

APPLICATION OF THE WASSERSTEIN METRIC TO SEISMIC SIGNALS *

BJÖRN ENGQUIST [†] AND BRITTANY D. FROESE [‡]

Abstract. Seismic signals are typically compared using travel time difference or L_2 difference. We propose the Wasserstein metric as an alternative measure of fidelity or misfit in seismology. It exhibits properties from both of the traditional measures mentioned above. The numerical computation is based on the recent development of fast numerical methods for the Monge-Ampère equation and optimal transport. Applications to waveform inversion and registration are discussed and simple numerical examples are presented.

Key words.

Wasserstein metric, seismology, optimal transport, waveform inversion, registration.

1. Introduction A classical way of comparing seismic signals is to use the travel time difference. This can be done in the time domain or, more recently, by different measures of the phase shift; see for example [17]. The misfit or fidelity estimate can be between two measured signals or between a computed signal and a measured seismic signal. For more complex signals, travel time estimates may not be appropriate and L_2 estimates of misfit are often used in full waveform inversion [15, 16].

We propose the Wasserstein metric as a measure of misfit that combines many of the best features of the metrics given by travel time and L_2 . The Wasserstein metric measures the difference between two distributions by the optimal cost of rearranging one distribution into the other [18]. The mathematical definition of the distance between the distributions $f: X \rightarrow \mathbb{R}$, $g: Y \rightarrow \mathbb{R}$ can be formulated as

$$W_2^2(f, g) = \inf_{T \in \mathcal{M}} \int_X |x - T(x)|^2 f(x) dx \quad (1.1)$$

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

We consider two simple one-dimensional examples to show the relation of the Wasserstein metric to travel time and L_2 . First, we compare the hat functions $f(x)$ and $g(x, s) = f(x - s)$ with

$$f(x) = \max\{1 - |x|, 0\}. \quad (1.2)$$

For small s , the L_2 distance is

$$\|f - g\|_{L_2}^2 = 2s^2 + \mathcal{O}(s^3),$$

and the Wasserstein metric measures the misfit by

$$W_2^2(f, g) = s^2.$$

*November 20, 2013

[†]Department of Mathematics and ICES, The University of Texas at Austin, 1 University Station C1200, Austin, TX 78712 USA, (engquist@math.utexas.edu).

The first author was partially supported by NSF DMS-1217203.

[‡]Department of Mathematics and ICES, The University of Texas at Austin, 1 University Station C1200, Austin, TX 78712 USA, (bfroese@math.utexas.edu).

The second author was partially supported by an NSERC PDF.

For large s , on the other hand, the L_2 distance is given by

$$\|f - g\|_{L_2}^2 = \|f\|_{L_2}^2 + \|g\|_{L_2}^2 = 2,$$

which is independent of s and of no value in minimisation. However, the Wasserstein distance preserves the ideal $\mathcal{O}(s^2)$ scaling.

For an example that better resembles seismic signals, consider the misfit between the simple wavelet f in Figure 1.1(a) and another wavelet shifted by a distance s . Figures 1.1(b)-1.1(c) clearly illustrate the advantage of using the Wasserstein distance in minimisation processes.

Earlier algorithms for the numerical computation of the Wasserstein distance required a large number of operations [3, 4, 1]. The optimal transportation problem can be rigorously related to the following Monge-Ampère equation [5, 14], which enables the construction of more efficient methods for computing the Wasserstein distance.

$$\begin{cases} \det(D^2u(x)) = f(x)/g(\nabla u(x)) + \langle u, & x \in X \\ \nabla u(X) = Y \\ u \text{ is convex.} \end{cases} \quad (1.3)$$

The Wasserstein distance is then given by

$$W_2^2(f, g) = \int_X |x - \nabla u(x)|^2 f(x) dx. \quad (1.4)$$

There are now fast and robust numerical algorithms for the solution of (1.3), and thus for the computation of W_2^2 [2].

