach

W_2 for 1D data

1D Optimal transport (trace by trace)

The explicit formulation for the 1D Wasserstein metric is:

$$W_2^2(f,g) = \int_0^1 |F^{-1}(x) - G^{-1}(x)|^2 dx. \tag{4}$$

where $F(t) = \int_{-\infty}^{t} \tilde{f}(\tau)d\tau$ and $G(t) = \int_{-\infty}^{t} \tilde{g}(\tau)d\tau$. $\tilde{f}$ and $\tilde{g}$ are normalized signals that have positivity and conservation of mass. The optimal map is $G^{-1} \circ F$.

Get F^{-1} in $O(N)$ complexity. N is the total number of data points in time.

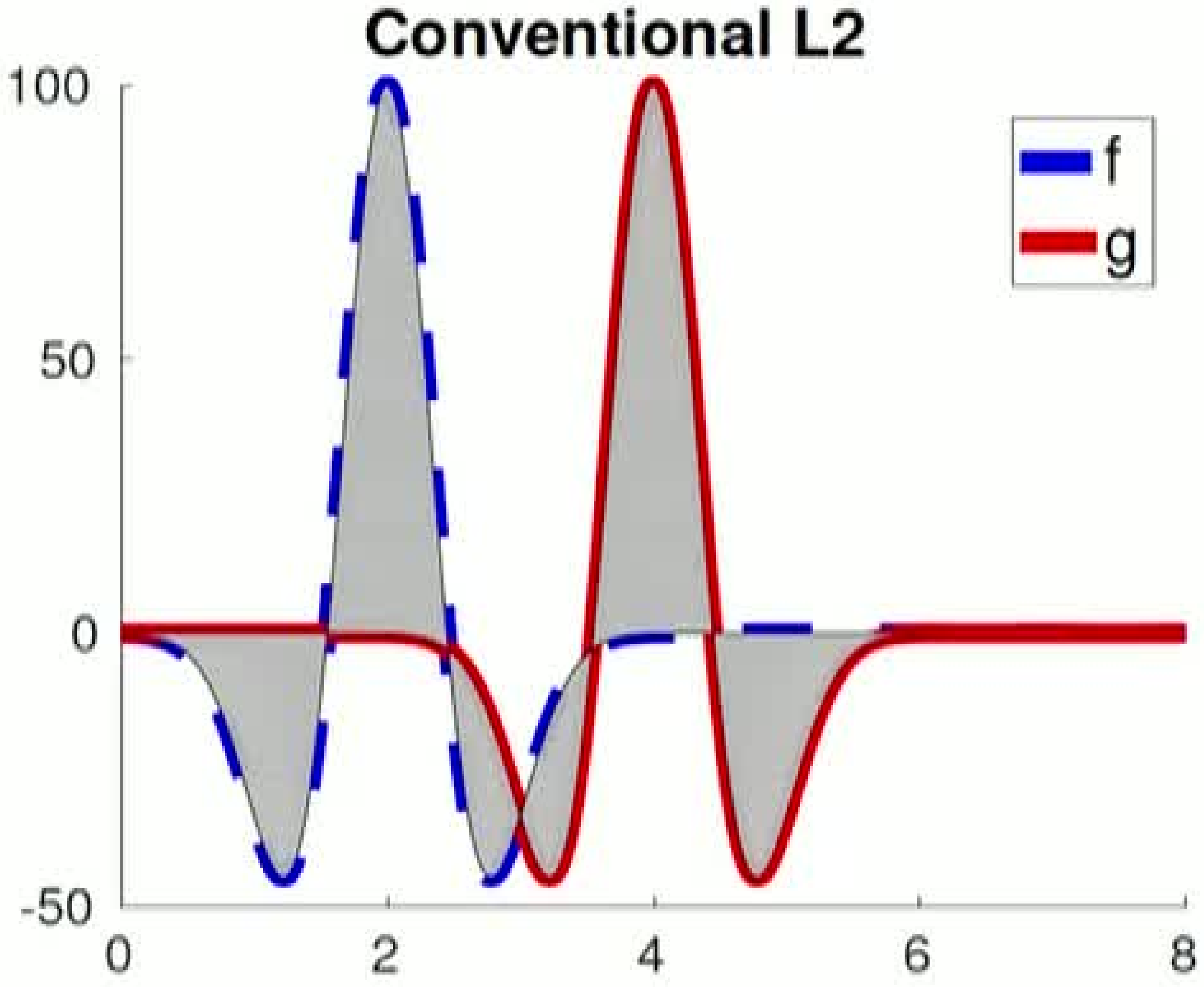
Data transformation

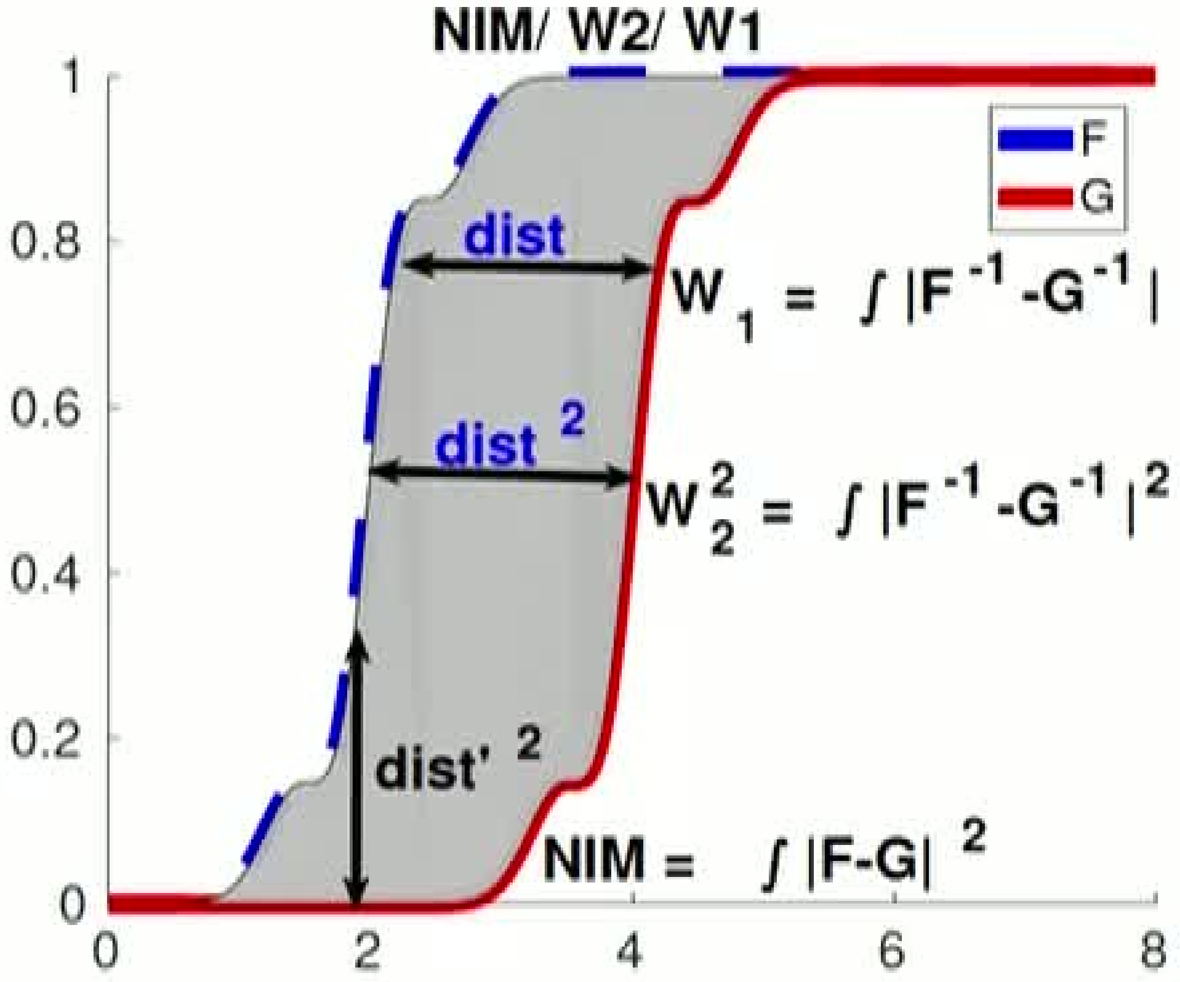
$$f \xrightarrow[\textit{very important!}]{\text{normalize}} \tilde{f} \xrightarrow{\text{integrate in time}} F \xrightarrow{\text{inverse function}} F^{-1}$$

[Yang et al., 2016]

The fundamental difference between L^2 and 1D optimal transport

Signal g (red) is a shift of Ricker wavelet f (blue).





