Seismic imaging and optimal transport

Bjorn Engquist

In collaboration with Brittany Froese, Sergey Fomel
and Yunan Yang

Brenier60, Calculus of Variations and Optimal Transportation,
Paris, January 10-13, 2017

Curriculum vitae of Yann Brenier, 2012

Born January 1st 1957, Saint-Chamond, France, French citizen and resident.

Centre de mathématiques Laurent Schwartz,
Ecole Polytechnique, FR-91128 Palaiseau, France
(new address, since sept. 1, 2012)

brenieratmath.polytechnique.fr

website : http : //www.math.polytechnique.fr/ brenier/

Education and positions

2012– Senior researcher (Directeur de recherches) CNRS,
Centre de mathématiques Laurent Schwartz, Ecole Polytechnique, FR-91128 Palaiseau, France

2000-2012 Senior researcher (Directeur de recherches) CNRS, Université de Nice (permanent member of CNRS since 2005, formerly on leave from U. Paris 6).

1990-2005 Professor, Université Paris 6 (including 7 years as full-time professor at Ecole Normale Supérieure, rue d'Ulm, Paris, 1990-1997).

1986-1990, Senior researcher (Directeur de recherches), INRIA, Rocquencourt.

1985-1986 J.R. Hedrick Assistant Professor, University of California (UCLA).

Doctorat ès Sciences, 1986, Université Paris-Dauphine, thesis committee: P.-A. Raviart, C. Bardos, G. Chavent, B. Engquist, J.-M. Lasry, M. Schatzman, L. Tartar.

1979-1985, junior researcher (chercheur), INRIA, Rocquencourt (including 15 months of national service, as researcher at the IIMAS-Universidad Nacional Autonoma de México, Mexico city)

Doctorat de 3ème cycle (PhD), Paris-Dauphine, 1982, advisor : Guy Chavent.

Curriculum vitae of Yann Brenier, 2012

Born January 1st 1957, Saint-Chamond, France, French citizen and resident.

Centre de mathématiques Laurent Schwartz,
Ecole Polytechnique, FR-91128 Palaiseau, France
(new address, since sept. 1, 2012)

brenieratmath.polytechnique.fr

website : http : //www.math.polytechnique.fr/ brenier/

Education and positions

2012– Senior researcher (Directeur de recherches) CNRS,
Centre de mathématiques Laurent Schwartz, Ecole Polytechnique, FR-91128 Palaiseau, France

2000-2012 Senior researcher (Directeur de recherches) CNRS, Université de Nice (permanent member of CNRS since 2005, formerly on leave from U. Paris 6).

1990-200
fessor at

1985-1986 J.R. Hedrick Assistant Professor, University of California (UCLA).

1986-1990, Senior researcher (Directeur de recherches), INRIA, Rocquencourt.

1985-1986 J.R. Hedrick Assistant Professor, University of California (UCLA).

Doctorat ès Sciences, 1986, Université Paris-Dauphine, thesis committee: P.-A. Raviart, C. Bardos, G. Chavent, B. Engquist, J.-M. Lasry, M. Schatzman, L. Tartar.

1979-1985,
15 months of
cional Auto

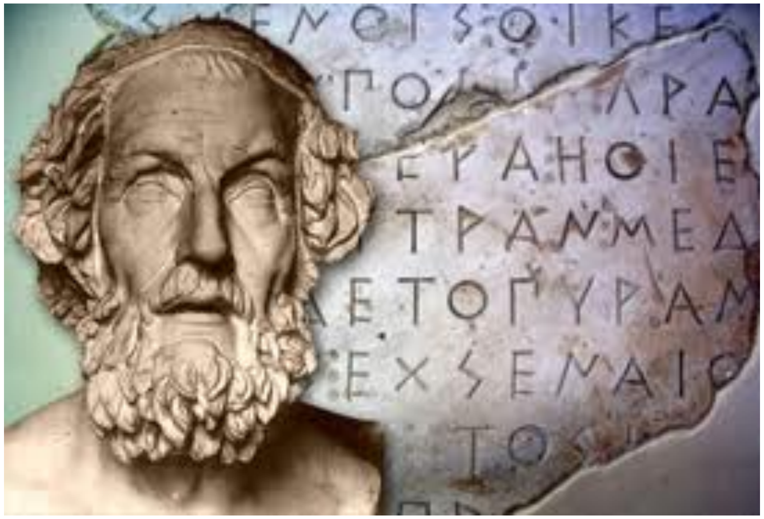
Doctorat ès Sciences, 1986, Université Paris-Dauphine, thesis committee: P.-A. Raviart, C. Bardos, G. Chavent, B. Engquist, J.-M. Lasry, M. Schatzman, L. Tartar.

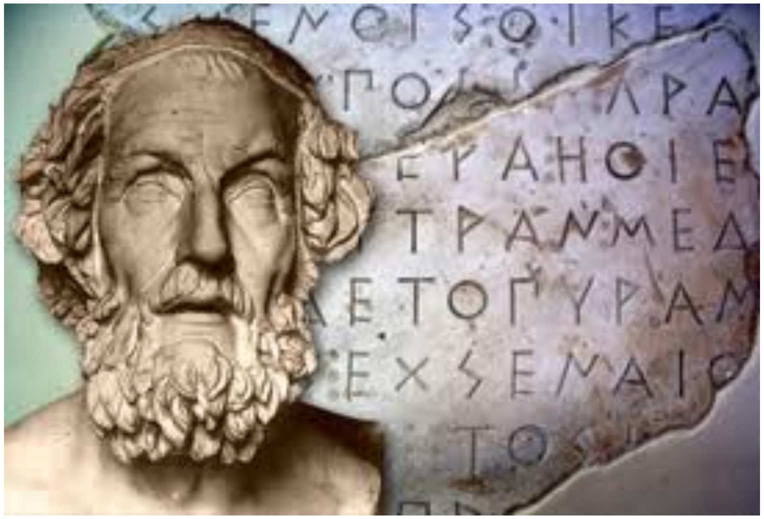
Doctorat de 3

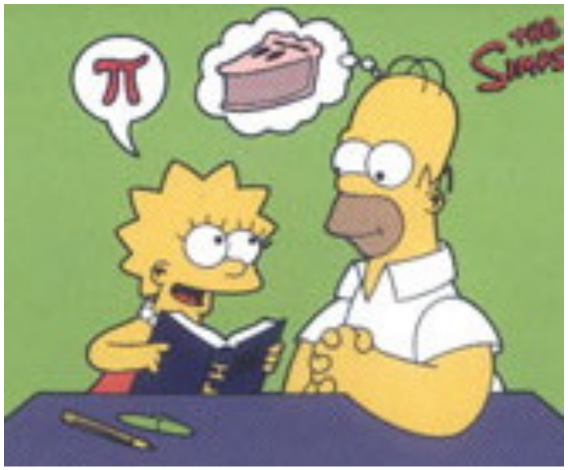
ABEL PRISEN

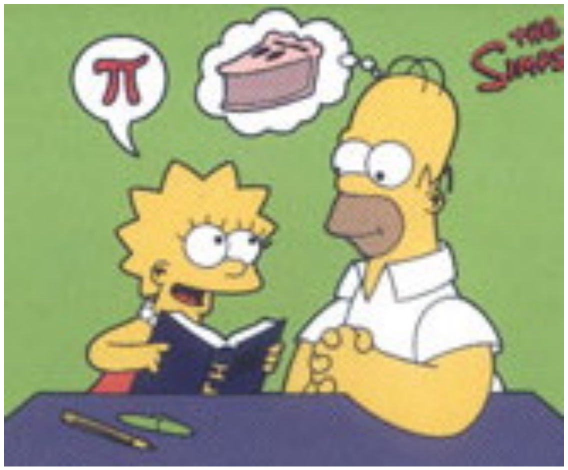
Abel Prize 2016
Andrew J Wiles

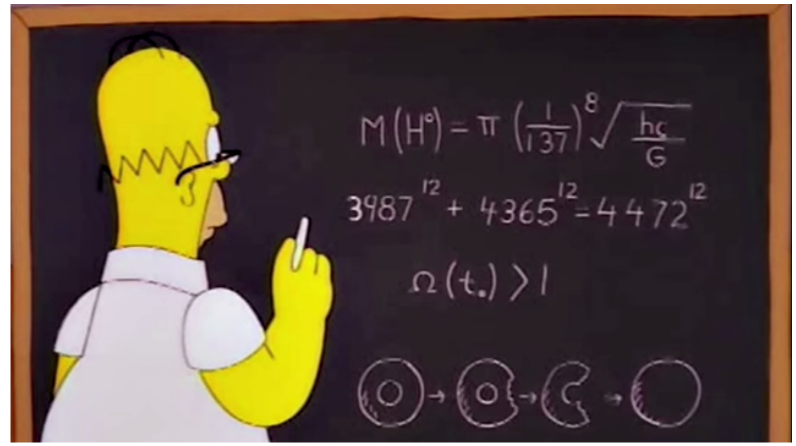
ABEL
PRISEN

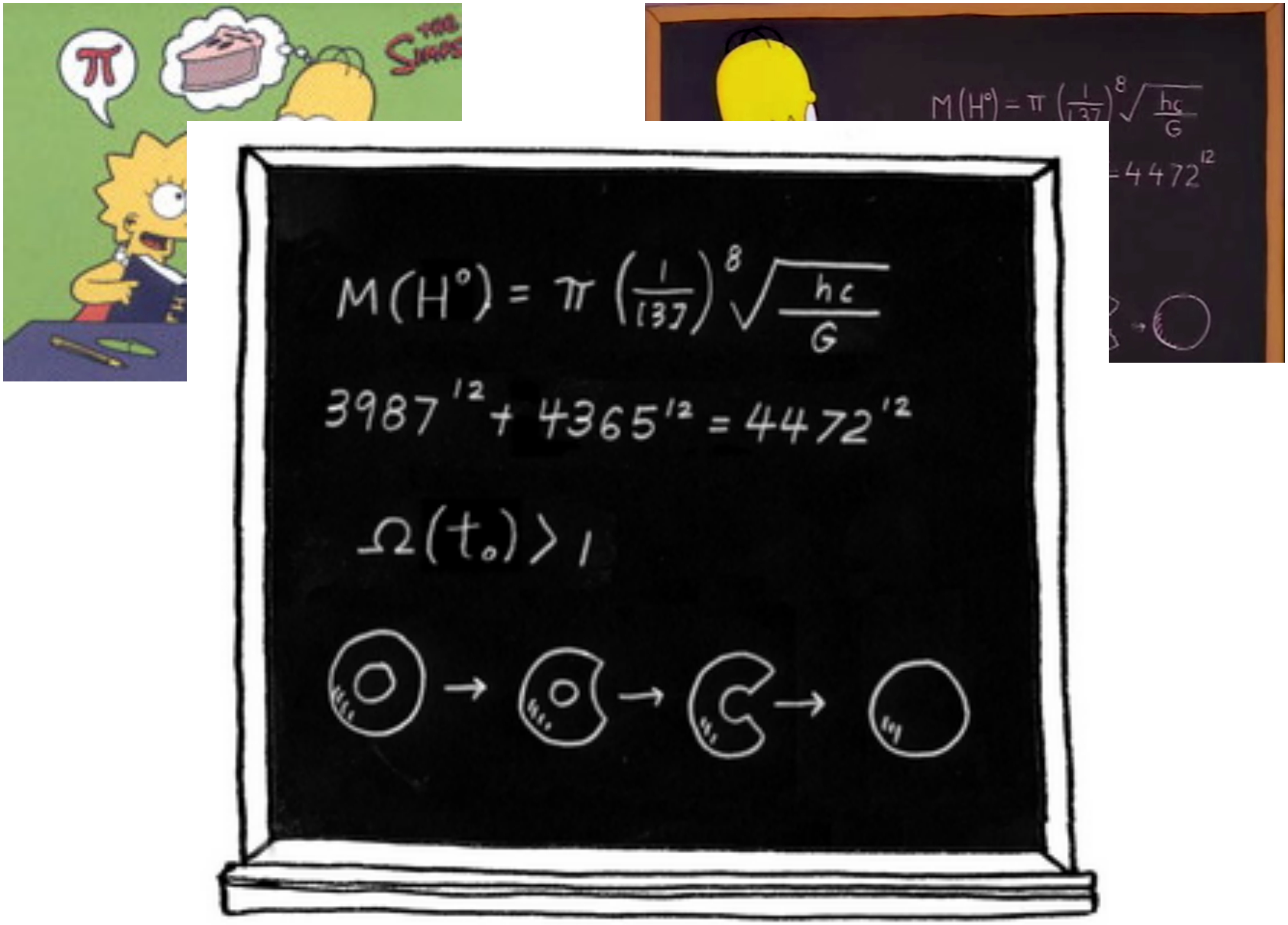





π


π

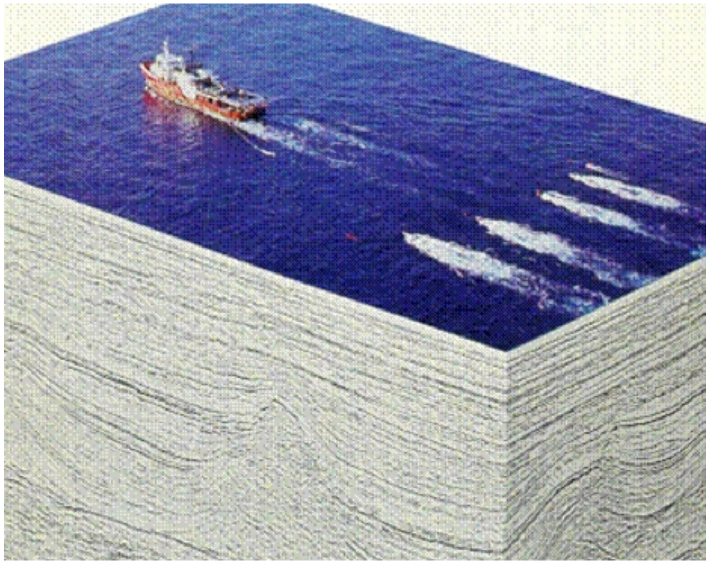

$M(H^{0}) = \pi \left(\frac{1}{137}\right)^{8} \sqrt{\frac{hc}{G}}$
$3987^{12} + 4365^{12} = 4472^{12}$
$\Omega(t_{0}) > 1$