The solution u of the Monge-Ampère equation contains additional information since the vector $\nabla u(x) - x$ indicates which parts of the distributions f and g are connected under the optimal transport map. This information is useful for problems in image registration [13], meteorology [7], mesh generation [6], reflector design [12], and astrophysics [8]. As we will see, it can also be of great value in seismology.

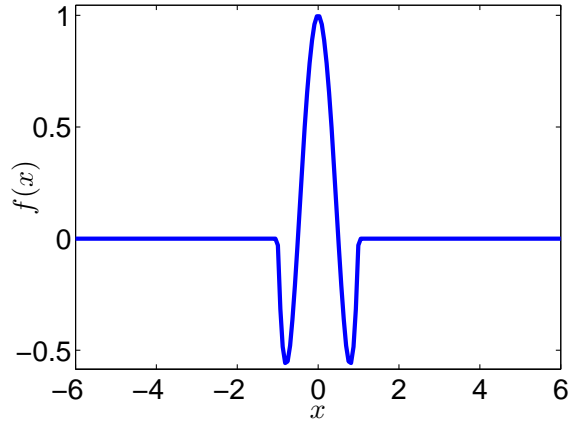
2. Challenges in application to seismology While the Wasserstein distance has many excellent properties, there remain challenges that need to be addressed due to the specific nature of seismic signals. Some of these difficulties come from the formulation of the Wasserstein metric and some from the numerical algorithm used to solve the Monge-Ampère equation.

2.1. Positivity The Wasserstein metric requires $f, g \geq 0$, which is typically not the case with seismic signals. This could be achieved by adding positive constants to f and g , but this would distort the optimal transportation map. Another option is to compare envelopes of the functions. We have chosen to compare separately the positive and negative parts of $f = f^+ - f^-$, $g = g^+ - g^-$, then add the results. Then the misfit we compute is

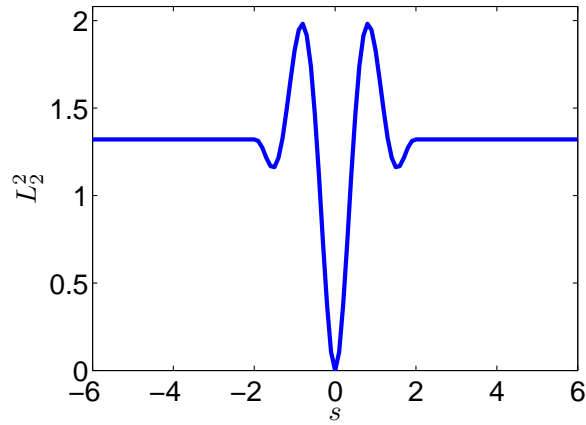
$$W_2^2(f^+, g^+) + W_2^2(f^-, g^-).$$

2.2. Mass conservation The Wasserstein metric also requires that mass is conserved,

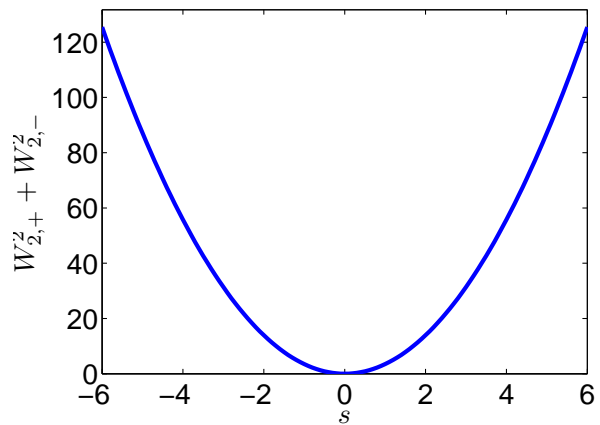
$$\int_X f(x) dx = \int_Y g(y) dy.$$



(a)



(b)



(c)

FIG. 1.1. (a) A wavelet profile $f(x)$. The distances between $f(x)$ and $g(x) = f(x-s)$ measured by (b) $L_2^2(f, g)$, and (c) $W_{2,+}^2(f^+, g^+) + W_{2,-}^2(f^-, g^-)$.