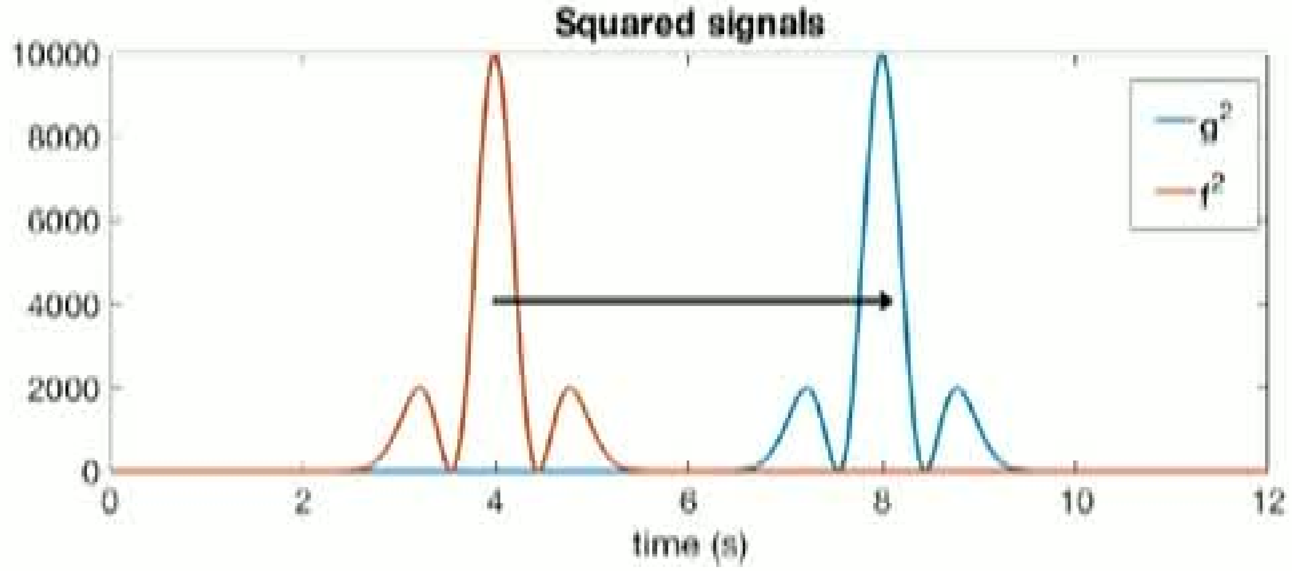
L^2 : $\int_0^T (f-g)^2 dt$

W_2 : $\int_0^1 |F^{-1}(x) - G^{-1}(x)|^2 dx.$

[Yang and Engquist, 2017]

Data Normalization: Bridging the Gap Between Seismic Signals and Probability Densities

How to normalize? The squaring scaling



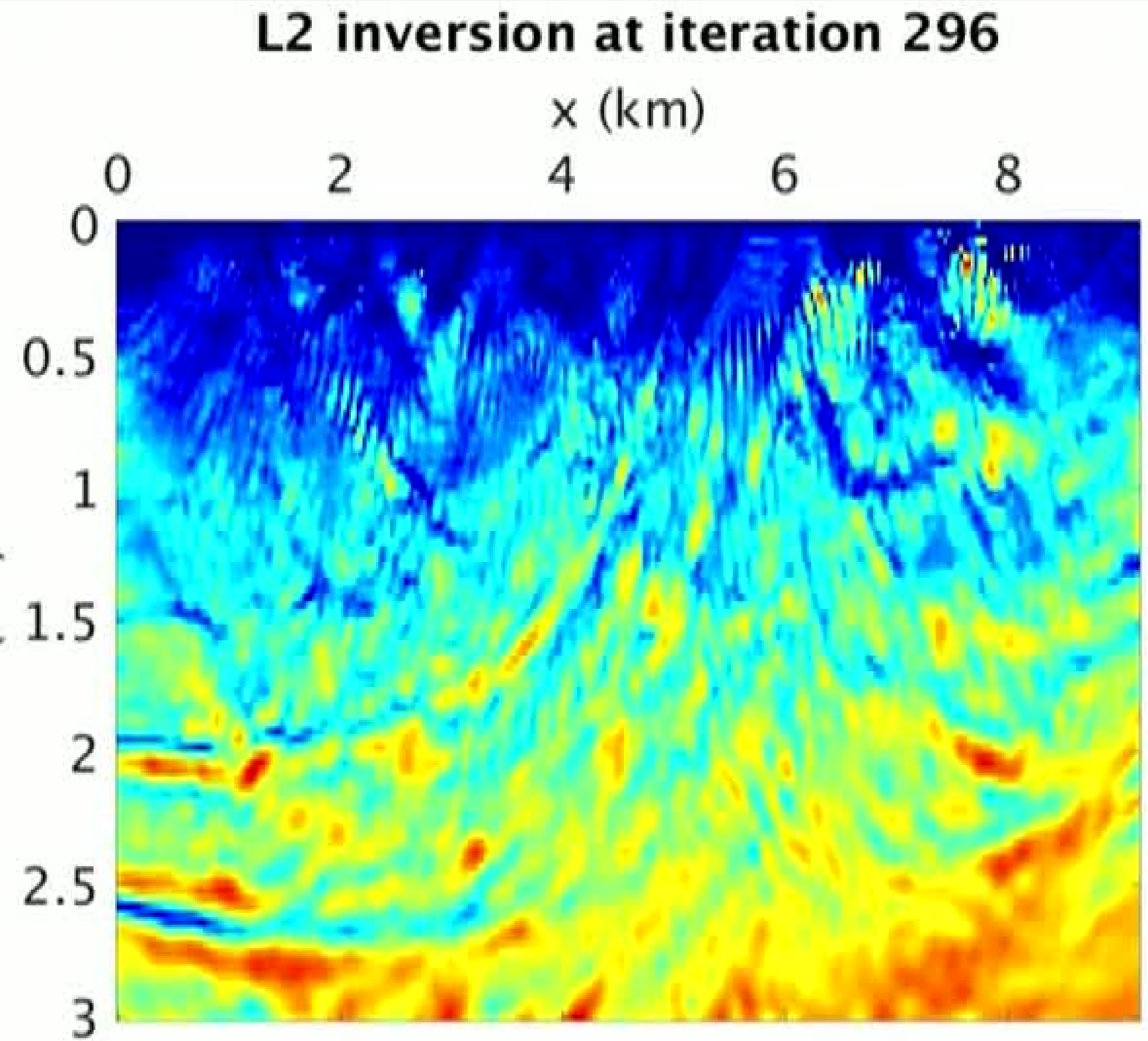
Square of the data: $f \to f^2$ and $g \to g^2$ (arrows indicate transport)

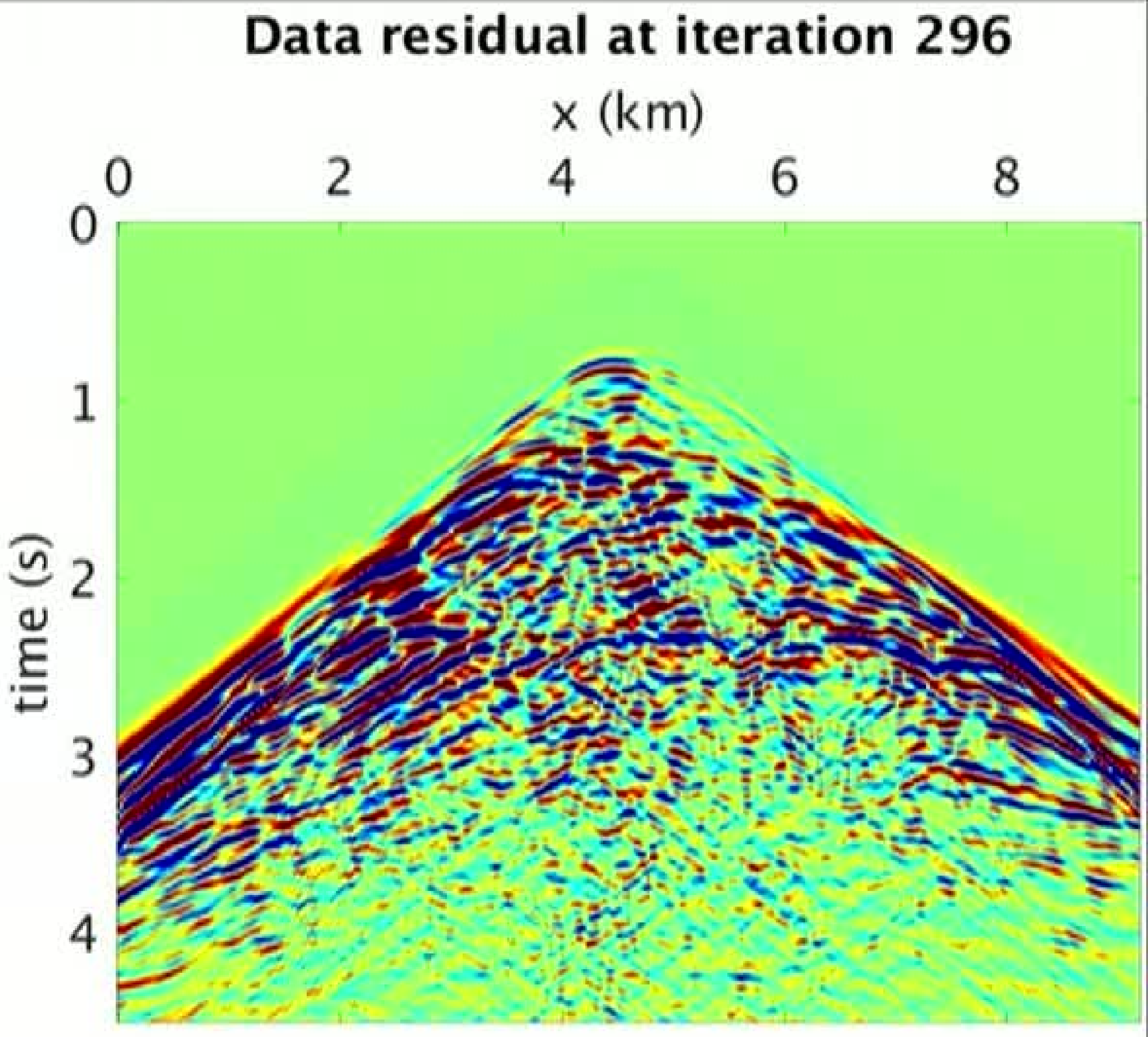
Why is it not working well in inversions?