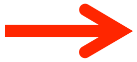

π
M(H°) = π (1/137)^8 √(hc/G)
3987^12 + 4365^12 = 4472^12
Ω(t₀) > 1

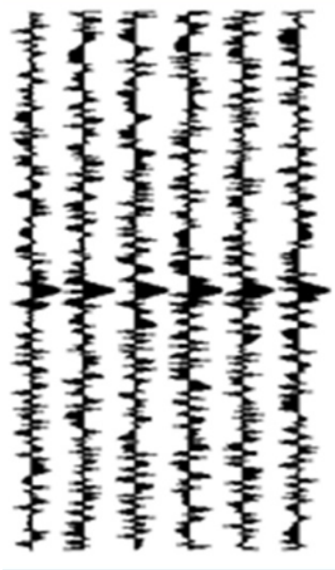
Outline

1. Remarks on seismic imaging
2. Measure of mismatch: optimal transport and the Wasserstein metric
3. Monge-Ampère equation and its numerical approximation
4. Application to full waveform inversion and registration
5. Conclusions

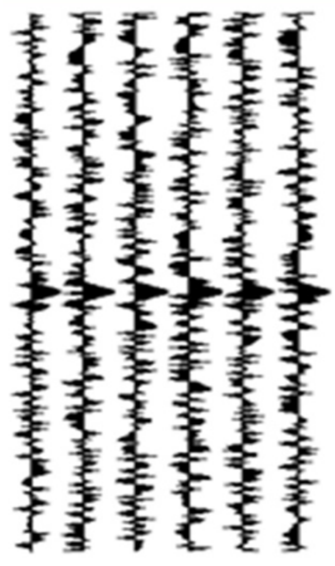
1. Remarks on seismic imaging

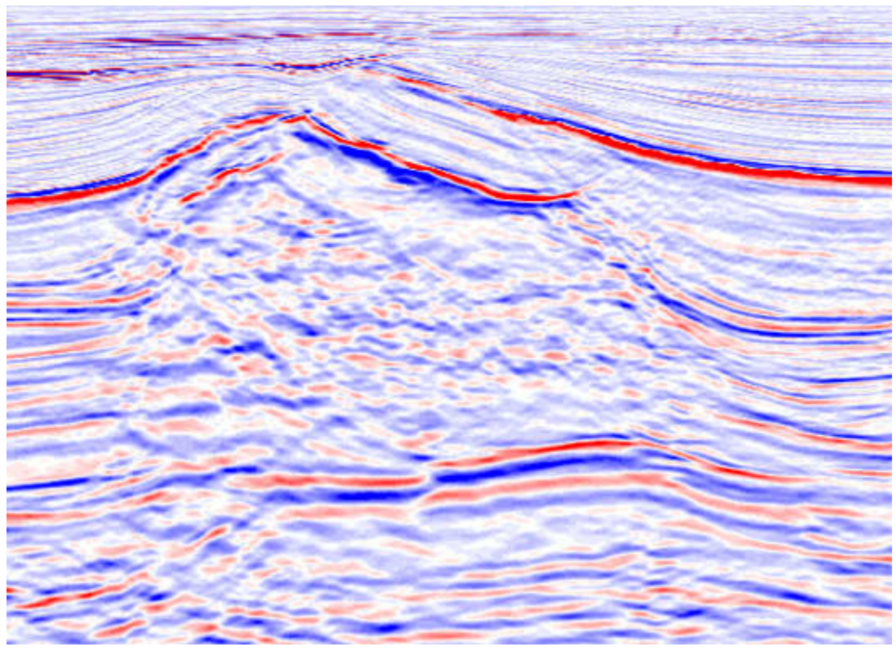






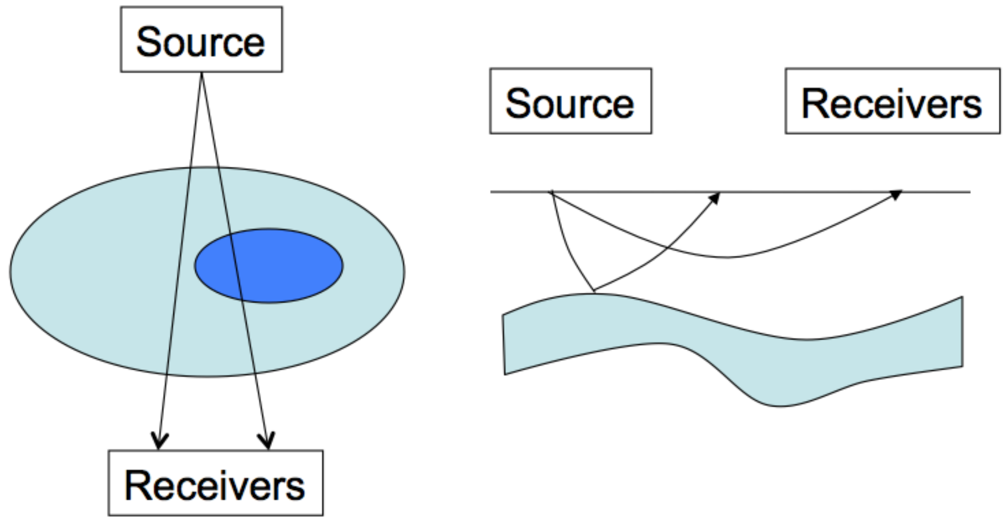
Seismic imaging





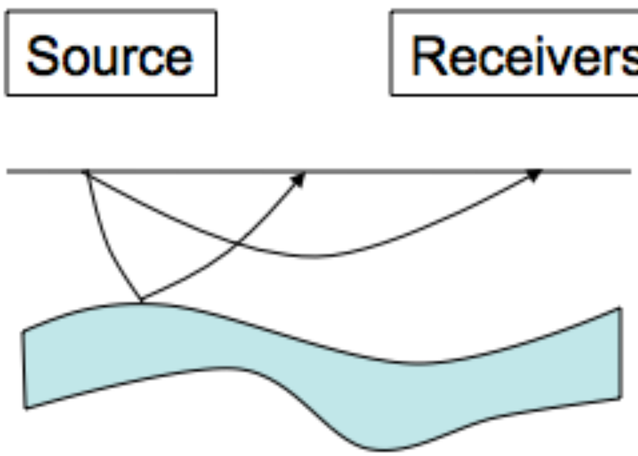
Compare tomography

- In seismic imaging no explicit formula of inverse Radon transform type (computed tomography or CT scan)



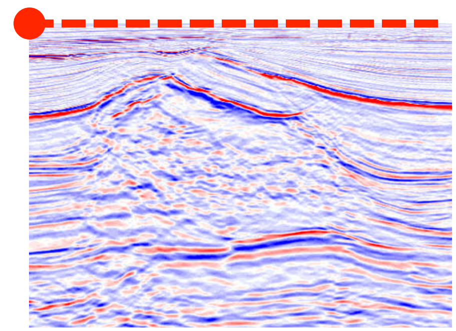
Seismic imaging

- Find seismic wave velocity and reflecting interfaces (or low and high frequency part of velocity field) separately
  - First velocity estimation
  - Then reflectivity (details too small for velocity estimation): determined by "migration"
- We will focus on the first step – velocity estimation



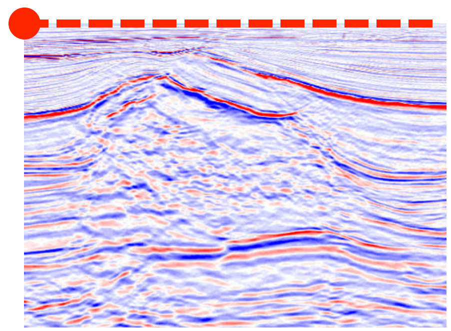
Mathematical and computational challenges

- Velocity estimation is typically done by PDE constrained optimization (classical inverse problem – compare Calderon)
  - Measured and processed data is compared to a computed wave field based on wave velocity to be determined
  - Important steps
    - Relevant measure of mismatch
    - Fast wave field solver
    - Optimization



Velocity estimation

- Velocity estimation is typically done by PDE constrained optimization.
  - Measured and processed data is compared to a computed wave field based on wave velocity to be determined
  - Important steps
    - Relevant measure of mismatch (✔)
    - Fast wave field solver
    - Optimization
- Example of forward problem: p - waves



$$p_{tt} = c(x)^2 \Delta p,$$

2. Measure of mismatch proposal: optimal transport and the Wasserstein metric

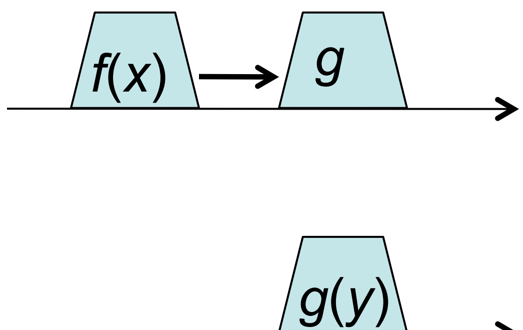
- Compare measured data to computed wave field in full waveform inversion
- In PDE-constrained optimization process: find parameters (velocity) that minimizes the mismatch

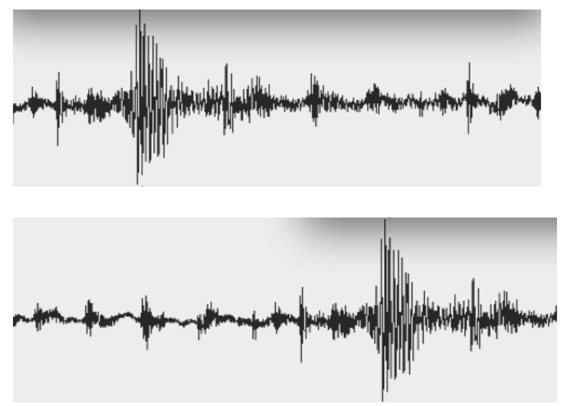
$$\min_{c(x)} \left(\left\| p_{data} - p_{comp}(c) \right\|_A + \lambda \left\| Lc \right\|_B \right)$$

- $c(x)$: velocity, u_{data} measured signal, u_{comp} computed signal based on velocity $c(x)$
- $\|\ .\ \|_A$ measure of mismatch: L_2 the standard choice
- $\|\ Lc\ \|_B$ potential regularization term (we will ignore this term, which is not common in exploration seismology

Optimal transport and Wasserstein metric

- Wasserstein metric measures the "cost" for optimally transport one measure (signal) *f* to the other, *g* – Monge-Kantorivich optimal transport measure

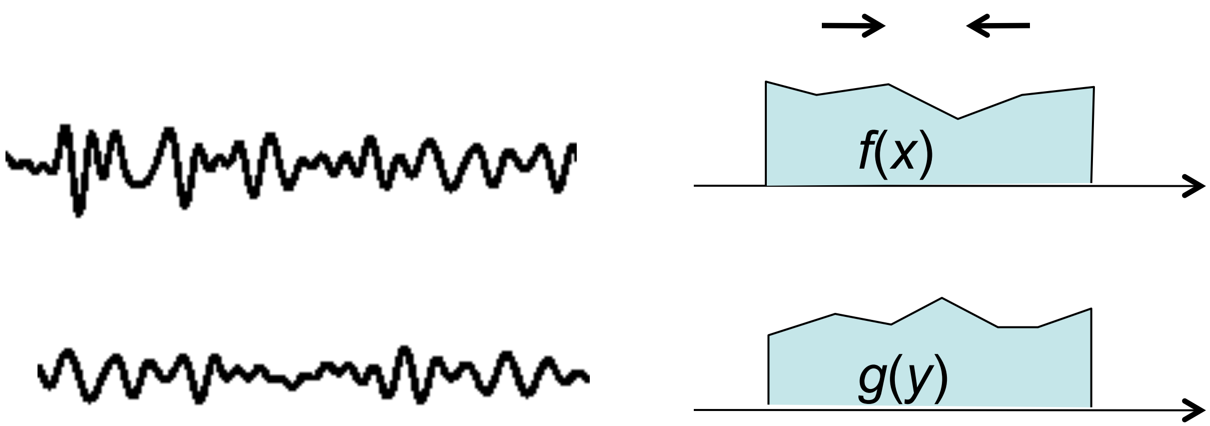




Compare travel time distance
Classic in seismology

Optimal transport and Wasserstein metric

- For some signals the "work" needed to optimally transport one distribution to the other is similar to L^p distance
- L^2 historically the standard in full waveform inversion



Wasserstein distance

$$W_p(f,g) = \left(\inf_{\gamma} \int_{X \times Y} d(x,y)^p \, d\gamma(x,y) \right)^{1/p}$$