This can be achieved most easily by a simple constant scaling of the densities f and g .

Other scalings are also possible when additional information about the distribution is available. For instance, if the signals f and g are close to each other, with each consisting of two separate components, we could rescale the components separately. This would reduce the risk that the optimal transportation plan $T(x)$ could transport some mass between separate components, which may not be desirable in certain applications.

2.3. Convexity Other challenges come from the solution of the Monge-Ampère equation. In particular, the set Y where the target density $g(y)$ is positive must be convex. This can be accomplished by first selecting a convex set \tilde{Y} containing Y , then preprocessing the data as follows,

$$\tilde{g}^+(y) = \begin{cases} g^+(y) + \theta, & y \in \tilde{Y}^+ \\ 0, & \text{otherwise.} \end{cases} \quad (2.1)$$

Here $\theta > 0$, \tilde{Y}^+ is a convex set containing the support of g^+ , and the same type of transformation can be applied to the other components f^+, f^-, g^- . In the examples below, we choose the convex sets to be rectangles. This θ -layer will introduce a small amount of artificial transport into and out of the layer. To reduce distortion of the optimal transportation plan, we will choose all rectangles to have the same size, with each one centred at the centre of mass of the corresponding distribution. If the optimal transportation plan itself is of interest, we can reduce some of the artificial transport by thresholding the transport vectors $\nabla u(x) - x$ to zero in the layer.

2.4. Regularity The numerical method used to solve the Monge-Ampère equation also requires that the ratio $f(x)/g(y)$ is Lipschitz continuous in the y variable. This leads to the requirement that θ cannot be too small, particularly if there are regions where $g=0$ and the corresponding gradient is not small. Appropriate choice of θ and convolution of g (and f for symmetry) with a regularising kernel will ensure the success of the numerical method.

3. Numerical algorithm We describe here a two-dimensional form of the algorithm we use to solve the Monge-Ampère equation. The method is based on the following variational characterisation of the Monge-Ampère equation (1.3) combined with the convexity constraint [9, 10].

$$\det^+(D^2u) = \min_{\{\nu_1, \nu_2\} \in V} \{ \max\{u_{\nu_1, \nu_1}, 0\} \max\{u_{\nu_2, \nu_2}, 0\} + \min\{u_{\nu_1, \nu_1}, 0\} + \min\{u_{\nu_2, \nu_2}, 0\} \} \quad (3.1)$$

where V is the set of all orthonormal bases for \mathbb{R}^2 . The transportation constraint $\nabla u(X) = Y$ can be re-expressed as the Hamilton-Jacobi equation

$$H(\nabla u(x)) = 0, \quad x \in \partial X \quad (3.2)$$

where $H(y)$ is the signed distance to the convex target set Y [2].

The discretisation described below leads to a large system of nonlinear equations, which is solved using Newton's method. The linear equations arising in each Newton iteration are solved using a direct sparse solver. In a typical example, fewer than ten iterations are required for convergence.

3.1. Monotone approximation of the Monge-Ampère equation Equation (3.1) can be discretised by computing the minimum over finitely many directions $\{\nu_1, \nu_2\}$, which may require the use of a wide stencil. For simplicity and brevity, we describe a compact version of the scheme and refer to [10] for complete details.