- Taking the squares boosts the higher frequency of the signal
- Squaring the signals may lose the important phase information (Refraction vs. Reflection)
- Not a one-to-one function; cannot recover the original signal after normalization

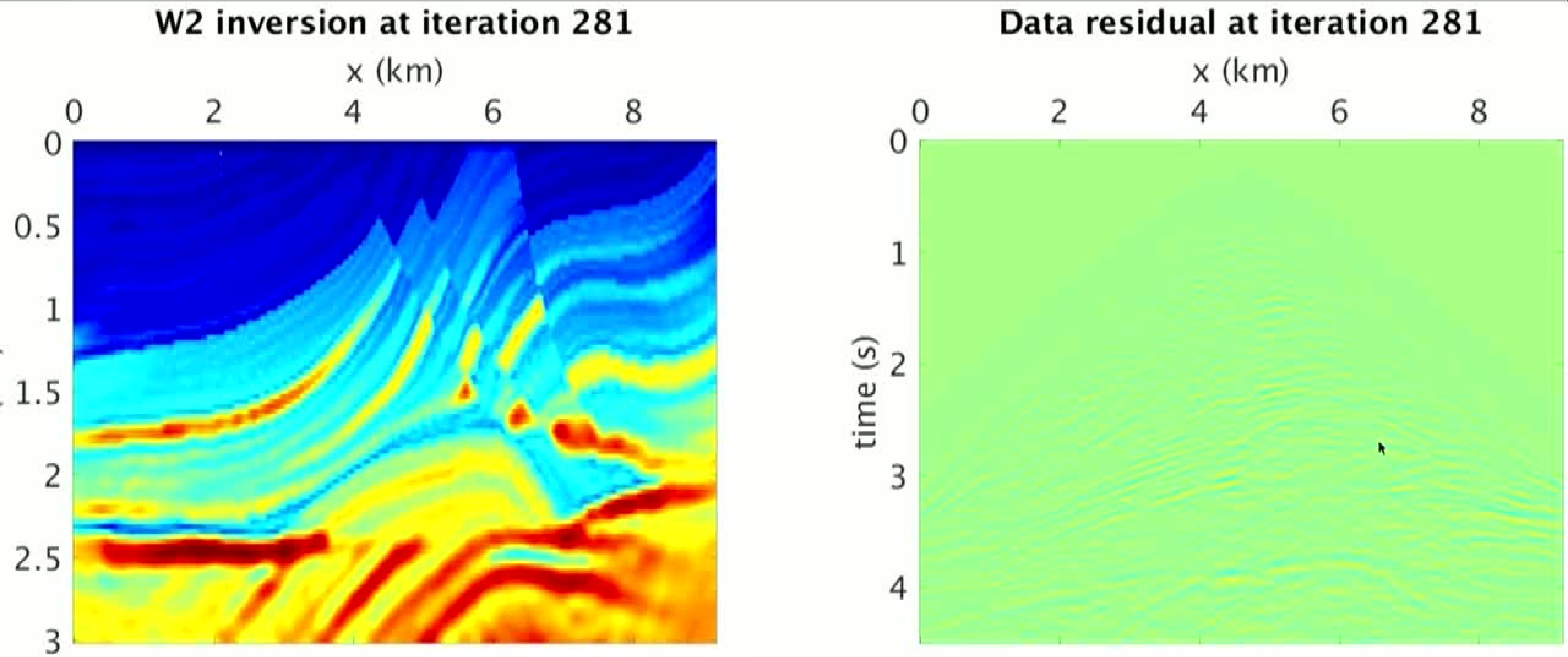
Inversion results

Marmousi model with L^2 norm



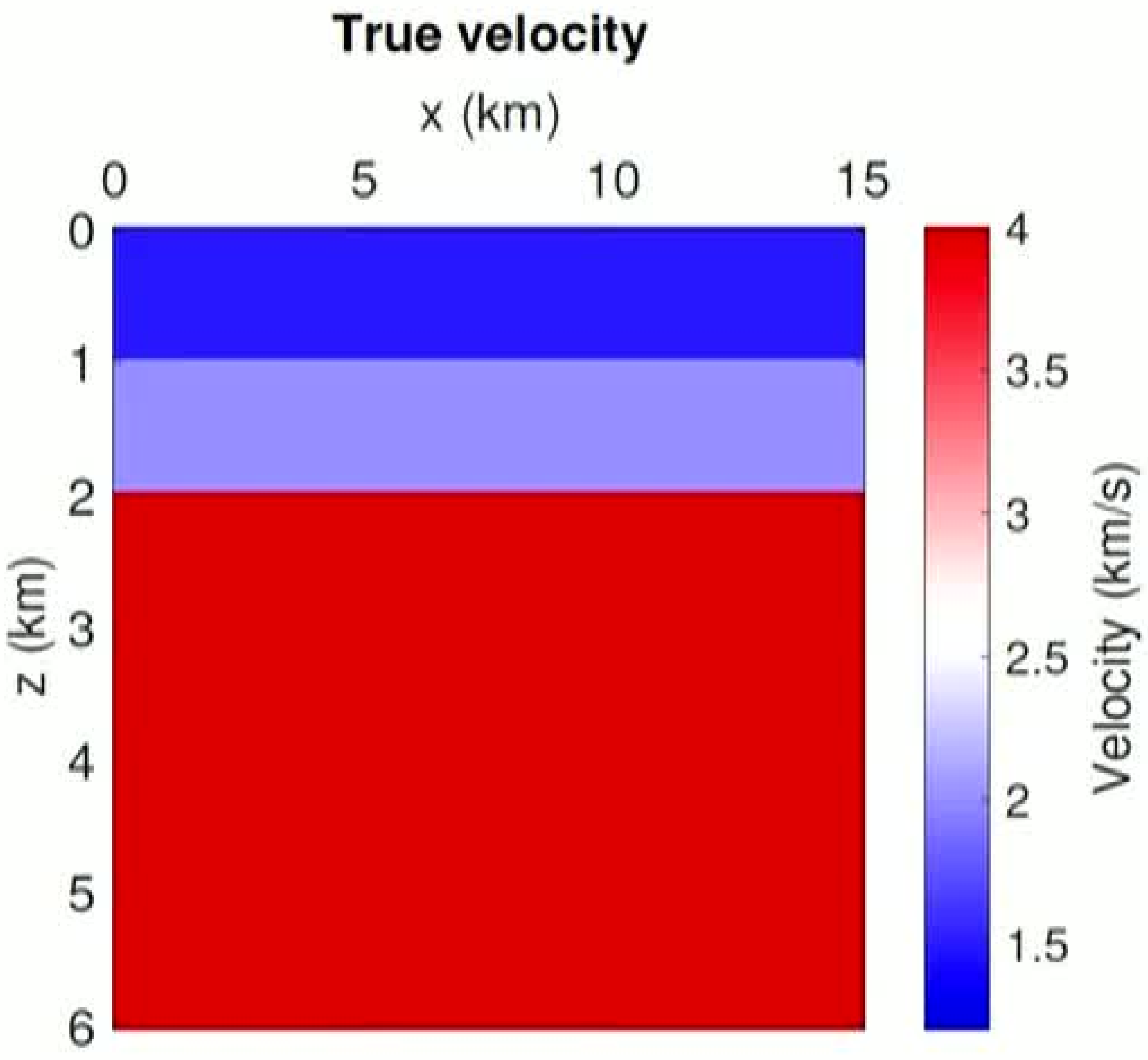


Marmousi model with W_2 norm