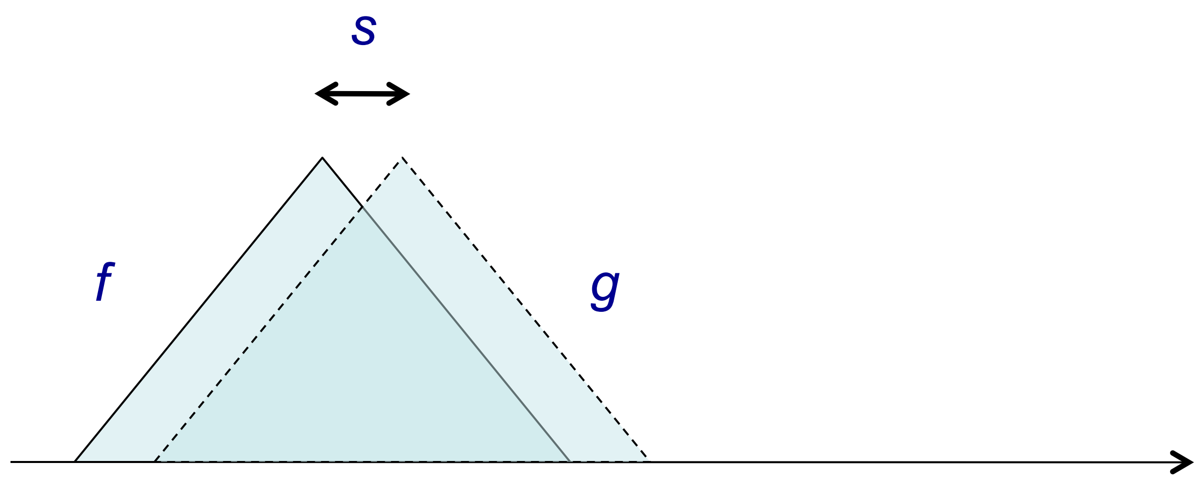
$$\gamma \in \Gamma \subset X \times Y, \textit{the set of product measure}: f \textit{ and } g$$

$$\int_X f(x)\,dx = \int_Y g(y)\,dy, \quad f,\ g \geq 0$$

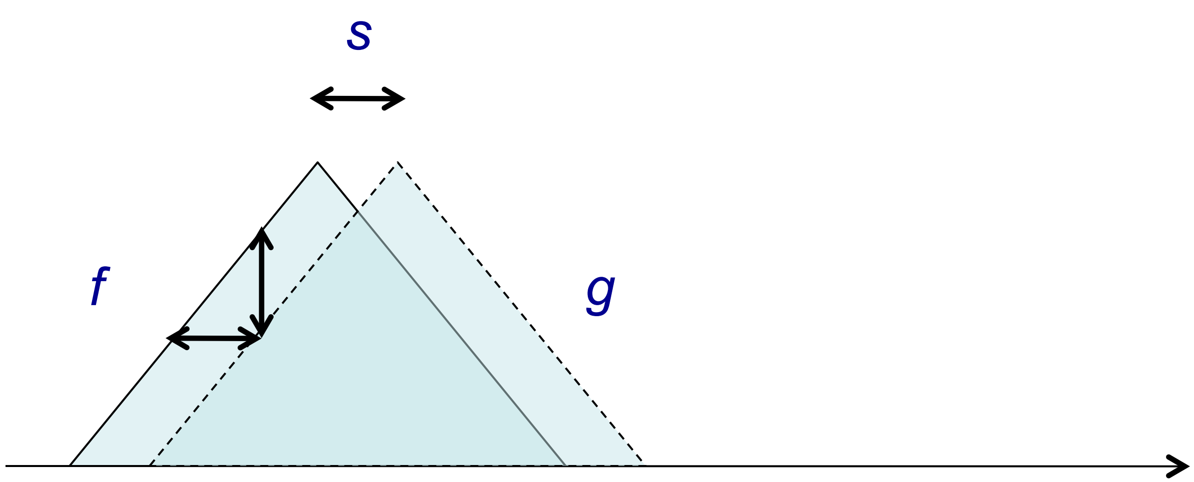
$$W_2(f,g) = \left(\inf_T \int_X \|x - T(x)\|_2^2 f(x)\,dx \right)^{1/2}$$

- Here T is the optimal transport map from f to g

Wasserstein distance

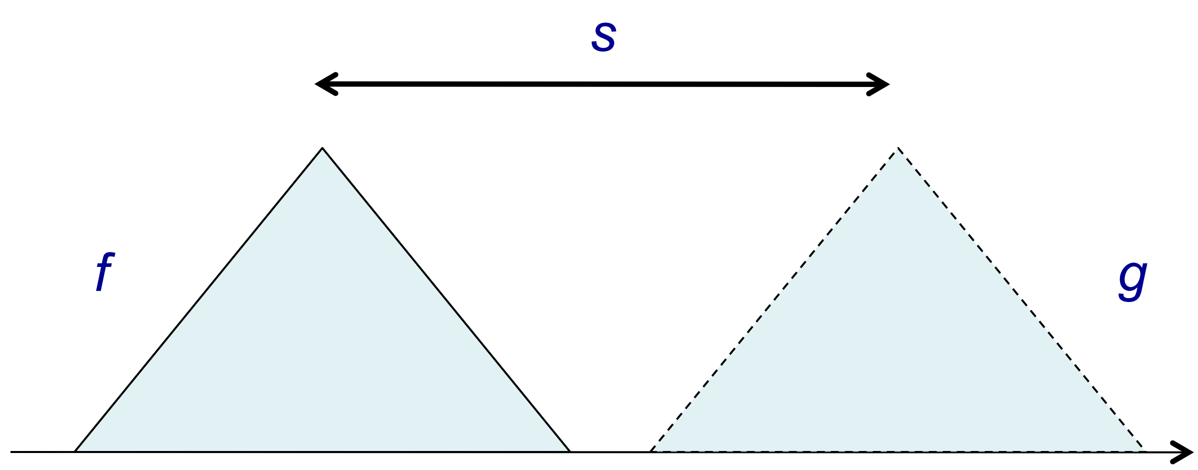


Wasserstein distance



- In this model example W_2 and L_2 is equal (modulo a constant) to leading order when separation distance *s* is small. Recall L_2 is the standard measure

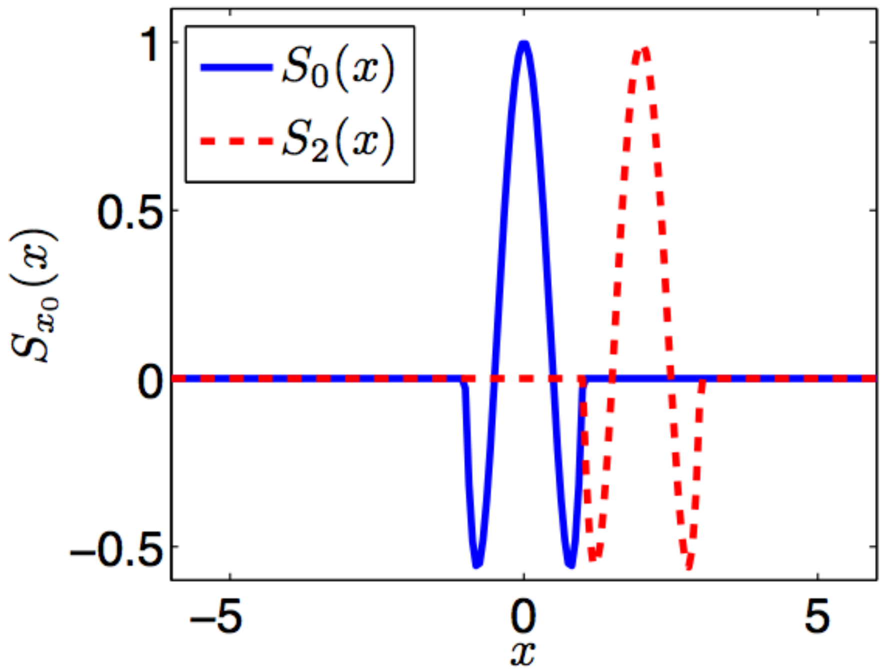
Wasserstein distance



- When s is large $W_2 = s$ = travel distance (time), (“higher frequency”), L_2 independent of s

Wasserstein distance vs L_2

- Fidelity measure, single seislet or Ricker wavelet

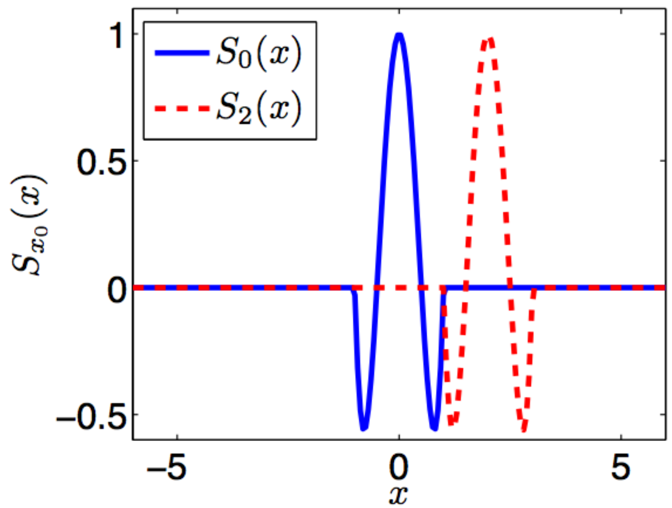


Wasserstein distance vs L_2

- Note that "shift" and also "dilation" are natural effects of difference in velocity c.

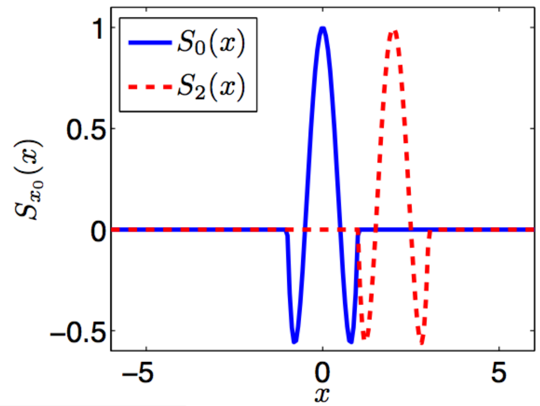
$$p_{tt} = c^2 p_{xx}, \quad x > 0, t > 0$$
$$p(0,t) = s(t) \rightarrow p = s(t - x/c)$$

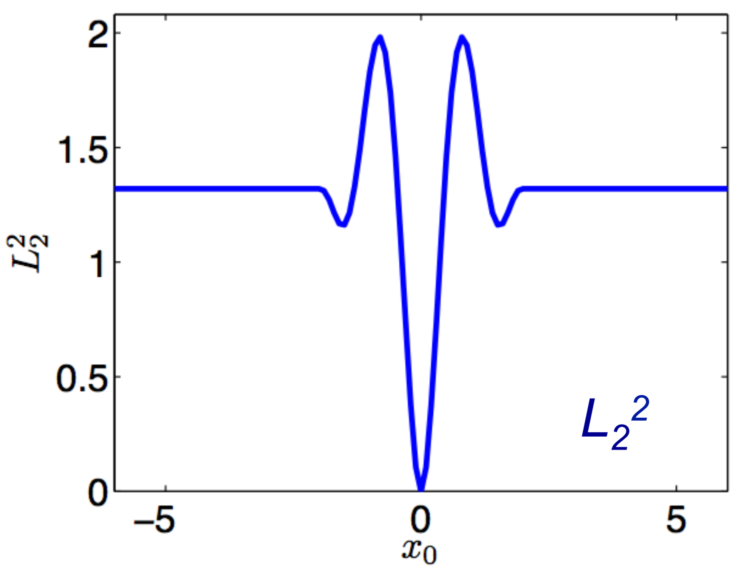
- Shift as a function of *t*, dilation as a function of *x*
- Natural effect of mismatch in velocity