We begin by introducing the finite difference operators

$$\begin{aligned} [\mathcal{D}_{x_1 x_1} u]_{ij} &= \frac{1}{dx^2} (u_{i+1,j} + u_{i-1,j} - 2u_{i,j}) \\ [\mathcal{D}_{x_2 x_2} u]_{ij} &= \frac{1}{dx^2} (u_{i,j+1} + u_{i,j-1} - 2u_{i,j}) \\ [\mathcal{D}_{x_1} u]_{ij} &= \frac{1}{2dx} (u_{i+1,j} - u_{i-1,j}) \\ [\mathcal{D}_{x_2} u]_{ij} &= \frac{1}{2dx} (u_{i,j+1} - u_{i,j-1}) \\ [\mathcal{D}_{vv} u]_{ij} &= \frac{1}{2dx^2} (u_{i+1,j+1} + u_{i-1,j-1} - 2u_{i,j}) \\ [\mathcal{D}_{v^\perp v^\perp} u]_{ij} &= \frac{1}{2dx^2} (u_{i+1,j-1} + u_{i-1,j+1} - 2u_{i,j}) \\ [\mathcal{D}_v u]_{ij} &= \frac{1}{2\sqrt{2}dx} (u_{i+1,j+1} - u_{i-1,j-1}) \\ [\mathcal{D}_{v^\perp} u]_{ij} &= \frac{1}{2\sqrt{2}dx} (u_{i+1,j-1} - u_{i-1,j+1}). \end{aligned}$$

In the compact version of the scheme, the minimum in (3.1) is approximated using only two possible values. The first uses directions aligning with the grid axes.

$$\begin{aligned} MA_1[u] &= \max\{\mathcal{D}_{x_1 x_1} u, \delta\} \max\{\mathcal{D}_{x_2 x_2} u, \delta\} \\ &\quad - \min\{\mathcal{D}_{x_1 x_1} u, \delta\} - \min\{\mathcal{D}_{x_2 x_2} u, \delta\} - f/g(\mathcal{D}_{x_1} u, \mathcal{D}_{x_2} u) - u_0. \end{aligned} \quad (3.3)$$

Here dx is the resolution of the grid, $\delta > K\Delta x/2$ is a small parameter that bounds second derivatives away from zero, u_0 is the solution value at a fixed point in the domain, and K is the Lipschitz constant in the y -variable of $f(x)/g(y)$.

For the second value, we rotate the axes to align with the corner points in the stencil, which leads to

$$\begin{aligned} MA_2[u] &= \max\{\mathcal{D}_{vv} u, \delta\} \max\{\mathcal{D}_{v^\perp v^\perp} u, \delta\} - \min\{\mathcal{D}_{vv} u, \delta\} - \min\{\mathcal{D}_{v^\perp v^\perp} u, \delta\} \\ &\quad - f/g\left(\frac{1}{\sqrt{2}}(\mathcal{D}_v u + \mathcal{D}_{v^\perp} u), \frac{1}{\sqrt{2}}(\mathcal{D}_v u - \mathcal{D}_{v^\perp} u)\right) - u_0. \end{aligned} \quad (3.4)$$

Then the compact monotone approximation of the Monge-Ampère equation is

$$-\min\{MA_1[u], MA_2[u]\} = 0. \quad (3.5)$$

3.2. Monotone approximation of the Hamilton-Jacobi boundary condition We describe the boundary conditions in the special case where the source and target sets are rectangles and refer to [2] for details of the more general setting. In this setting, the Hamilton-Jacobi equation can be written as the Neumann boundary condition

$$u_{x_1}(x_1^{min}) = y_1^{min}, u_{x_1}(x_1^{max}) = y_1^{max}, u_{x_2}(x_2^{min}) = y_2^{min}, u_{x_2}(x_2^{max}) = y_2^{max}, \quad (3.6)$$

which transports each side of the domain X to the corresponding side of the target rectangle Y . A monotone discretisation is easily constructed; on the left side of the domain, for example,

$$\frac{u_{2,j} - u_{1,j}}{dx} = y_1^{min}. \quad (3.7)$$

3.3. Filtered approximation For improved accuracy, we can combine the monotone scheme F_M just described with a scheme that is formally more accurate F_A . A second-order scheme is easily constructed using a centred difference discretisation of the two-dimensional Monge-Ampère equation

$$u_{xx}u_{yy} - u_{xy}^2 = f/g(u_x, u_y)$$

and a one-sided second-order approximation of the Neumann boundary conditions. We introduce a filter such as

$$S(x) = \begin{cases} x & |x| \leq 1 \\ 0 & |x| \geq 2 \\ -x+2 & 1 \leq x \leq 2 \\ -x-2 & -2 \leq x \leq -1. \end{cases} \quad (3.8)$$

Then a convergent, higher-order scheme is given by

$$F_F[u] = F_M[u] + \epsilon S \left(\frac{F_A[u] - F_M[u]}{\epsilon} \right). \quad (3.9)$$

In many cases, the higher order accuracy can be achieved even when the filtered scheme is based on a compact monotone scheme; see [11].