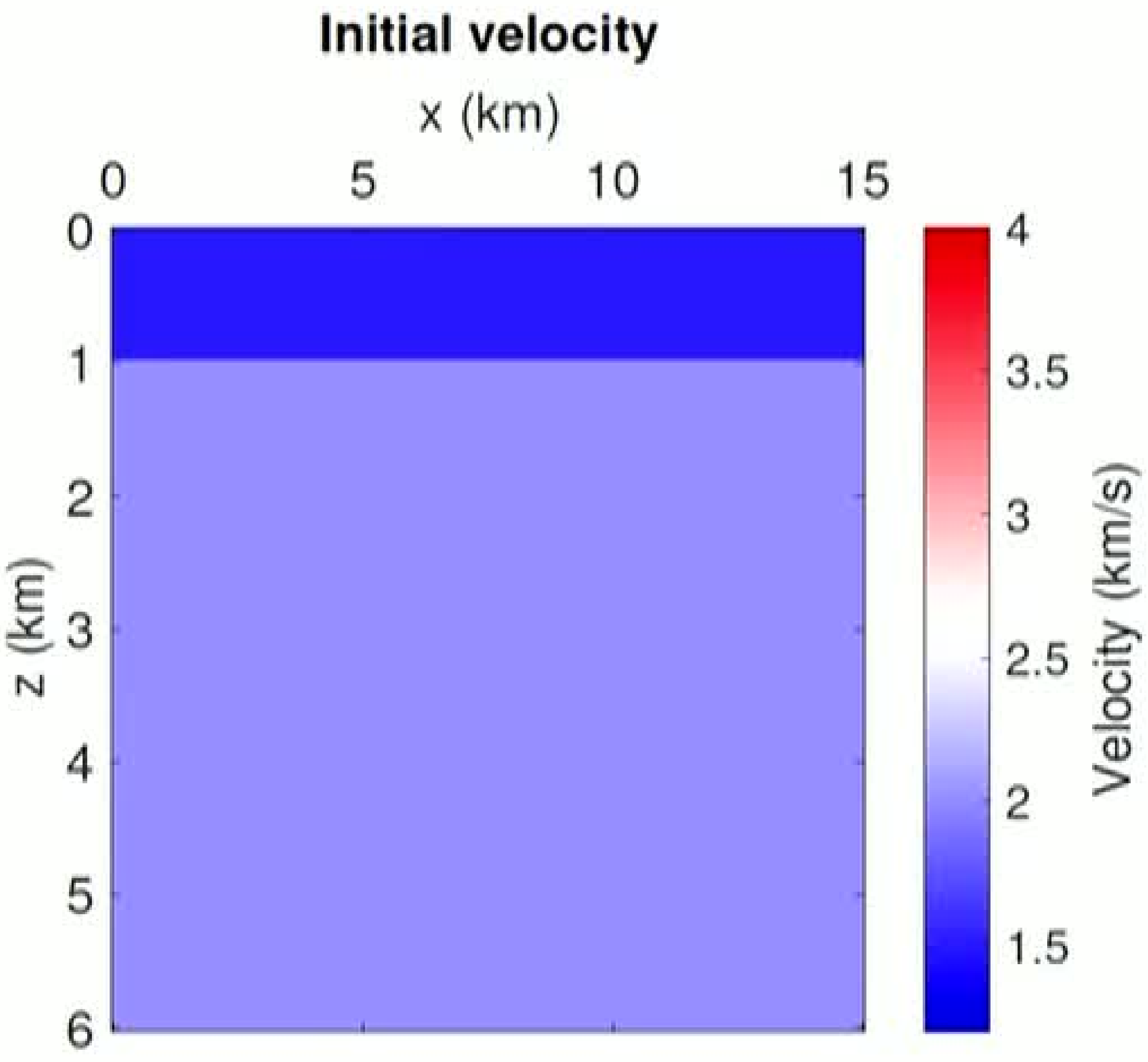


Reflection dominated FWI: local minima beyond cycle skipping

The difficulty of L^2 norm in Reflection FWI



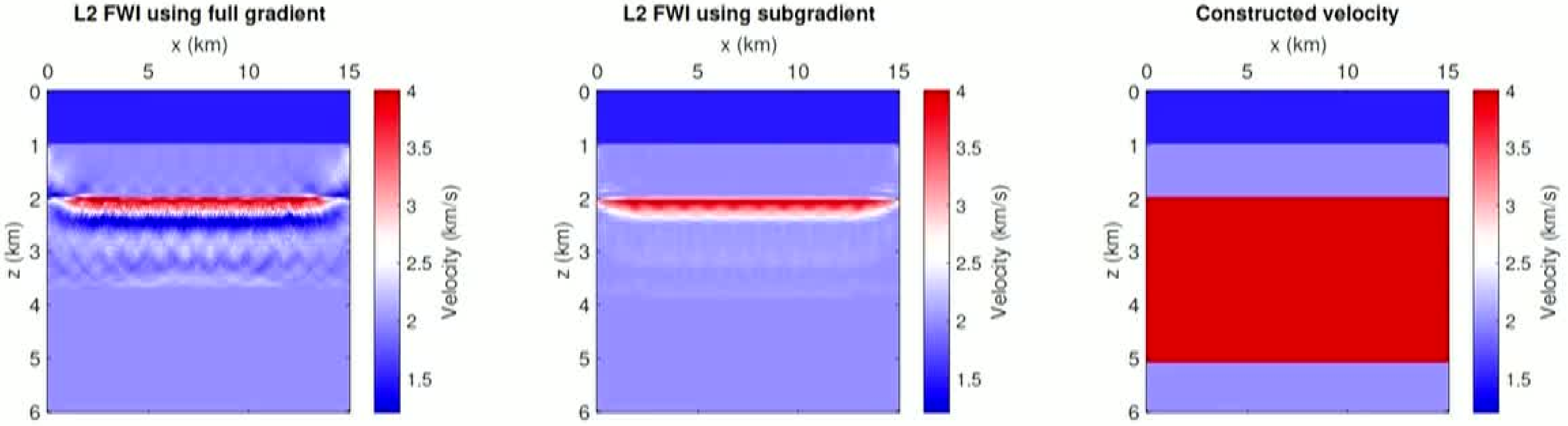
true velocity



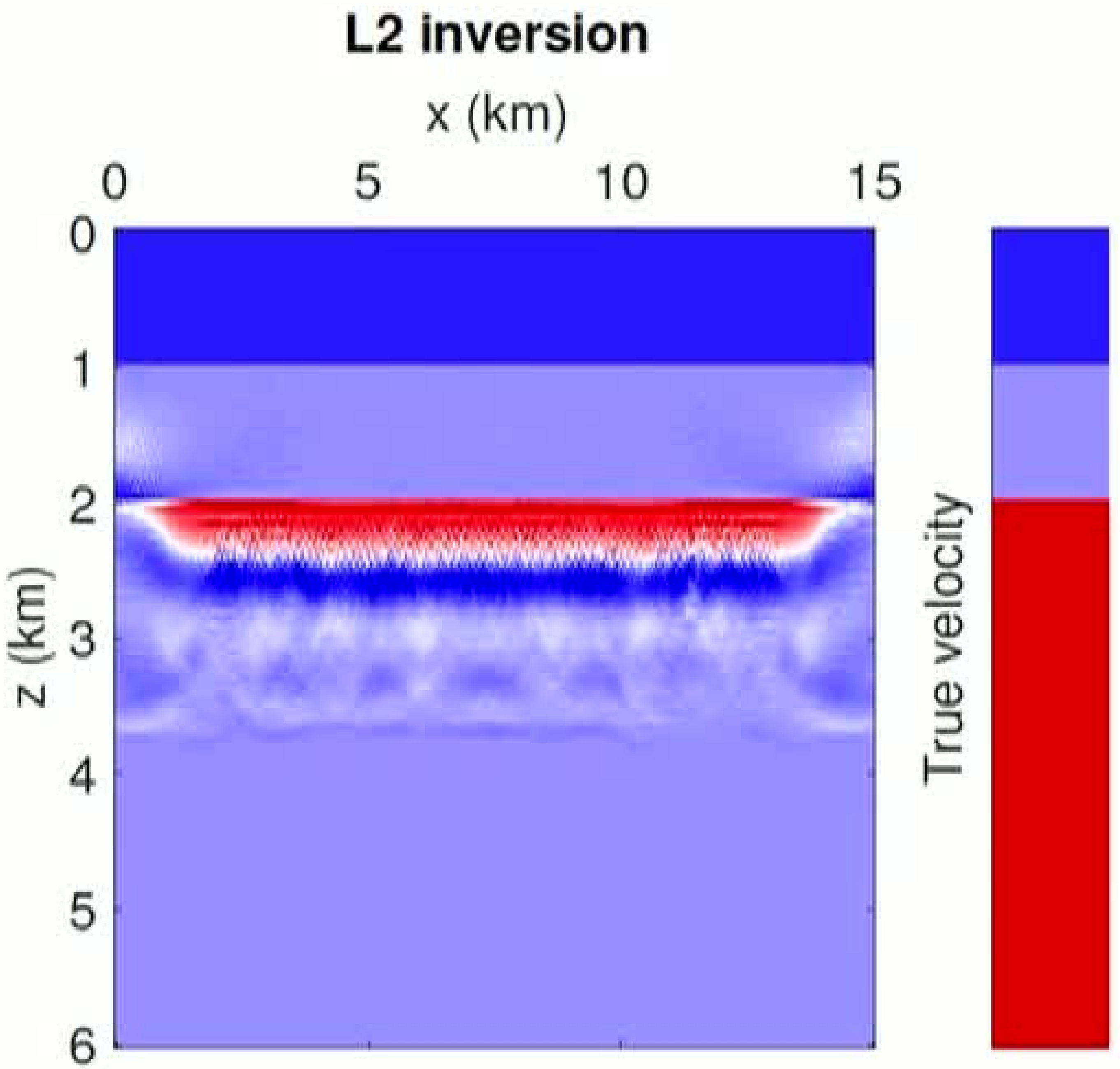
initial velocity

The difficulty of L^2 norm in Reflection FWI: local minima

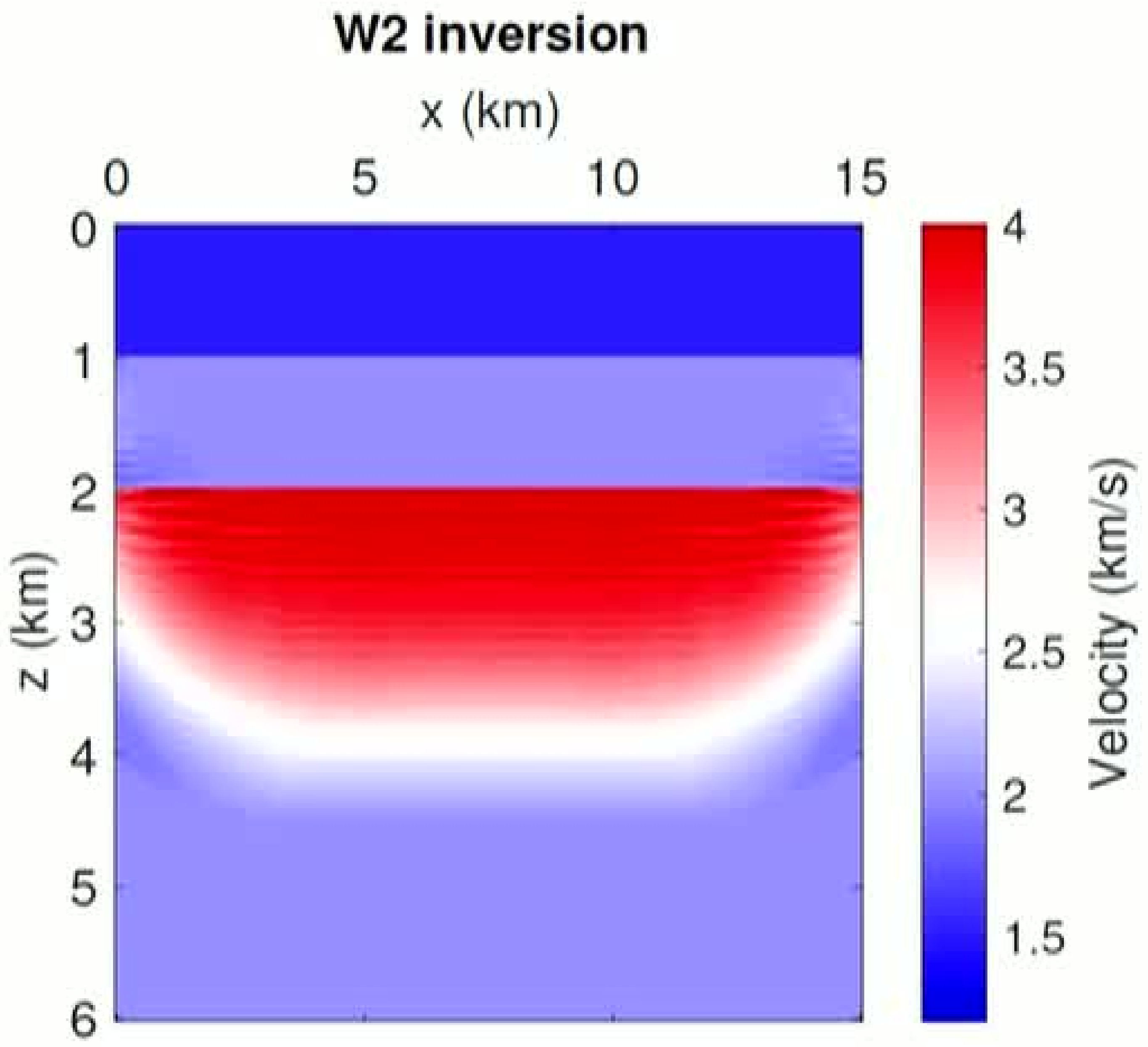
Three velocity models share the same L^2 data misfit