Wasserstein distance vs L_2

- Fidelity measure



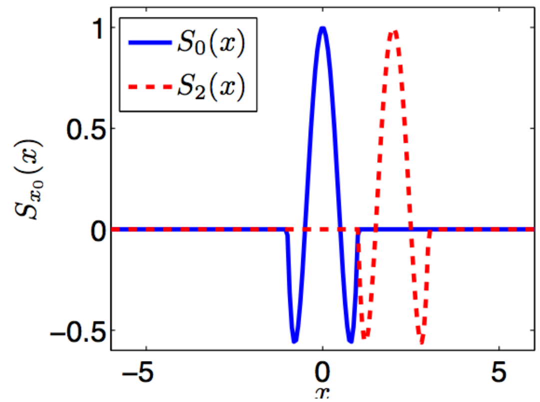


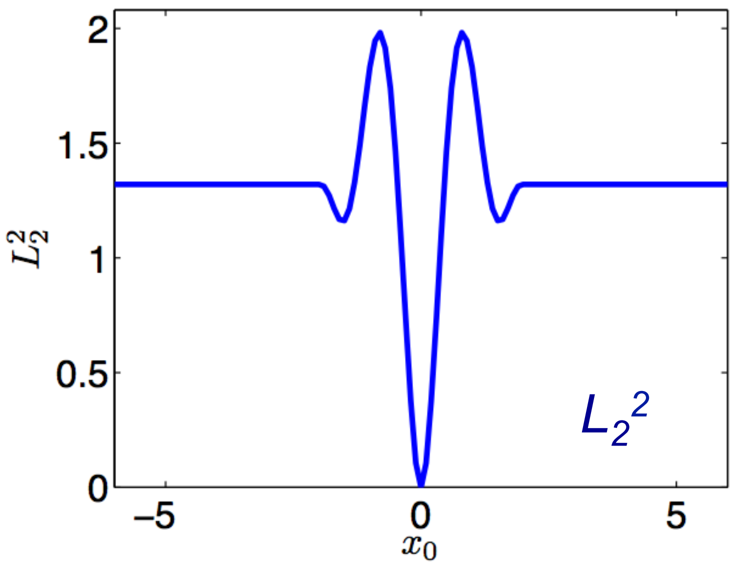
"Cycle skipping"
Local minima

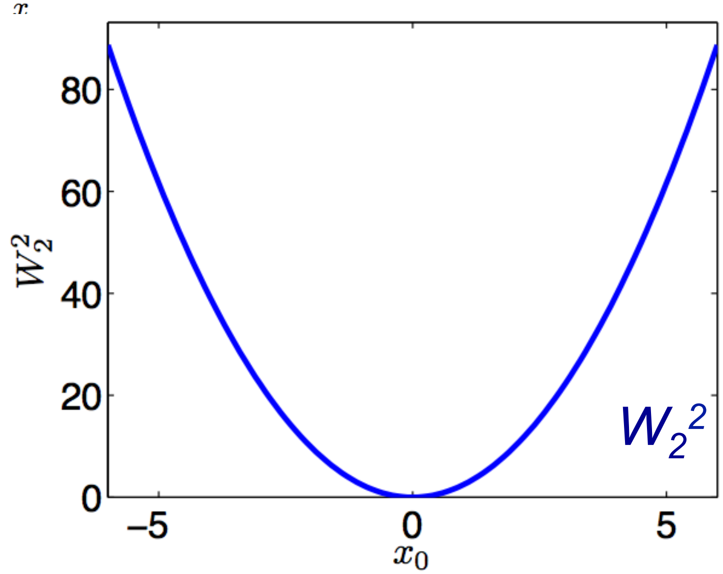
Function of displacement

Wasserstein distance vs L_2

- Fidelity measure

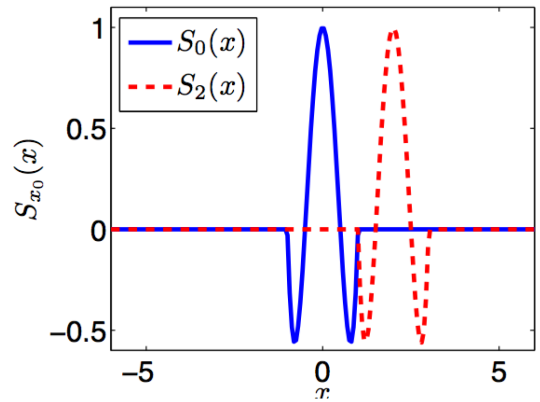


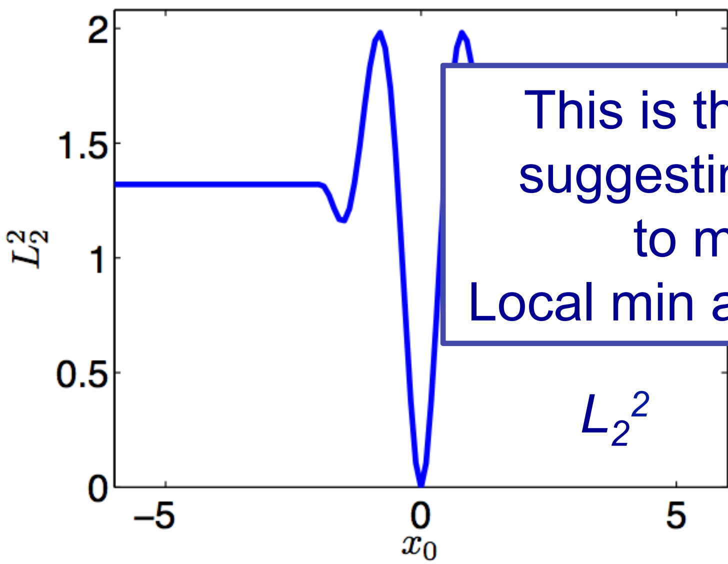


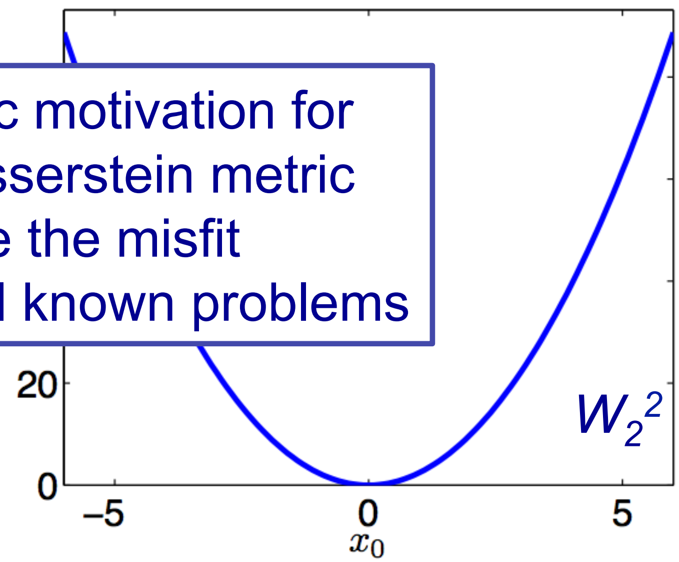


Wasserstein distance vs L_2

- Fidelity measure



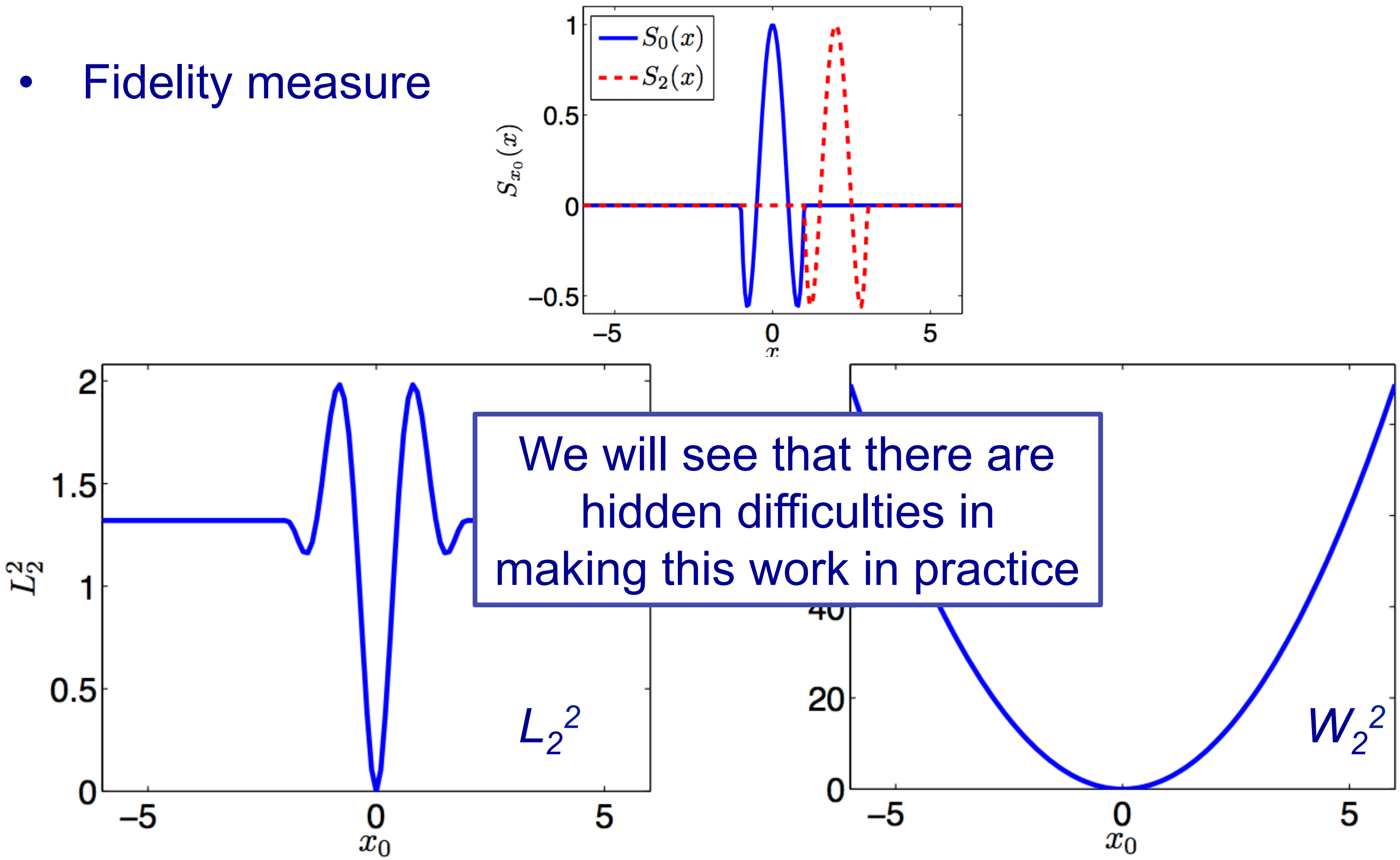




This is the basic motivation for suggesting Wasserstein metric to measure the misfit
Local min are well known problems

Wasserstein distance vs L_2

- Fidelity measure



We will see that there are hidden difficulties in making this work in practice

Analysis

- Theorem 1: $W_2^{\ 2}$ is convex with respect to translation, *s* and dilation, *a,*

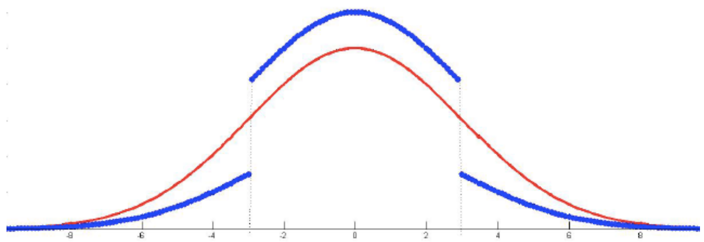
$$W_2^2(f,g)[\alpha,s],\ \ f(x)=g(\alpha x-s)\alpha^d,\ \ \alpha>0,\ x,\ s\in R^d$$

- Theorem 2: $W_2^{\ 2}$ is convex with respect to local amplitude change, *λ*

$$W_2^2(f,g)[\beta],\ \ f(x)=\begin{cases} g(x)\lambda,\ \ x\in\Omega_1 \\ \beta g(x)\lambda,\ \ x\in\Omega_2 \end{cases}\quad \beta\in R,\ \ \Omega=\Omega_1\cup\Omega_2$$

$$\lambda=\int_\Omega g\,dx\Big/\left(\int_{\Omega_1} g\,dx+\beta\int_{\Omega_2} g\,dx\right)$$

- (L_2 only satisfies 2nd theorem)



Remarks

- The scalar dilation *ax* can be generalized to *Ax* where *A* is a positive definite matrix. Convexity is then in terms of the eigenvalues
- The proof of theorem 1 is based on c-cyclic monotonicity

$$\left\{\left(x_j, x_j\right)\right\} \in \Gamma \rightarrow \sum_j c\left(x_j, x_j\right) \leq \sum_j c\left(x_j, x_{\sigma(j)}\right)$$

- The proof of theorem two is based on the inequality

$$W_2^2(sf_1 + (1-s)f_2, g) \leq sW_2^2(f_1, g) + (1-s)W_2^2(f_2, g)$$

Illustration: discrete proof (theorem 1)

- Equal point masses then weak limit
- Brenier: back of the envelope for laymen at Banff

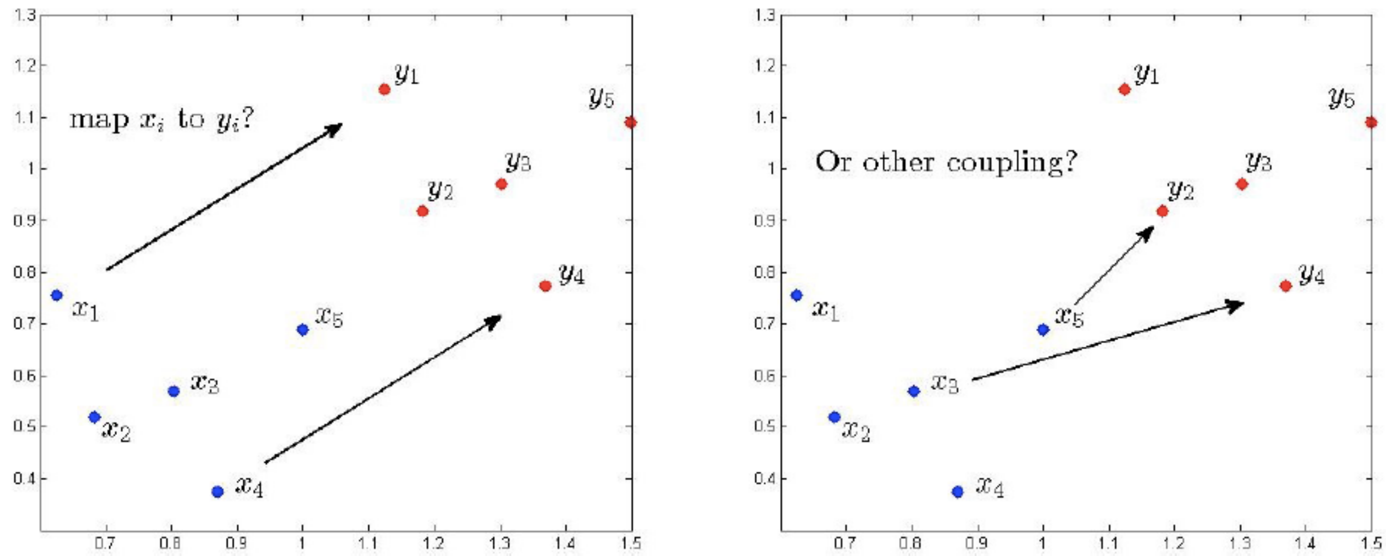


Illustration: discrete proof

$$W_2^2 = \min_{\sigma} \sum_{j=1}^{J} \left| x_{o_j} - (x_j - s\xi) \right|^2 = \quad (\sigma : \ permutation)$$

$$\min_{\sigma} \left(\sum_{j=1}^{J} \left| x_{o_j} - x_j \right|^2 - 2s \sum_{j=1}^{J} \left(x_{o_j} - x_j \right) \cdot \xi + J \left| s\xi \right|^2 \right) =$$

$$\min_{\sigma} \left(\sum_{j=1}^{J} \left| x_{o_j} - x_j \right|^2 + J \left| s\xi \right|^2 \right), \quad from \ \sum_{j=1}^{J} x_{o_j} = \sum_{j=1}^{J} x_j$$

$$\rightarrow \ x_{o_j} = x_j \rightarrow \sigma_j = j$$

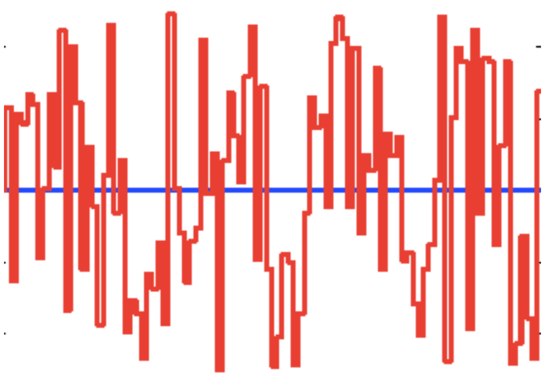
Noise

- W_2^2 less sensitive to noise than L_2
- Theorem 3: $f = g + \delta$, δ uniformly distributed uncorrelated random noise, ($f > 0$), discrete i.e. piecewise constant: N intervals

$$\|f-g\|_{L_2}^2 = O\ (1), \quad W_2^2\left(f-g\right) = O(N^{-1})$$

$$f = \left(f_1, f_2, \ldots, f_J\right)$$

- Proof by "domain decomposition" dimension by dimension and standard deviation estimates using closed 1D formula



3. Monge-Ampère equation and its numerical approximation

- In 1D, optimal transport is equivalent to sorting with efficient numerical algorithms O($N\log(N)$) complexity, N data points
- In higher dimensions such combinatorial methods as the Hungarian algorithm are very costly O(N^3), Alternatives: linear programming, sliced Wasserstein, ADMM
- Fortunately the optimal transport related to W_2 can be solved via a Monge-Ampère equation [Brenier,..]