4. Numerical examples for a two-layer material

We consider the example of the response from a two-layer material with upper and lower depths d_1, d_2 and wave speeds v_1, v_2 respectively; see Figure 4.1(a). Typical seismic signals are pictured in Figure 4.1(b)-4.1(c) in the offset-time domain. After preprocessing, we compute the optimal transportation plan between these two distributions.

4.1. Inversion One potential application of optimal transportation is full waveform inversion. To accomplish this, it is necessary to determine the unknown parameters that minimise the misfit between the observed and synthetic signals. In this example, the unknown parameters are d_1, d_2, v_1, v_2 . To demonstrate the advantage of the Wasserstein metric as a measure of misfit, we fix one signal g that is computed using $d_1 = 1, d_2 = 0.5, v_1 = 1, v_2 = 1.5$. We then compute the distance $W_2^2(f, g)$, where f is computed for a range of different parameter values. We plot several cross-sections of this distance in Figure 4.2. For comparison, we also plot several cross-sections of the L_2^2 distance $\|f - g\|_{L_2}^2$. It is clear that the Wasserstein distance is much more suitable for minimisation. In preliminary computations, the minimisation has been successfully accomplished using a simple Matlab implementation of the Nelder-Mead simplex method.

4.2. Registration A second application we have in view is seismic registration. With this in mind, we consider the (scaled) displacement vectors $\nabla u(x) - x$, which are pictured in Figure 4.1(d). This figure indicates that the two components of the initial

distribution are being transported towards the two corresponding components of the target distribution. (The small amount of artificial transport due to the preprocessing is truncated to zero). The vector $\nabla u(x) - x$ then gives the connection between f and g and $\det(D^2u)$ can be used to measure the registered amplitude difference.

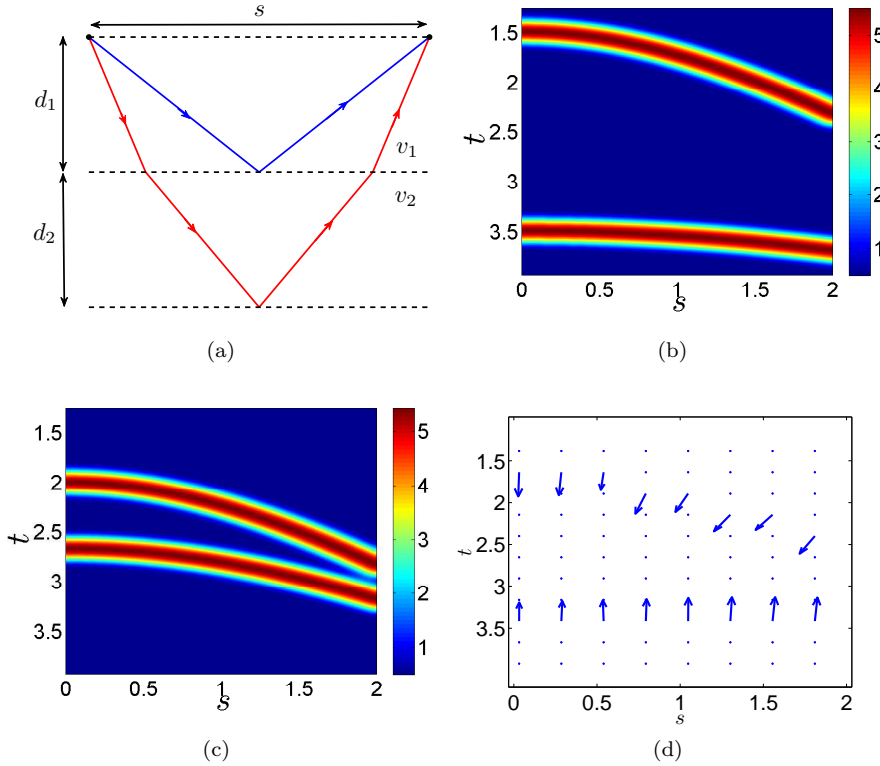


FIG. 4.1. (a) A two-layer material, (b),(c) seismic signals from different materials, and (d) the displacement $\nabla u(x) - x$ coming from the optimal transportation map that defines registration between the two signals.