The difficulty of L^2 norm in Reflection FWI

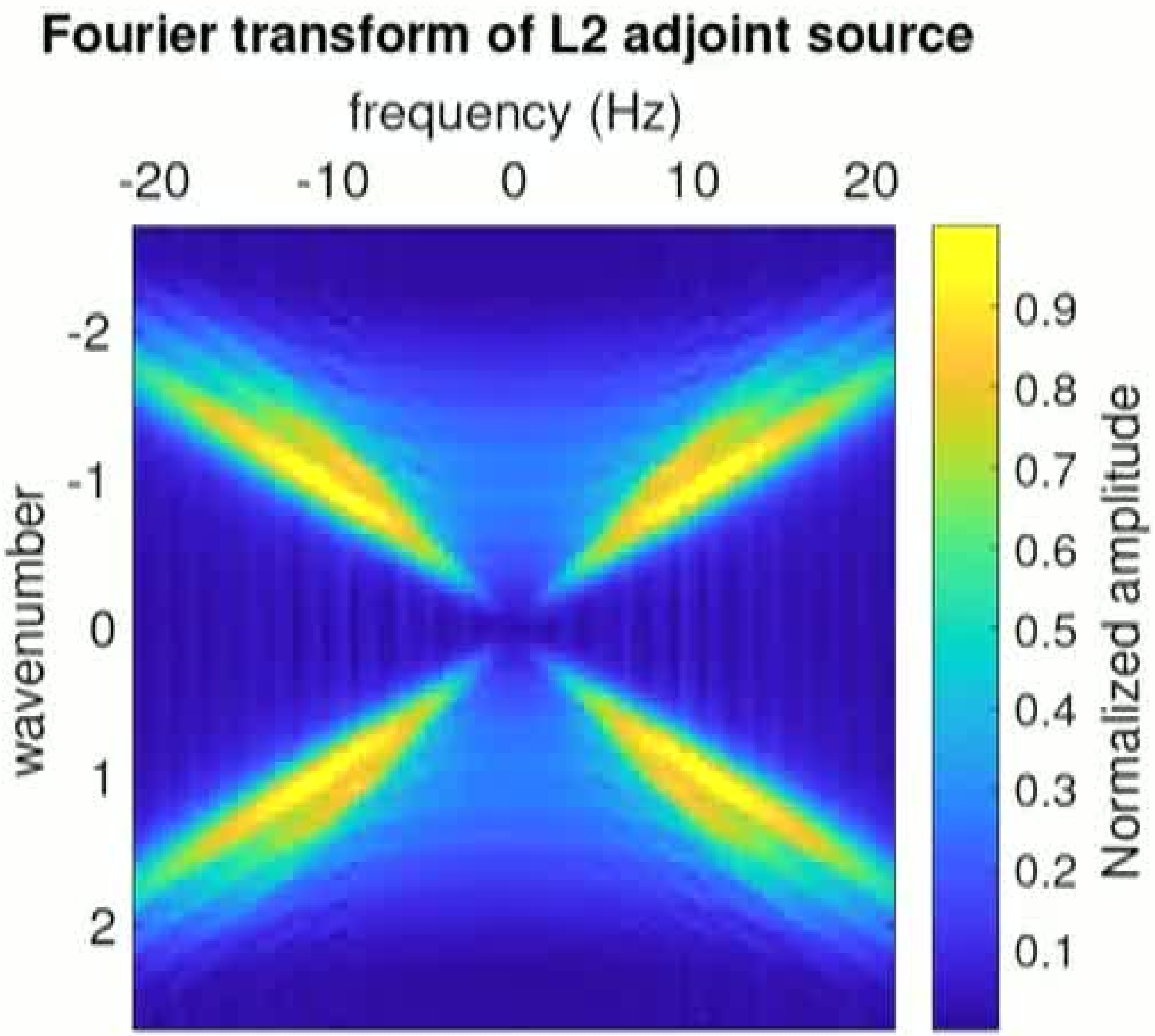


L^2 inversion

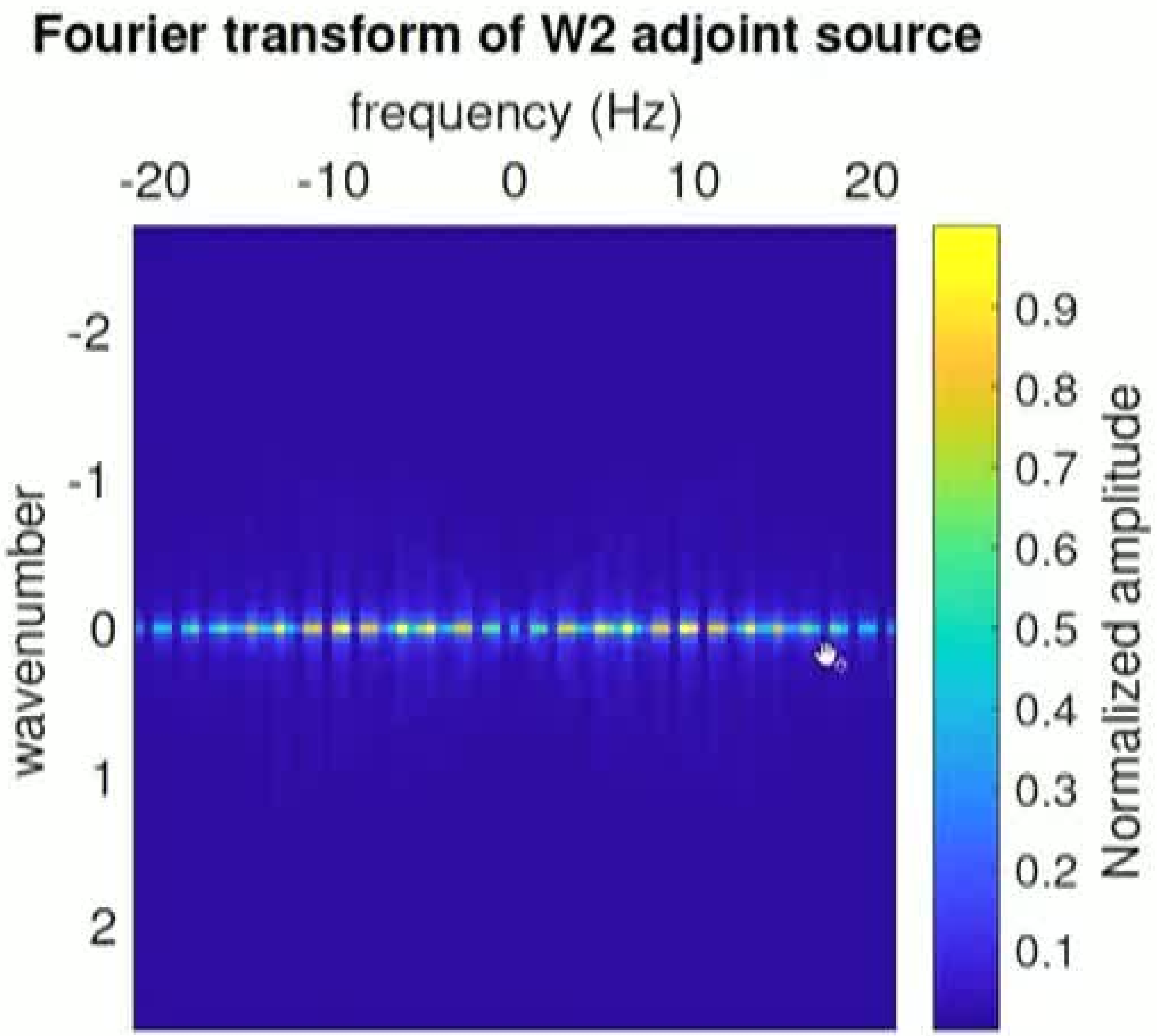


W_2 inversion

The difficulty of L^2 norm in Reflection FWI

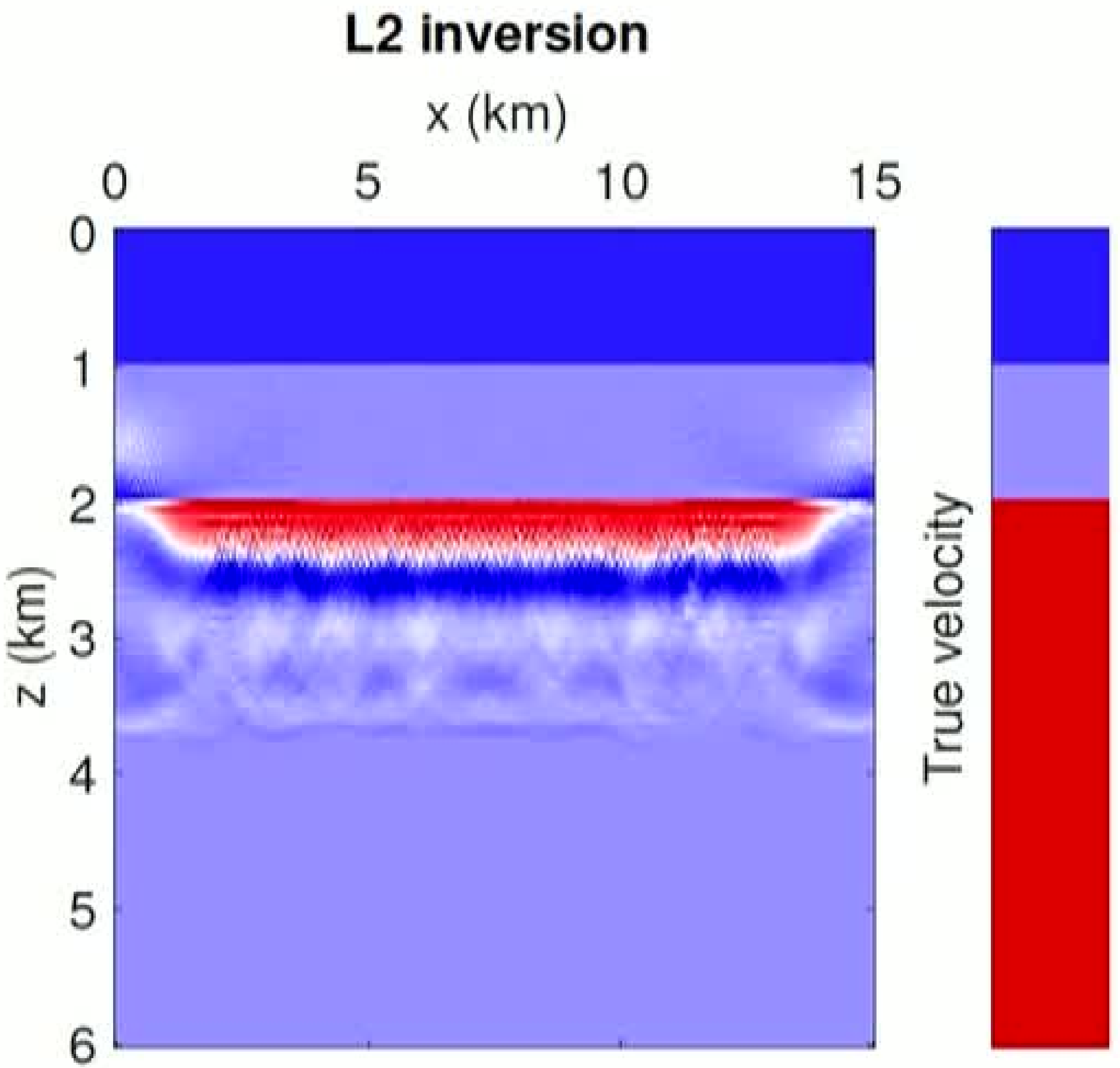


The Fourier transform of L^2 data gradient