$$W_2(f,g) = \left(\int_X \|x - \nabla u(x)\|_2^2 f(x)\,dx \right)^{1/2}$$

$$\det\left(D^2(u)\right) = f(x) / g(\nabla u(x))$$

Monge-Ampère equation

- Viscosity solution *u* if *u* is both a sub and super solution

$$\det\left(D^2(u)\right) - f(x) = 0,\ u\ convex,\ \ f \in C^0(\Omega)$$

- Sub solution (super analogous)

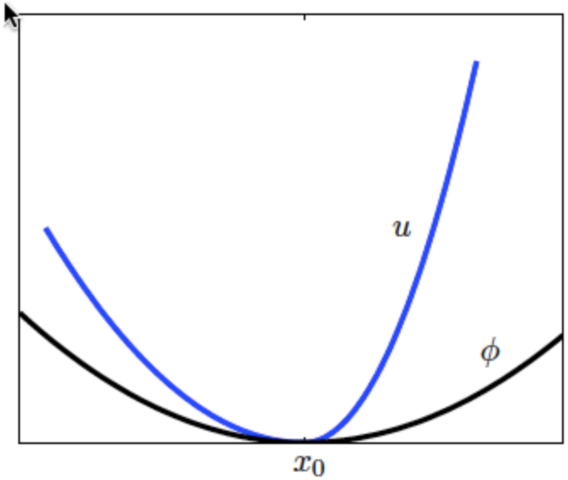
$$x_0 \in \Omega,\ if\ local \max\ of\ u - \phi,\ then$$

$$\det\left(D^2\phi\right) \le f(x_0)$$

- 1D

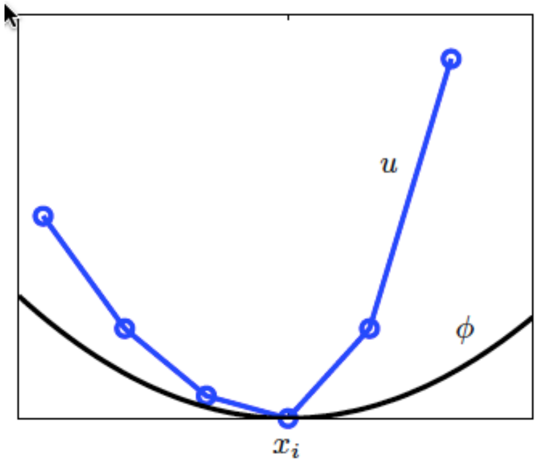
$$u_{xx} = f,\ \ \phi(x_0) = u(x_0), \phi(x_0) = u(x_0),$$

$$\phi(x) \le u(x) \rightarrow \phi_{xx} \le f$$



Numerical approximation

- Consistent, stable and monotone finite difference approximations will converge to Monge-Ampère viscosity solutions [Barles, Souganidis, 1991]

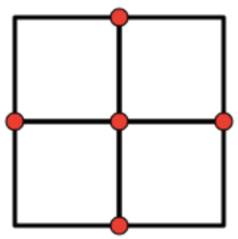


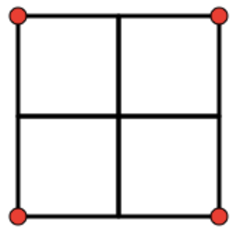
Numerical approximation

- Example, monotone scheme following [Benamou, Froese, Oberman, 2014], discrete maximum principle

$$u_{xx} \approx \left(u_{j+1,k} - 2u_{j,k} + u_{j-1,k}\right) / \Delta x^2 \quad \textit{monotone}$$

$$u_{xv} \approx \left(u_{j+1,k+1} + u_{j-1,k-1} - u_{j+1,k-1} - u_{j-1,k+1}\right) / 4\Delta x \Delta y \quad \textit{not monotone}$$

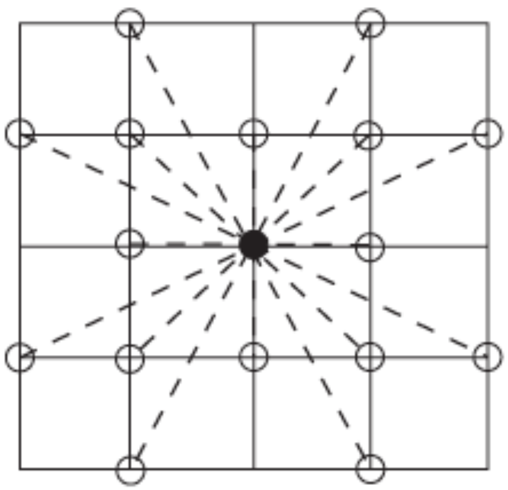




Monotone approximation

$$\det\left(D^2u\right) = \prod_{j=1}^{d}\left(u_{v_jv_j}\right)^+, \ \{v_j\}: set\ of\ eigenvectors\ of\ D^2u$$

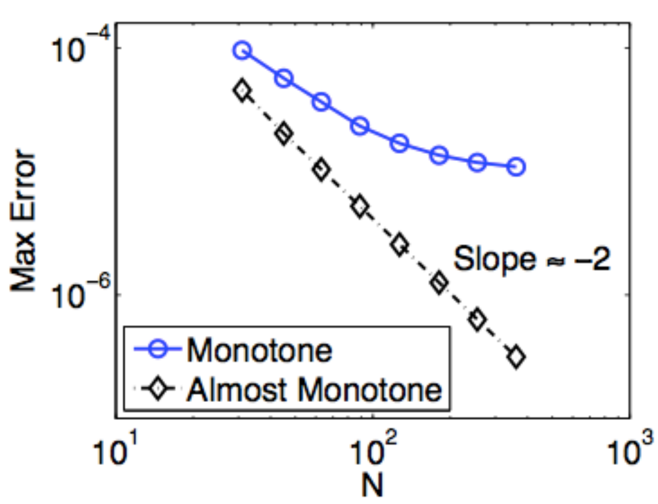
$$D_{vv} \approx \left(u(x+vh) - 2u(x) + u(x-vh)\right)/\left|vh\right|^2$$



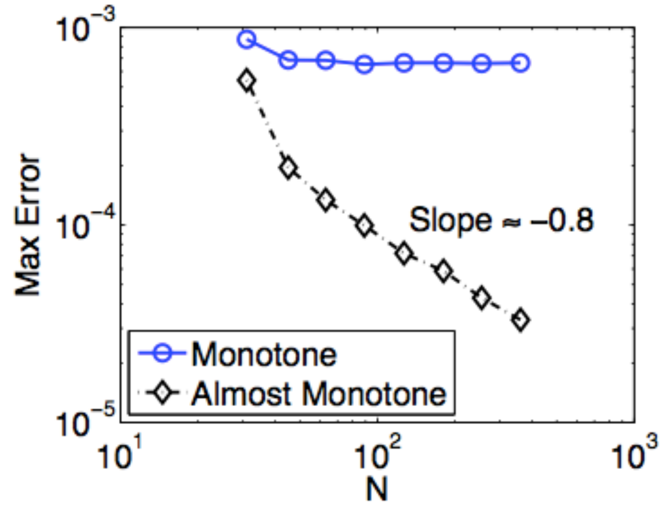
- Compare upwind or ENO adaptive stencils and limiters for nonlinear conservation laws
- WENO style smooth superposition improves Newton convergence
- MG improves linear solver

Numerical approximation

- Final algorithm with filter, almost monotone for higher accuracy (still converging)
- Newton's method for discretized nonlinear problem – added regularization in choice of stencil and limiters



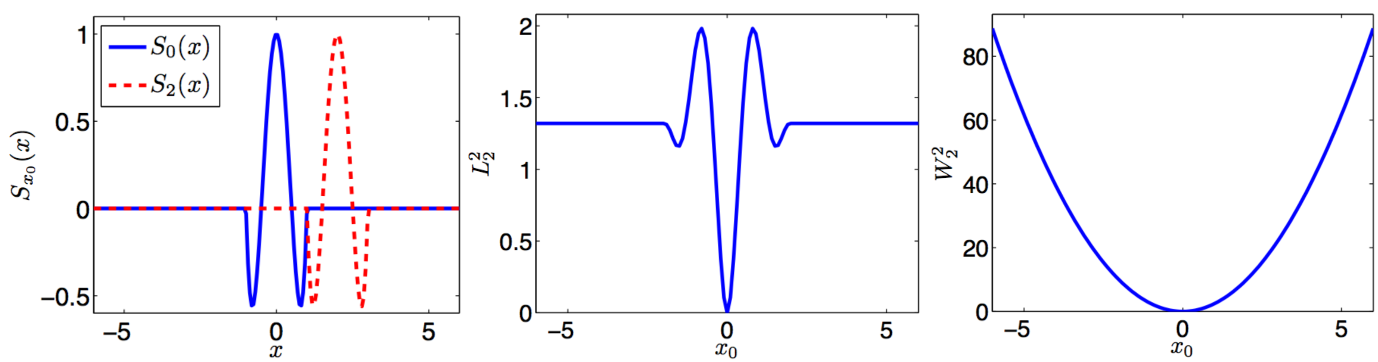
C^2 Example



C^1 Example

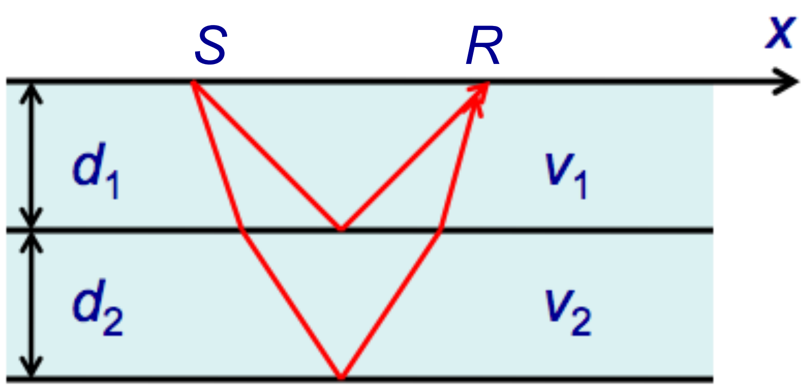
4. Applications to waveform inversion and registration

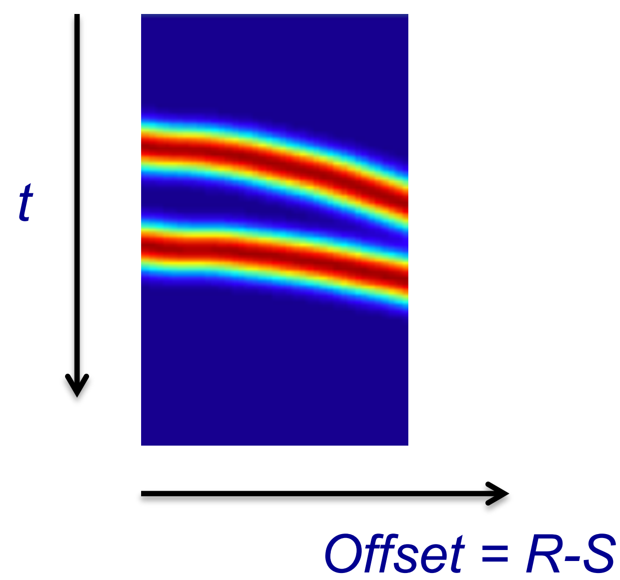
- Two natural seismic applications for optimal transport and Monge-Ampère
  - Measure of mismatch in the inverse problem of finding velocity: full waveform inversion
  - Registration: comparing different datasets
- Convexity relevant property



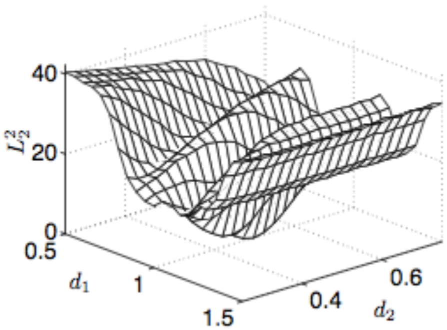
Reflections and inversion example

- Problem with reflection from two layers – dependence on parameters

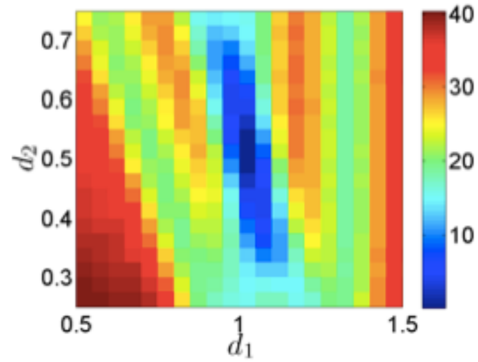




Reflections and inversion example

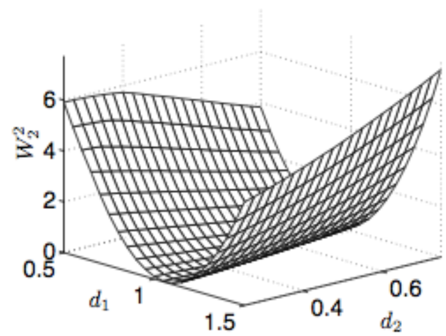


(e)

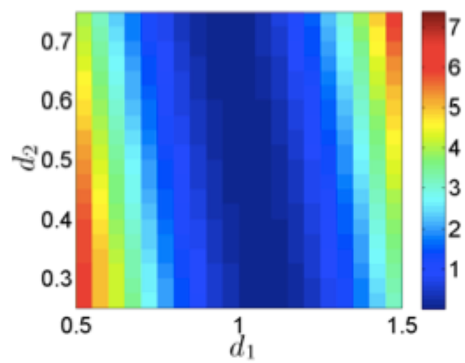


(f)

L_2



(g)



(h)

W_2

Gradient for optimization

- For large scale optimization, gradient of $J(f) = W_2^2(f,g)$ with respect to wave velocity is required in a quasi Newton method in the PDE constrained optimization step
- Based on linearization of J and Monge-Ampère equation resulting in linear elliptic PDE (adjoint source)

$$J + \delta J = \int (f + \delta f)\left\| x - \nabla(u_f + \delta u)\right\|^2 dx$$

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

$$L(v) = g(\nabla u_f)tr((D^2 u_f)^{\bullet} D^2(v)) + \det(D^2 u_f)g(\nabla u_f)\cdot \nabla v = \delta f$$

Remarks

+ Captures important features of distance in both travel time and L_2
+ There exists fast algorithms
+ Robust vs. noise
- Constraints that are not natural for seismology