4.3. Noise The W_2 metric is highly robust to noise. The difference between a noisy signal

$$f(x) + h(x) > 0, \quad \mathbb{E}[h(x)] = 0$$

and the clean signal $f(x) > 0$ is typically minimal owing to the strong cancellation between nearby positive and negative values of $h(x)$. The L_2 difference $\|h\|_{L_2}$ is typically substantially larger.

To demonstrate the insensitivity to noise, we repeat the computation of the $W_2^2(f(x), f(x-s))$ distance for the wavelet profiles in Figure 1.1. However, this time we add uniform random noise into either the source distribution f or both distributions f, g . While the noise has a clear effect on the computed values of the L_2^2 difference, the W_2^2 distance computed between the noisy profiles is nearly indistinguishable from the

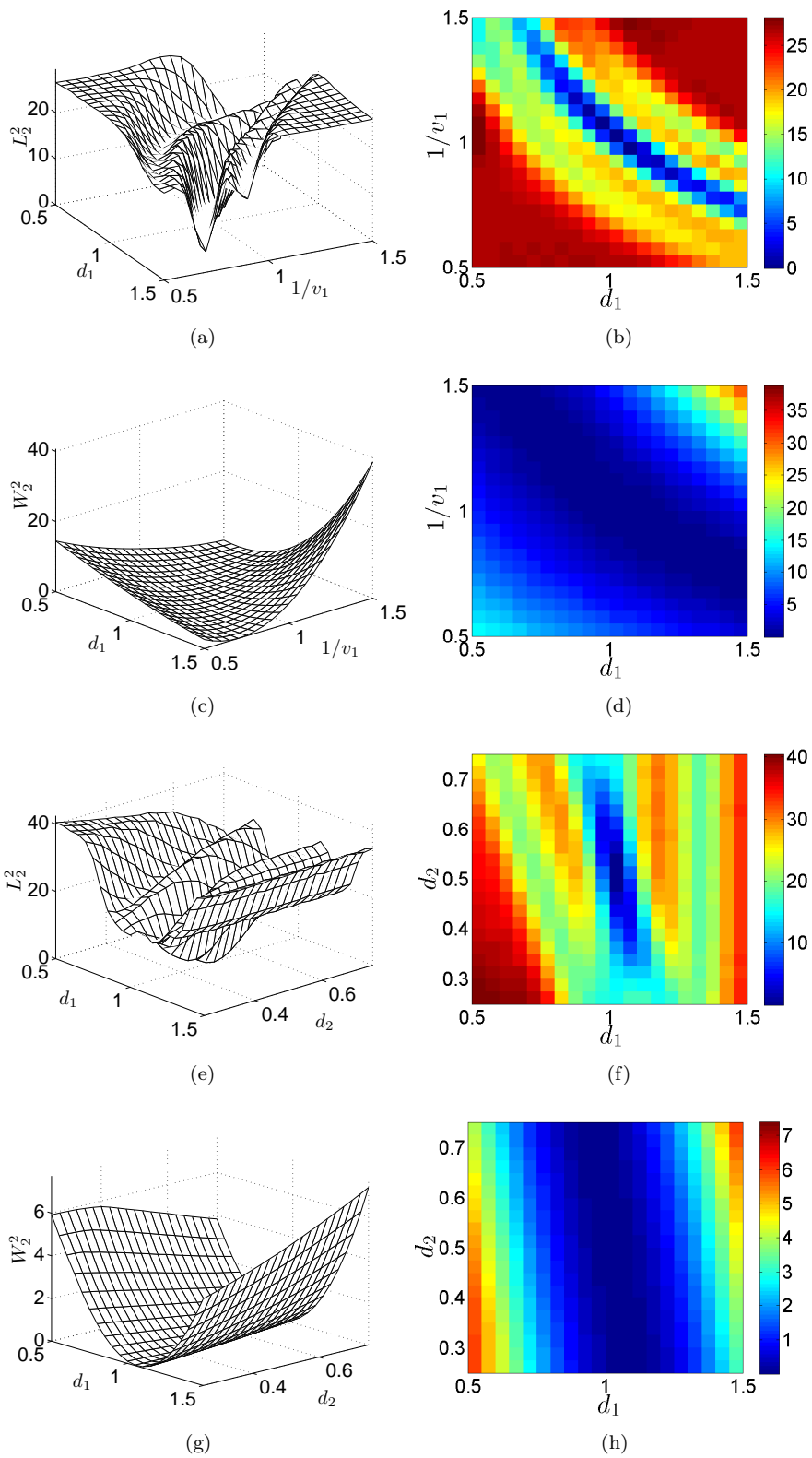


FIG. 4.2. Distances between profiles in Figure 4.1. Cross-sections of (a),(b) L_2^2 and (c),(d) W_2^2 for $d_2=0.5, v_2=1.5$. Cross-sections of (e),(f) L_2^2 and (g),(h) W_2^2 for $v_1=1, v_2=1.5$.

original setting; see Figure 4.3. For clarity, we restrict this presentation to one dimension, but similar results are observed when we introduce noise into two-dimensional distributions.