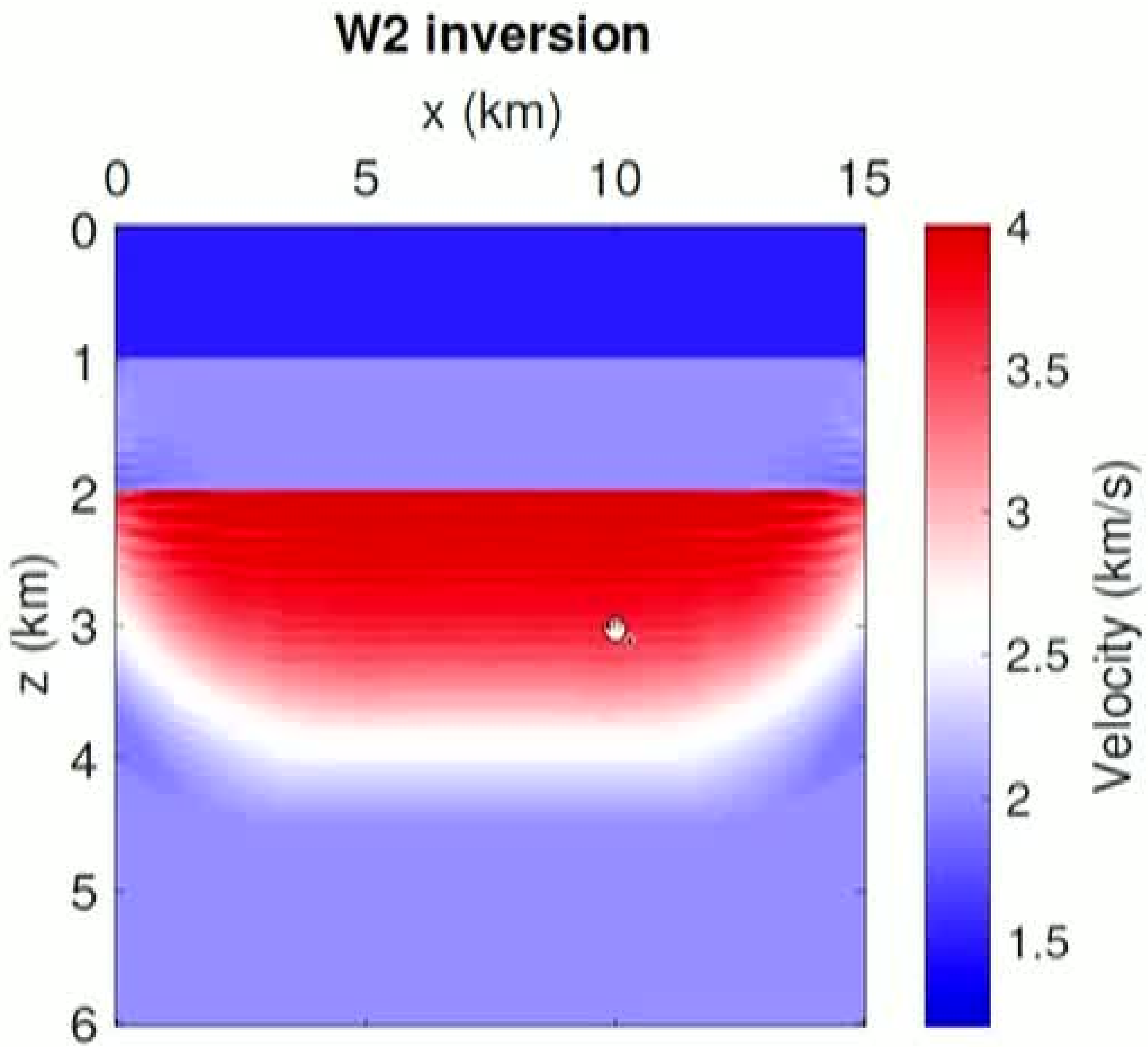


The Fourier transform of W_2 data gradient

The difficulty of L^2 norm in Reflection FWI

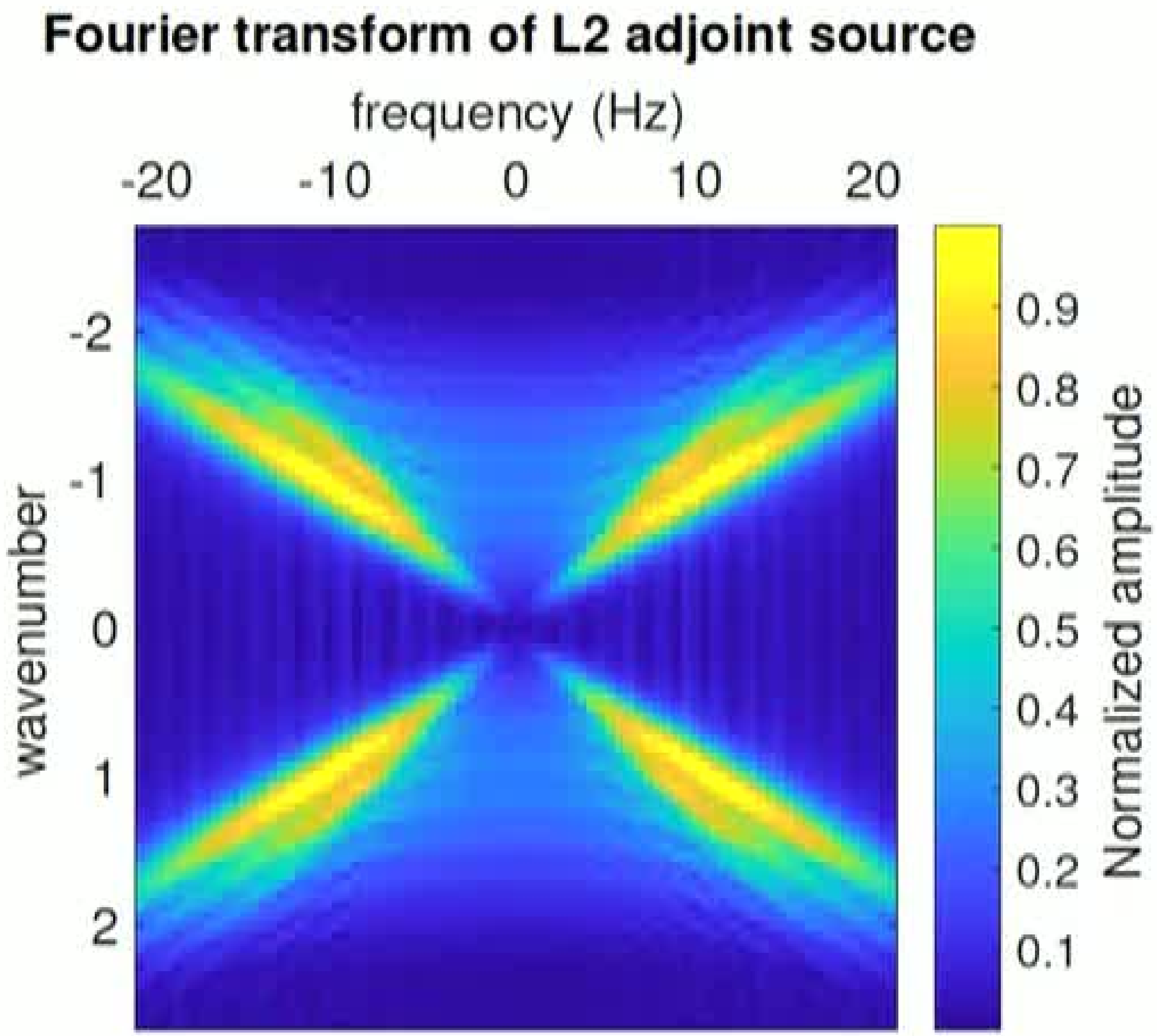


L^2 inversion

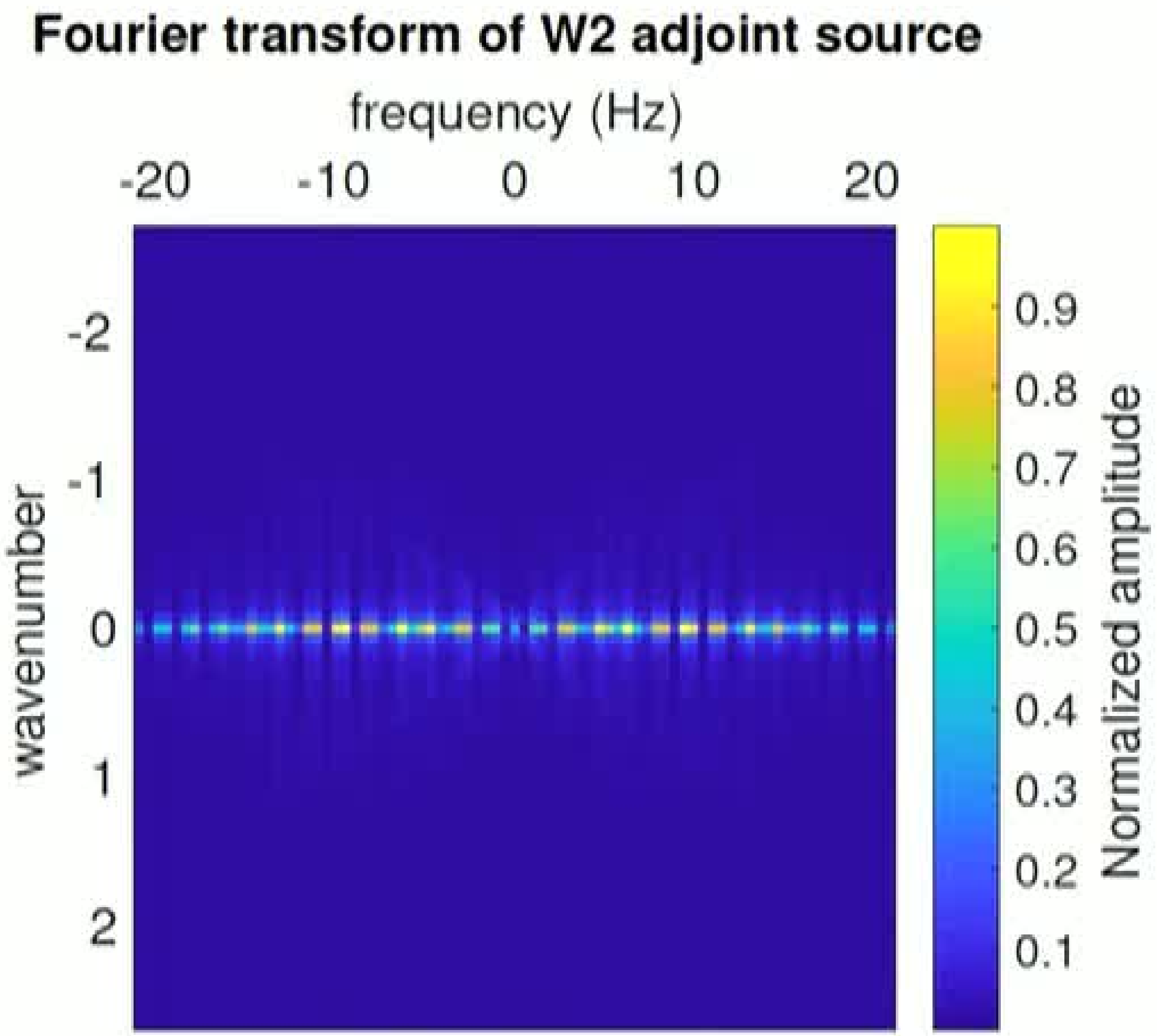


W_2 inversion

The difficulty of L^2 norm in Reflection FWI