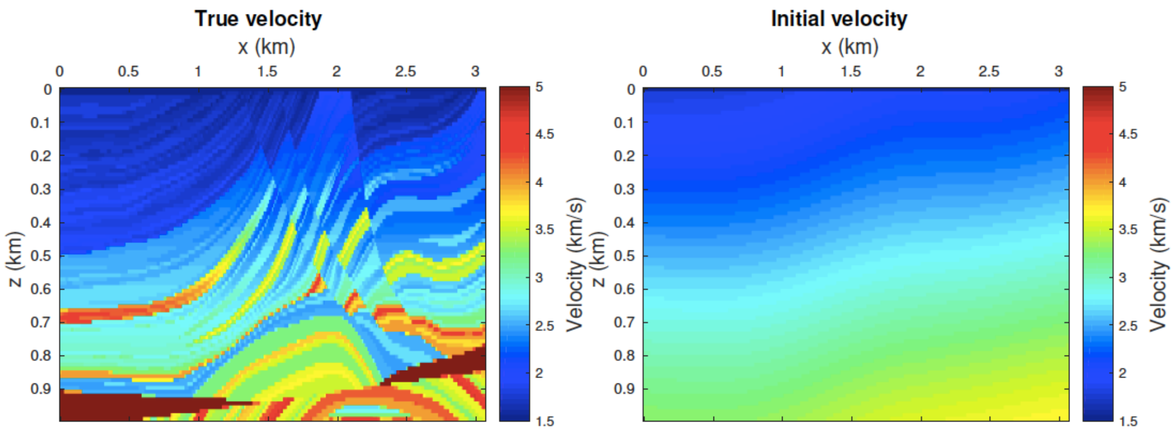
$$\int_X f(x)\,dx = \int_Y g(y)\,dy, \quad f,\ g \geq 0$$

$$g > 0, \ \textit{convex domain}$$

- Consider positive and negative parts of *f* and *g* separately and (regularize) – not appropriate for adjoin field gradient technique
- Normalize and regularize: add small constant where $g = 0$
- Alternative W_2: W_1 trace by trace W_2 (1D)

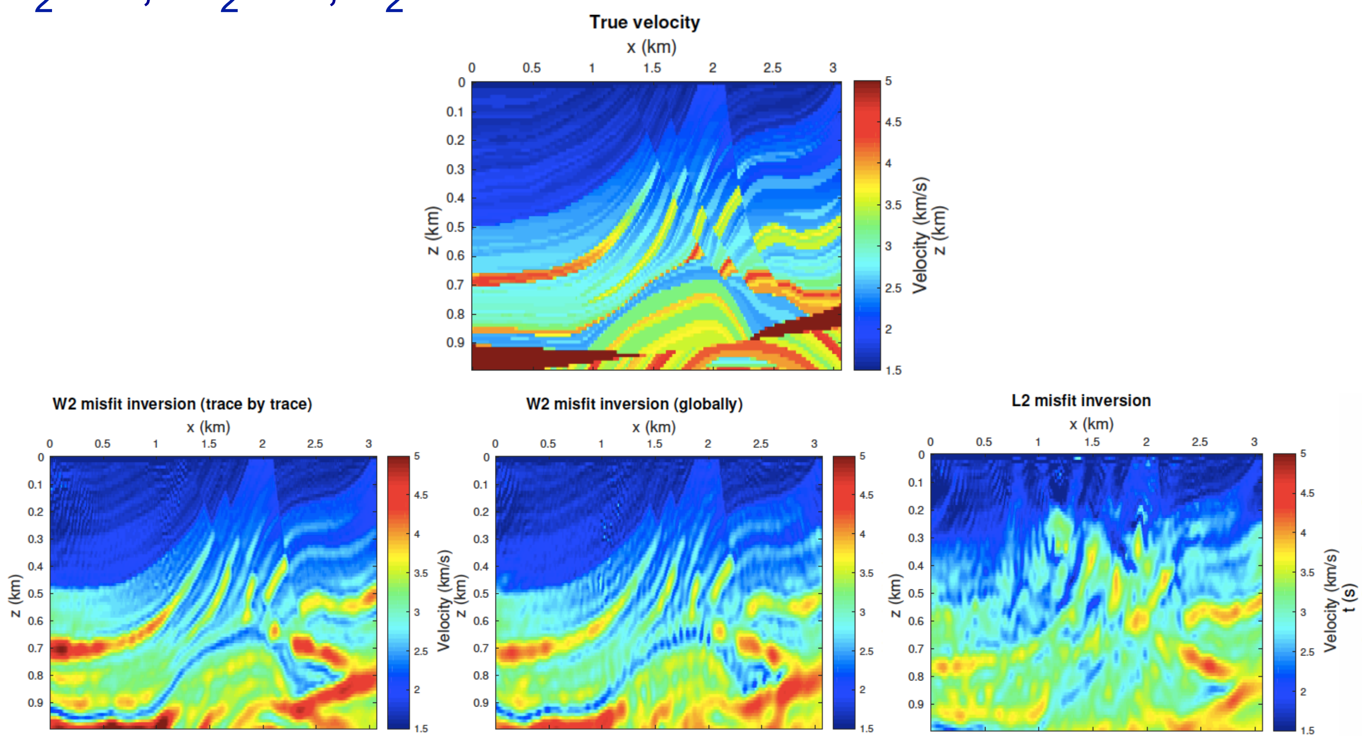
Applications Seismic test cases

- Marmousi model (velocity field)
- Original model and initial velocity field to start optimization



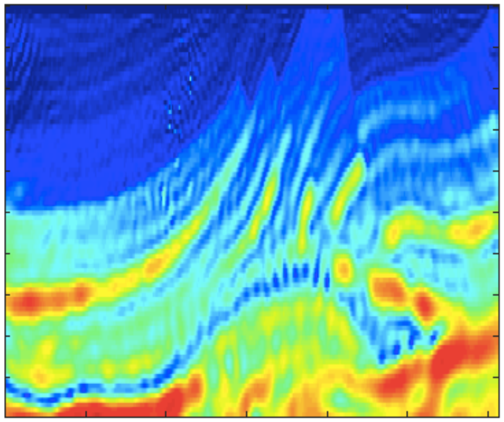
Marmousi model

- Original and FWI reconstruction with different initializations: W_2-1D, W_2-2D, L_2



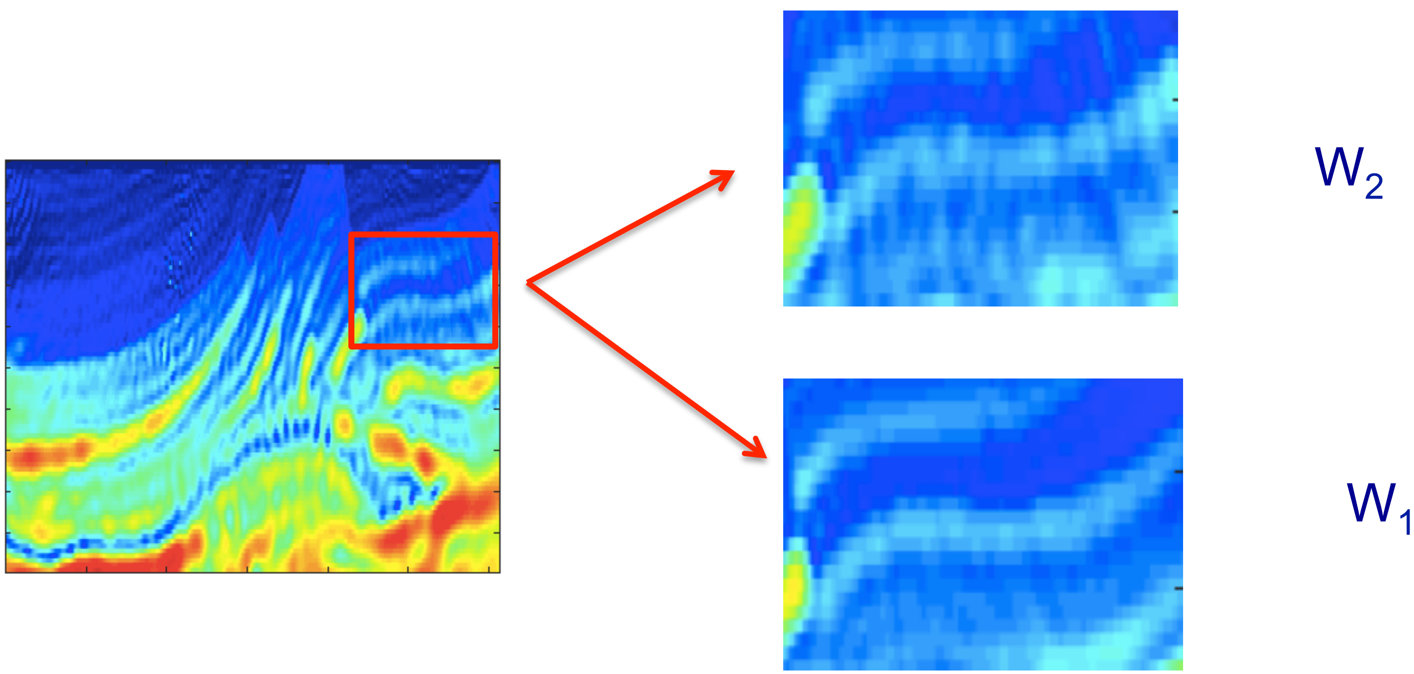
Remark

- Robustness to noise: good for data but allows for oscillations in "optimal" computed velocity
- Remedy: trace by trace, TV - regularization



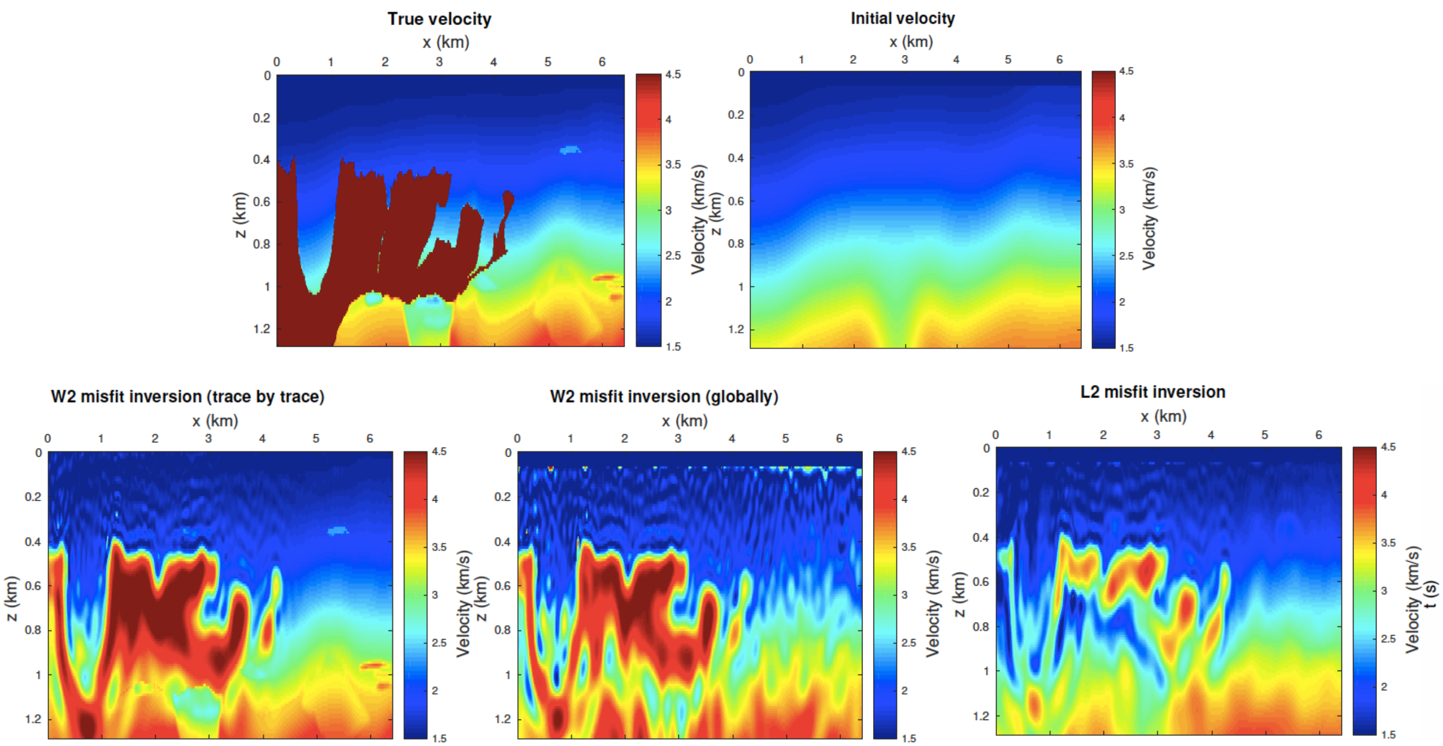
Remark

- Robustness to noise: good for data but allows for oscillations in "optimal" computed velocity
- Remedy: trace by trace, TV - regularization