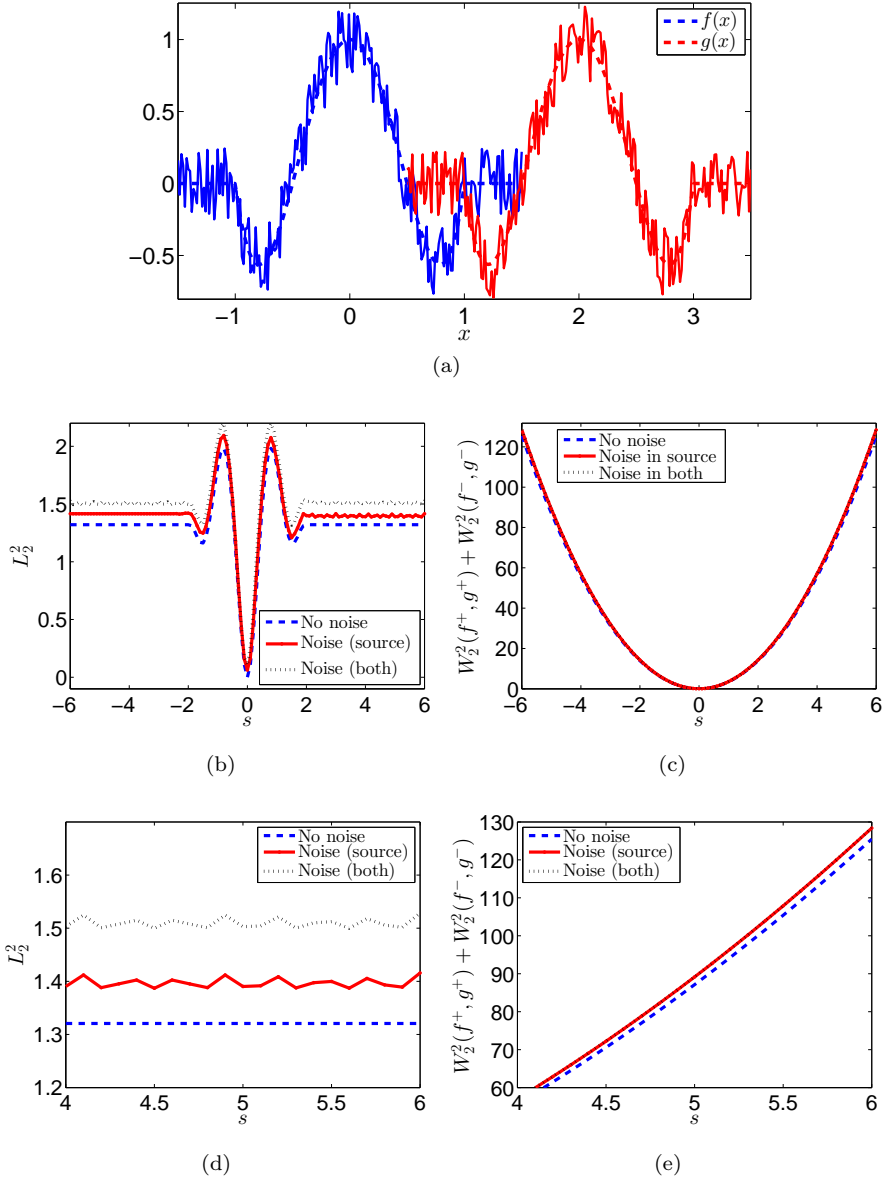


FIG. 4.3. (a) Sample noisy profiles $f(x)$ and $g(x)$. The distances between $f(x)$ and $g(x)$ measured by (b) $L_2^2(f, g)$, and (c) $W_2^2(f^+, g^+) + W_2^2(f^-, g^-)$. Enlarged view of a portion of the (d) $L_2^2(f, g)$, and (e) $W_2^2(f^+, g^+) + W_2^2(f^-, g^-)$ distances.

5. Conclusions We have introduced the Wasserstein metric W_2^2 as a measure of fidelity or misfit in seismology. It can be seen as incorporating the most desirable properties from both the travel time difference and the L_2^2 difference. We exploit recent progress in the theory of optimal transport and in fast, robust numerical methods for the Monge-Ampère equation. We further present solutions to specific challenges coming from seismic signals. A set of simple numerical examples illustrates the advantages of this approach for potential application to full waveform inversion and registration. Our Monge-Ampère based techniques are easily generalised to higher dimensions.

REFERENCES

- [1] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.
- [2] J.-D. Benamou, B. D. Froese, and A. M. Oberman. Numerical solution of the optimal transportation problem using the Monge-Ampère equation. 2013. Submitted.
- [3] D. P. Bertsekas. Auction algorithms for network flow problems: a tutorial introduction. *Comput. Optim. Appl.*, 1(1):7–66, 1992.