The Fourier transform of L^2 data gradient



The Fourier transform of W_2 data gradient

The data gradient of W_2 norm in FWI

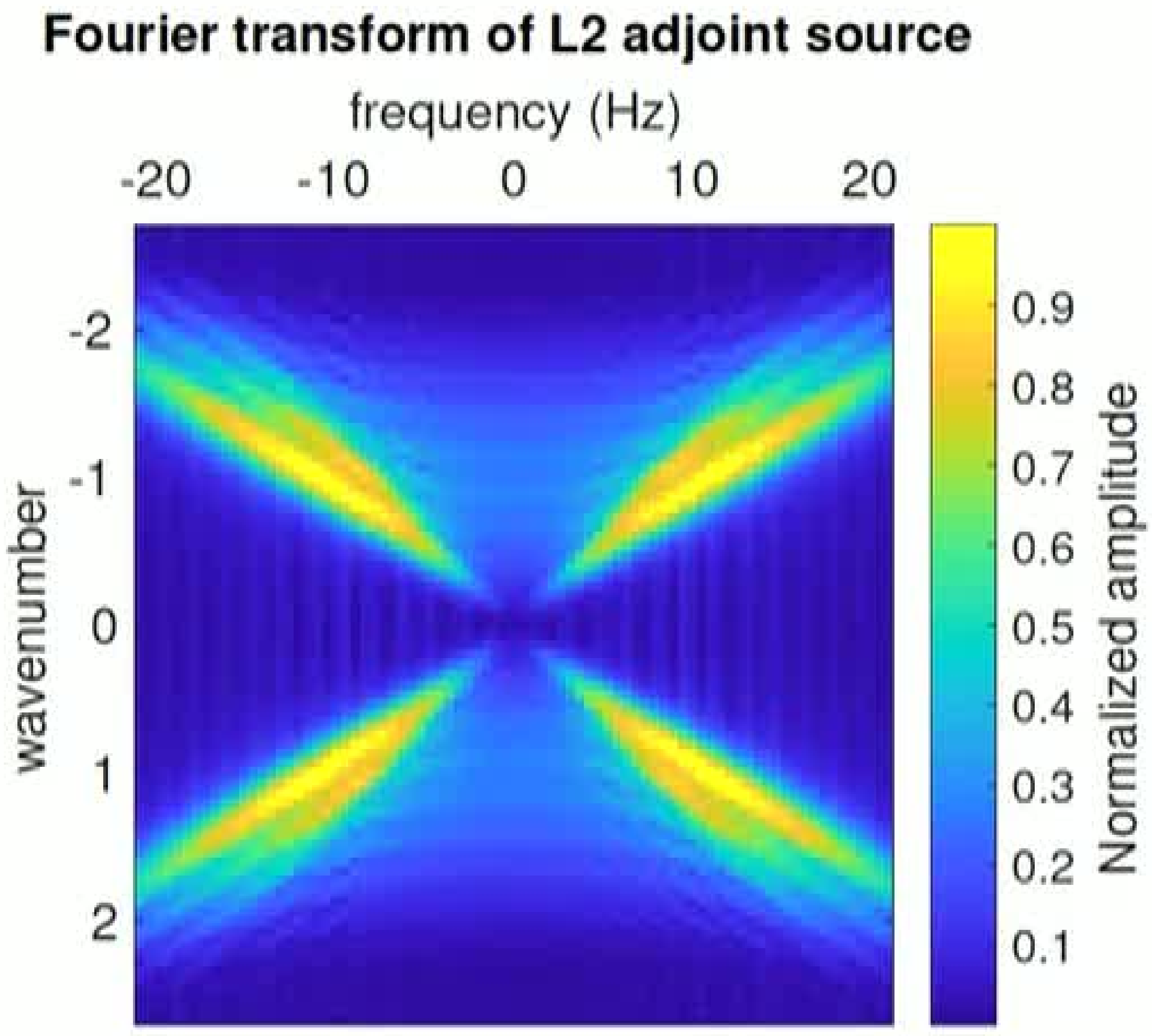
The corresponding Fréchet derivative, i.e. the adjoint source term in the backward propagation, has the following expression:

$$\frac{\partial W_2^2(f,g)}{\partial f} = \left(\int_t^{T_0} -2(s - G^{-1}(F(s)) \frac{dG^{-1}(y)}{dy}\bigg|_{y=F(s)} f(s)ds \right) dt + |t - G^{-1}(F(t))|^2 dt. \tag{5}$$

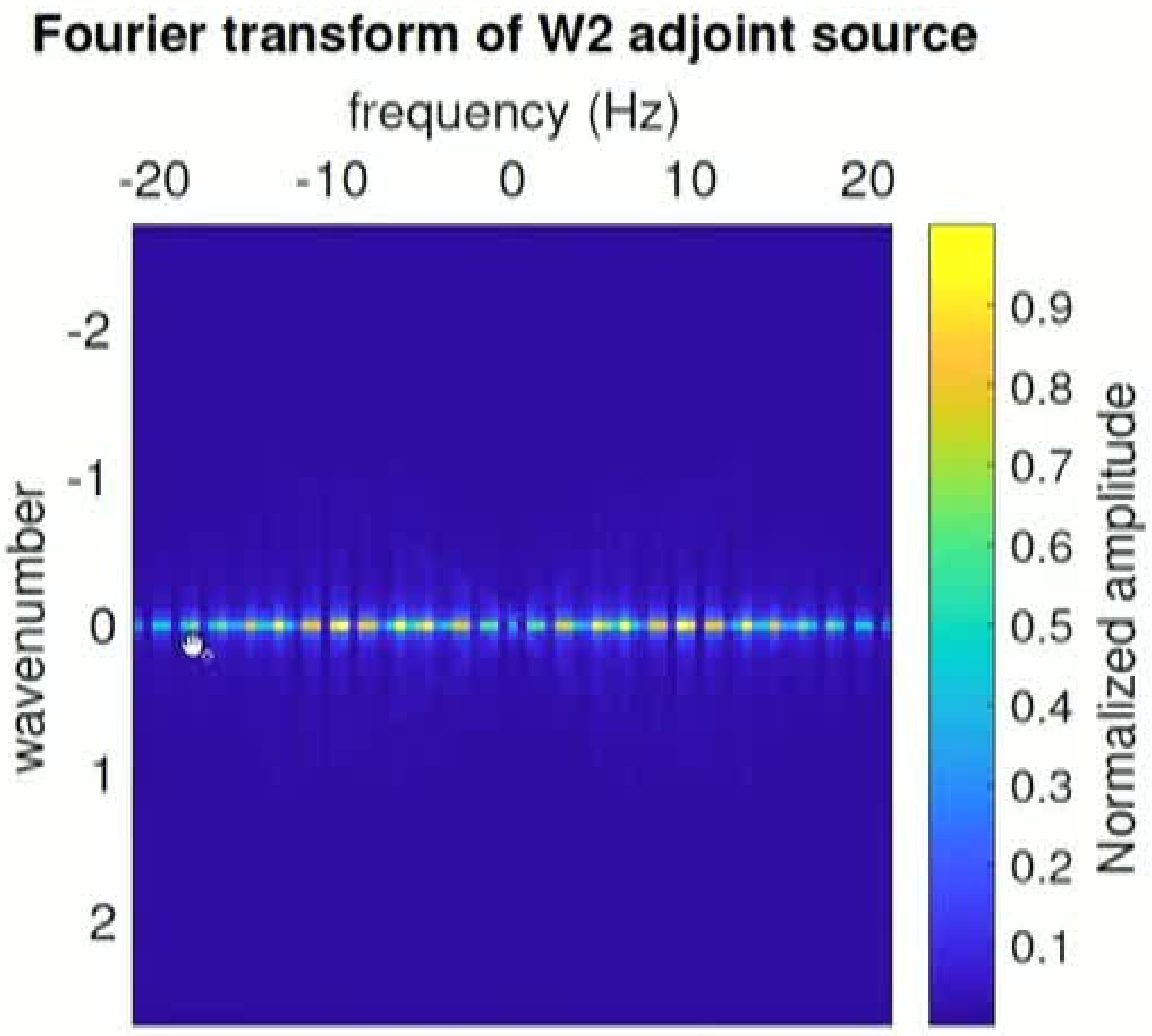
The corresponding Fréchet derivative of L^2 norm:

$$\frac{\partial L_2^2(f,g)}{\partial f} = f - g \tag{6}$$

The difficulty of L^2 norm in Reflection FWI



The Fourier transform of L^2 data gradient



The Fourier transform of W_2 data gradient

The data gradient of W_2 norm in FWI

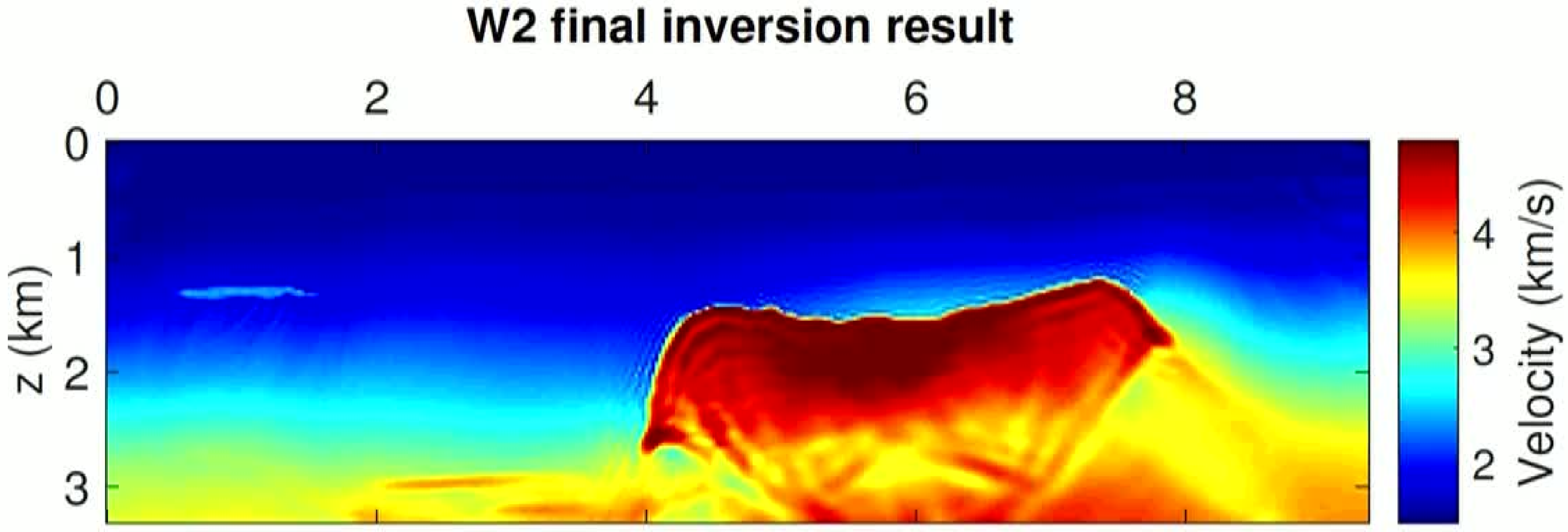
The corresponding Fréchet derivative, i.e. the adjoint source term in the backward propagation, has the following expression:

$$\frac{\partial W_2^2(f,g)}{\partial f} = \left(\int_t^{T_0} -2(s - G^{-1}(F(s)) \frac{dG^{-1}(y)}{dy}\bigg|_{y=F(s)} f(s)ds \right) dt + |t - G^{-1}(F(t))|^2 dt. \tag{5}$$

The corresponding Fréchet derivative of L^2 norm:

$$\frac{\partial L_2^2(f,g)}{\partial f} = f - g \tag{6}$$

Salt Inversion: 15 Hz Ricker



W_2 final inversion

Conclusion

Summary

Motivation

Consider the misfit globally (with an optimal map) instead of point-by-point (L^2).

Ideal properties

- Convexity;
- Insensitivity to noise;
- Working for transmission, refraction, reflection, etc.

Synthetic examples

2004 BP & Marmousi

Field data examples

Collaboration with PGS, data from the North Sea

Future work

Data normalization; Adding regularization; Multi-parameter FWI (e.g. elastic wave equation); Application to other fields, etc.