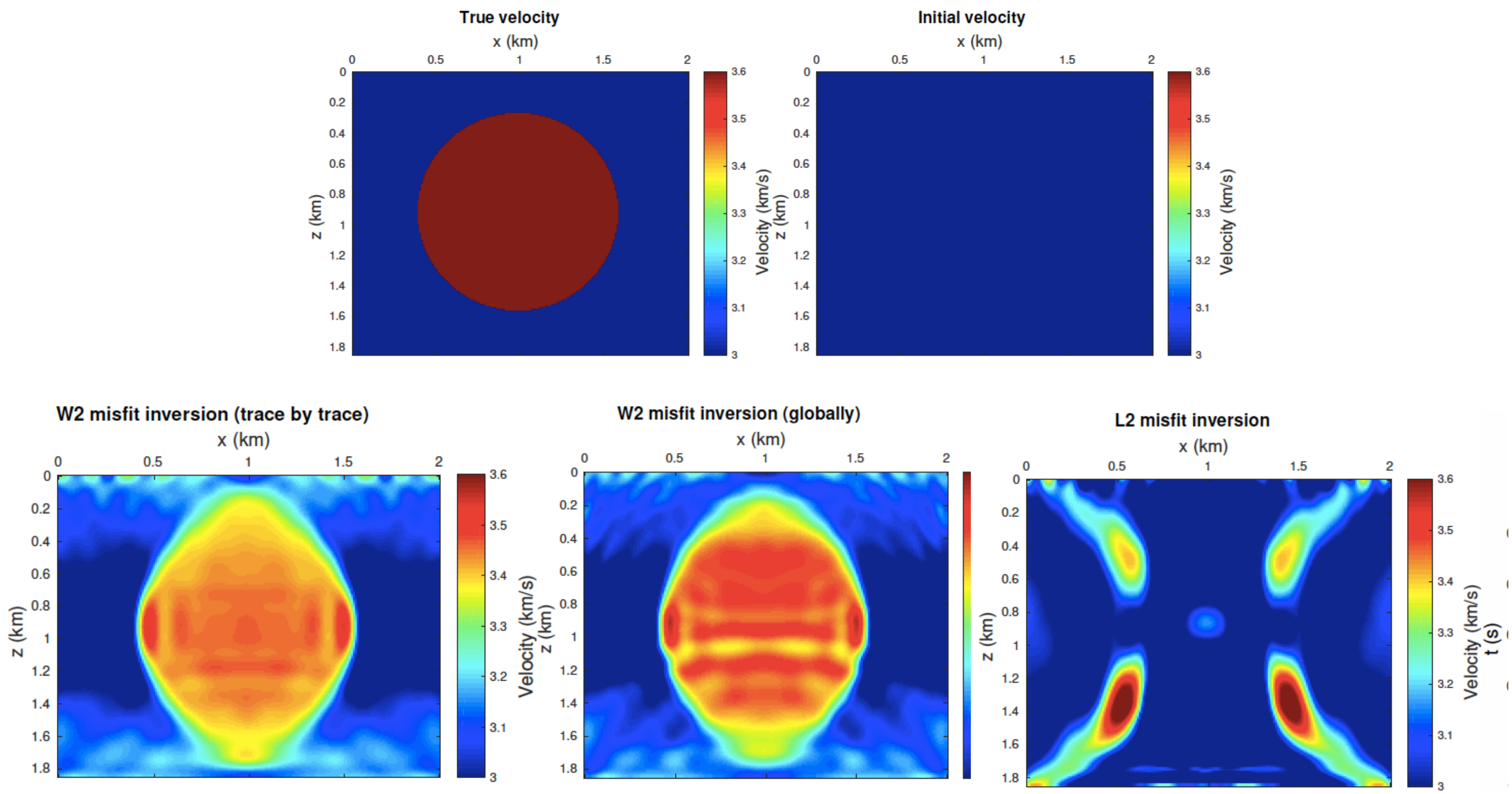


BP 2004 model

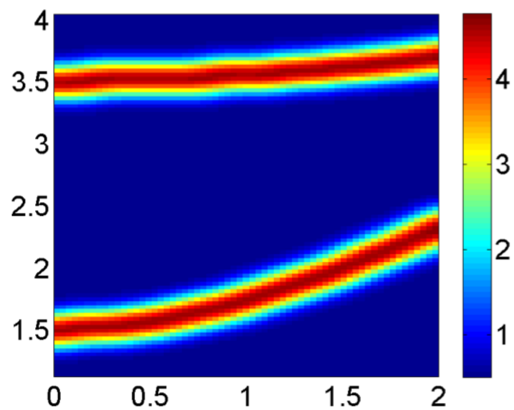
- High contrast salt deposit, W_2-1D, W_2-2D, L^2



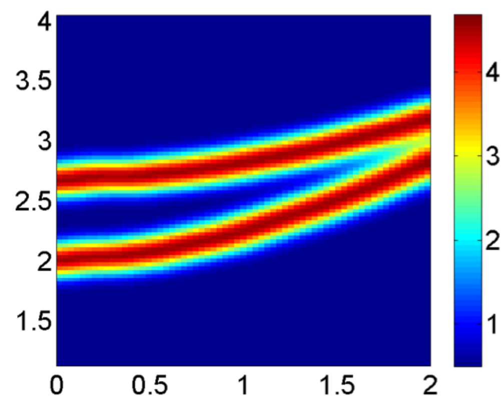
Camembert


True velocity
Initial velocity
W2 misfit inversion (trace by trace)
W2 misfit inversion (globally)
L2 misfit inversion
x (km)
z (km)
Velocity (km/s)
t (s)

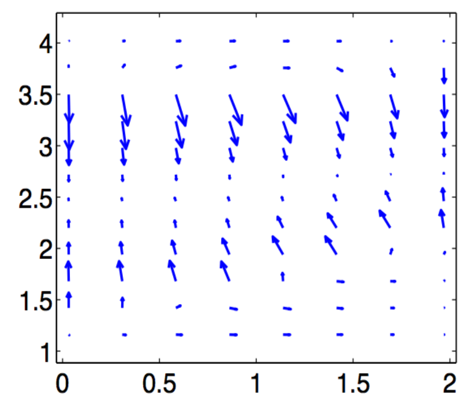
Additional information from Monge-Ampère solution: T=grad(u) for registration



Source



Target

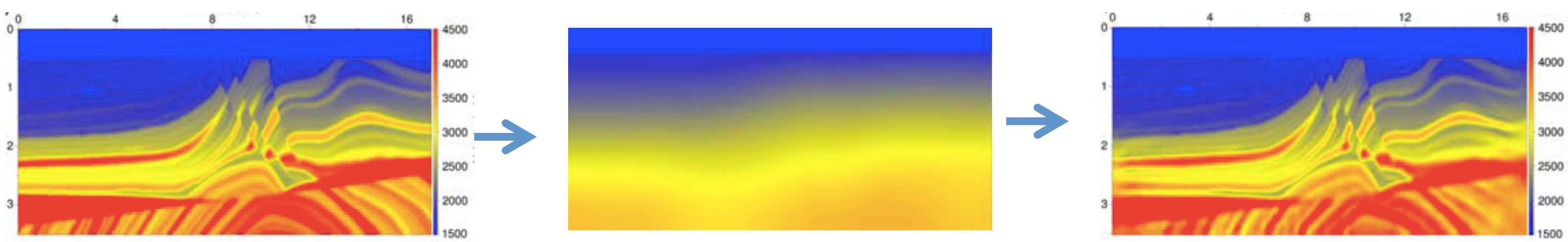


grad(*u*)-*x*

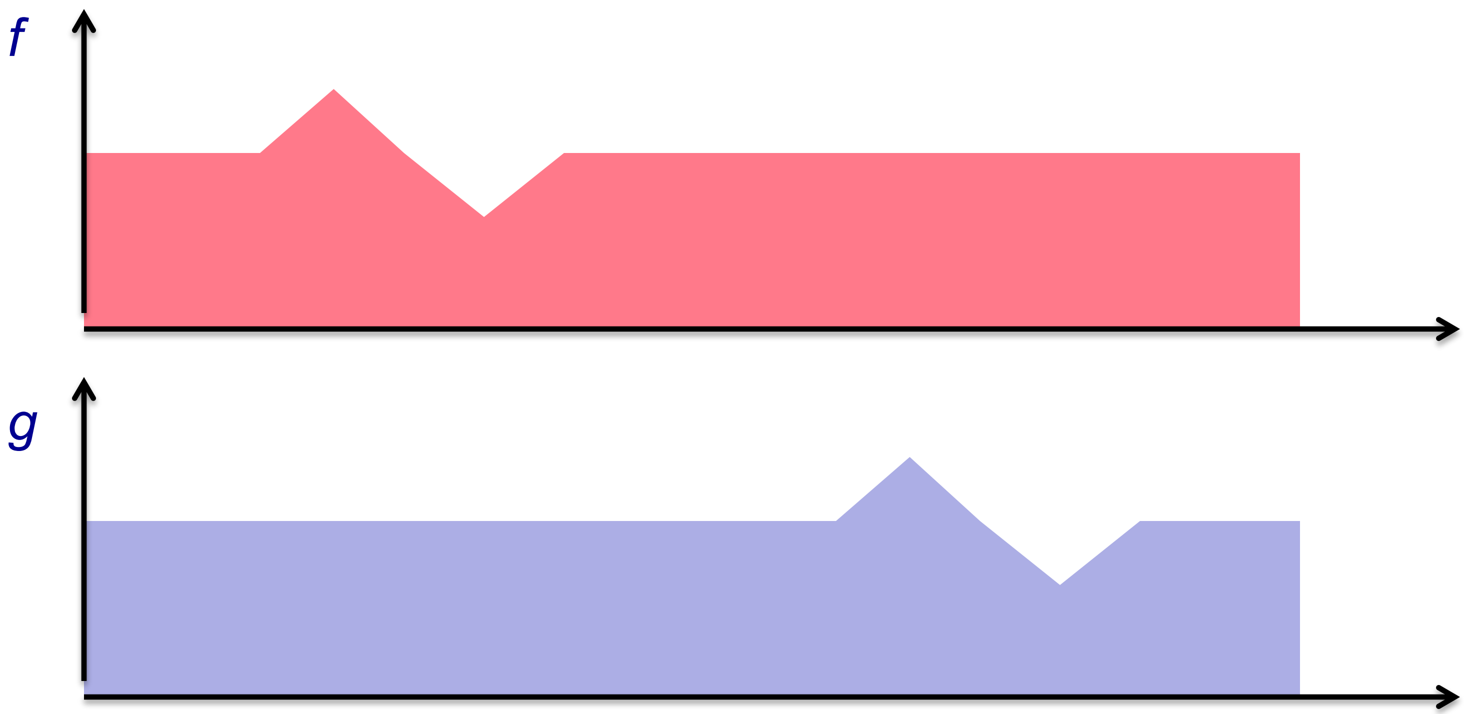
- Seismic applications
  - Matching different measurements (well log – seismic)
  - Monitor reservoir year by year
- Common in image processing (often 1D)

Other related work

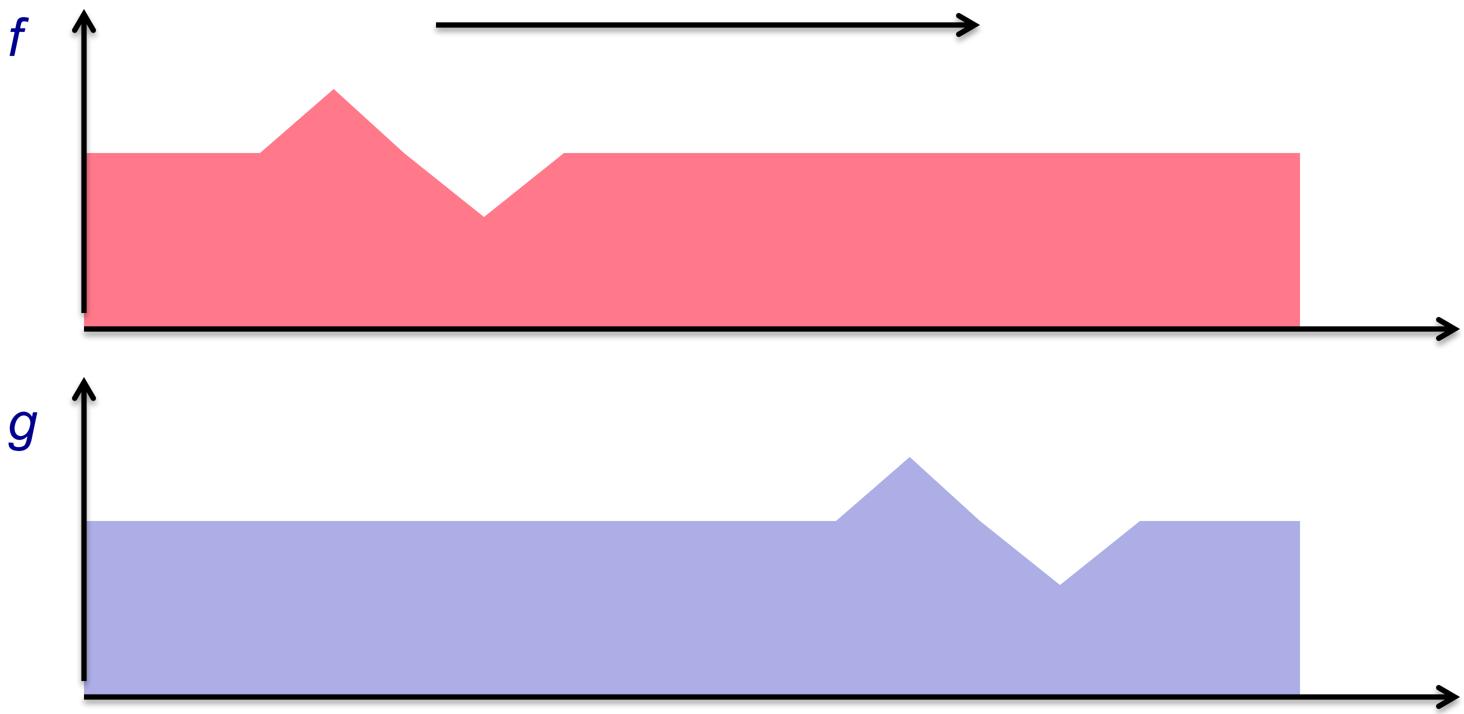
- Example below: W_1 measure and Marmousi p-velocity model [Metivier etr. Al, 2016]
- Current optimal transport based development: Schlumberger, Total



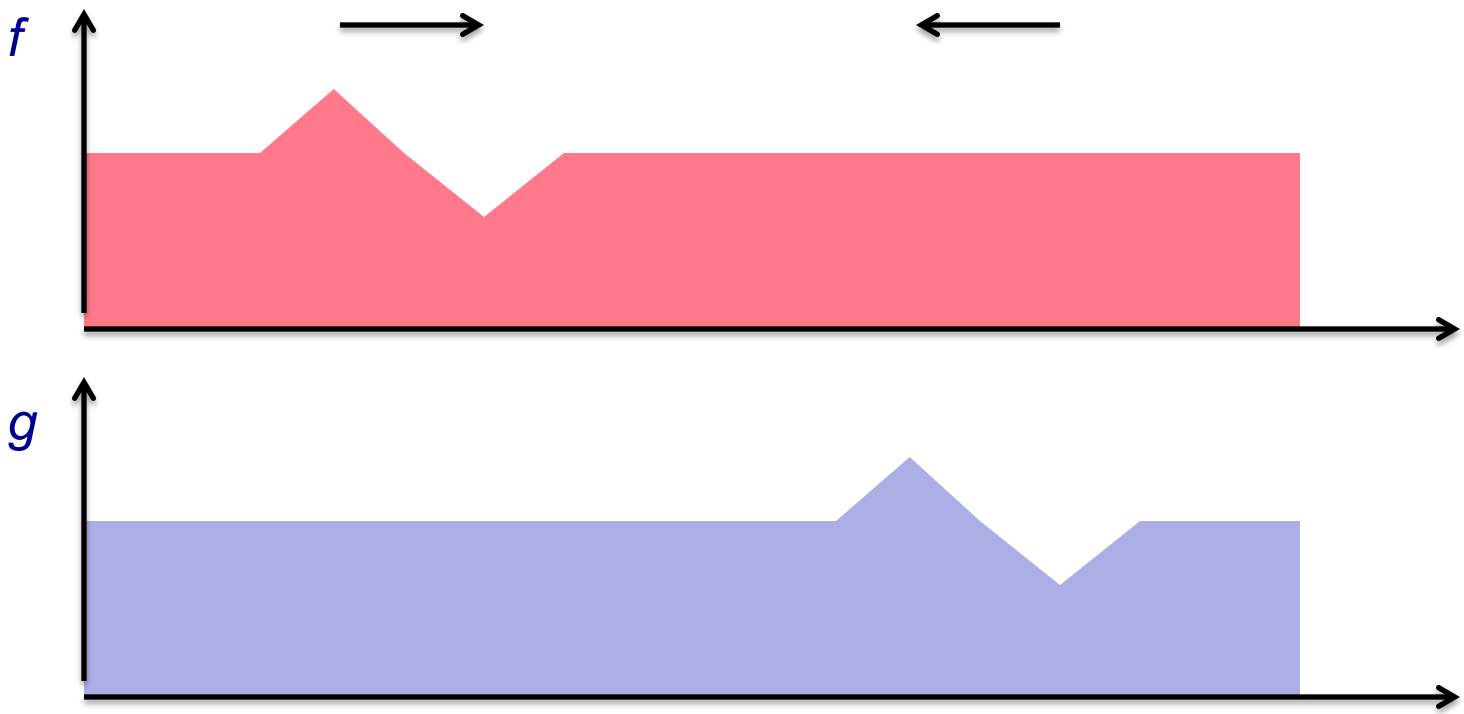
Registration



Registration

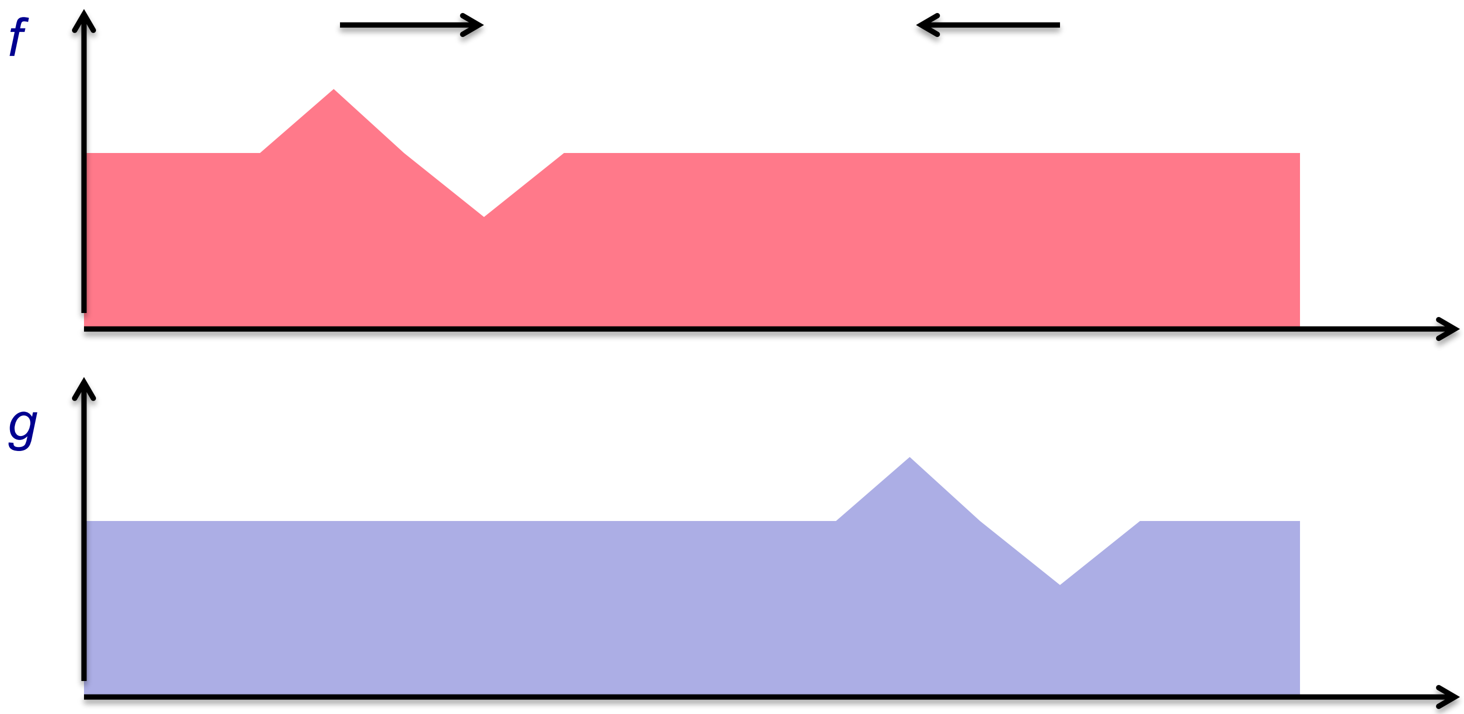


Registration



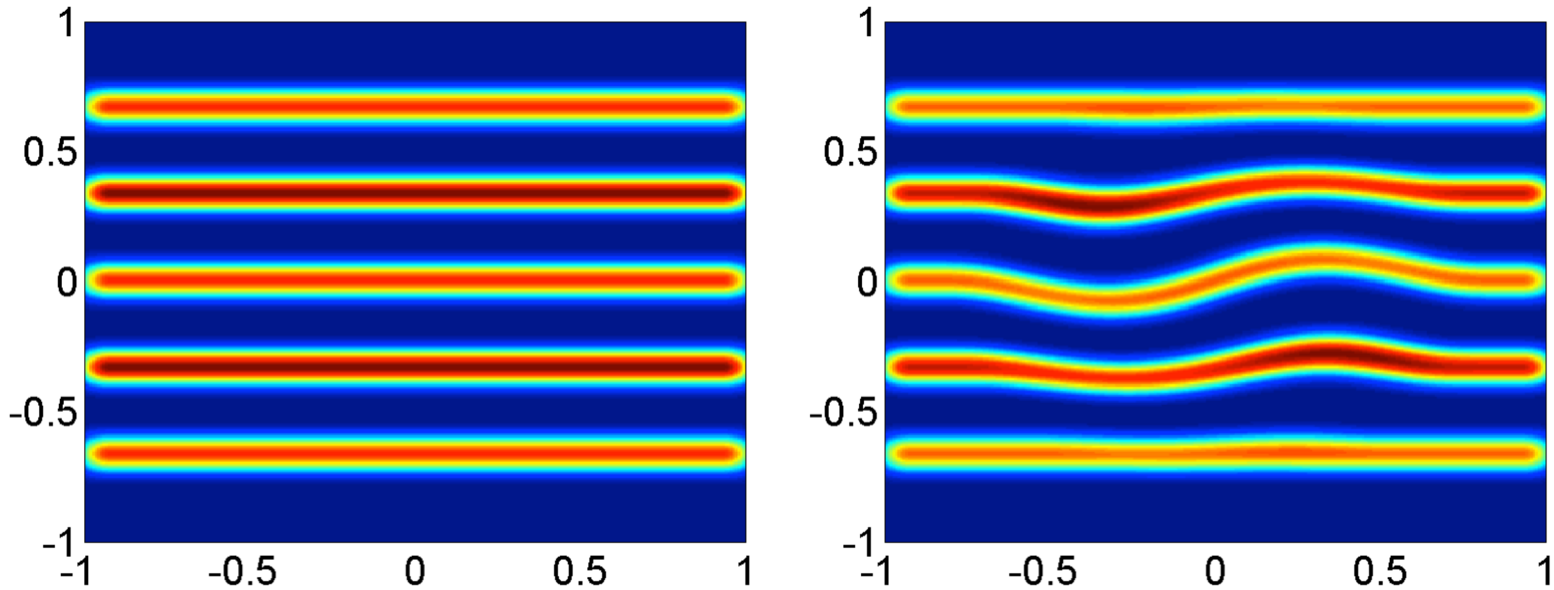
Optimal transport requirements may have unwanted effect on seismic registration

Registration

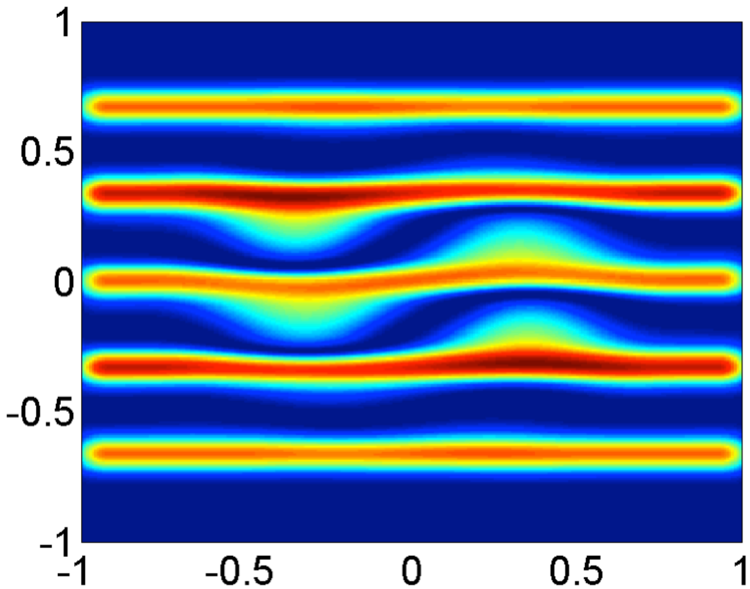


→ Need to modify algorithm

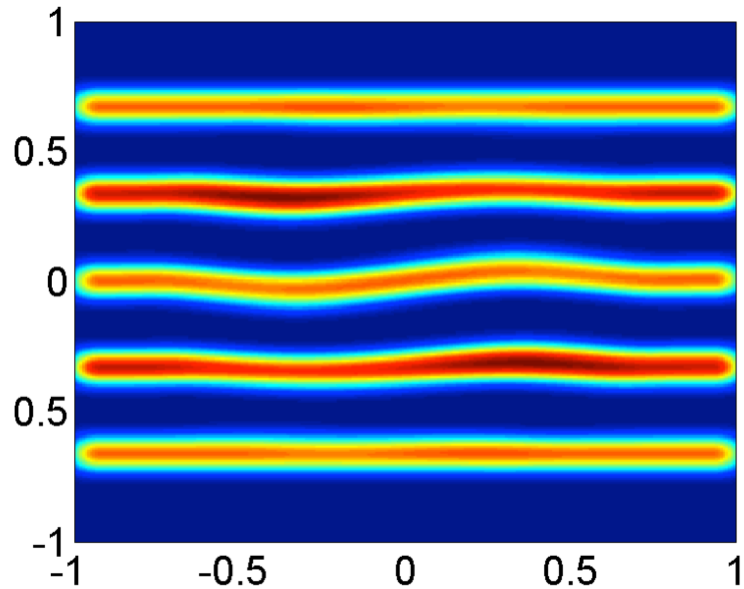
Iteration on truncated signals and maps



Using map based on Monge-Ampère solution for registration



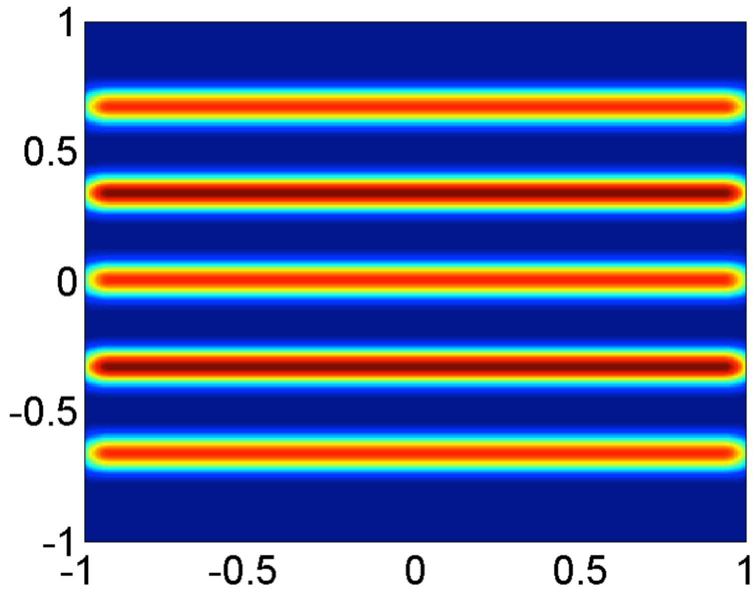
$T(x) = grad(u) - x$



$T_a(x) \approx grad(u) - x$

smooth, weighted L^2

Using map based on Monge-Ampère solution for registration



- The full algorithms based on cropping data and iterate over updated registered maps
- Applications commonly requires modification to the basic theory

5. Conclusions

- Improved seismic exploration requires progress in computational mathematics
- Optimal transport and the Wasserstein metric are promising tools in seismic imaging
- Theory and basic algorithms need to be substantially modified to handle realistic seismic data:
  - B. Engquist and B. Froese, Application of the Wasserstein metric to seismic signals, Comm. in Math. Sciences, 12. 979-988, 2014
  - B. Engquist, B. Froese and Y. Yang, Optimal transport for Seismic full waveform inversion, to appear
  - Y. Yang, B. Engquist and J. Sun, Convexity of the quadratic Wasserstein metric as a misfit function for full waveform inversion, SEG 2016, subm.

Happy Birthday Yann