- [4] D. Bosc. Numerical approximation of optimal transport maps. *SSRN*, 2010.
- [5] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.
- [6] C. J. Budd and J. F. Williams. Moving mesh generation using the parabolic Monge-Ampère equation. *SIAM J. Sci. Comput.*, 31(5):3438–3465, 2009.
- [7] M. J. P. Cullen, J. Norbury, and R. J. Purser. Generalised Lagrangian solutions for atmospheric and oceanic flows. *SIAM J. Appl. Math.*, 51(1):20–31, 1991.
- [8] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski. A reconstruction of the initial conditions of the universe by optimal mass transportation. *Nature*, 417, 2002.
- [9] B. D. Froese. A numerical method for the elliptic Monge-Ampère equation with transport boundary conditions. *SIAM J. Sci. Comput.*, 34(3):A1432–A1459, 2012.
- [10] B. D. Froese and A. M. Oberman. Convergent finite difference solvers for viscosity solutions of the elliptic Monge-Ampère equation in dimensions two and higher. *SIAM J. Numer. Anal.*, 49(4):1692–1714, 2011.
- [11] B. D. Froese and A. M. Oberman. Convergent filtered schemes for the Monge-Ampère partial differential equation. *SIAM J. Numer. Anal.*, 51(1):423–444, 2013.
- [12] T. Glimm and V. Olikar. Optical design of single reflector systems and the Monge-Kantorovich mass transfer problem. *J. Math. Sci. (N. Y.)*, 117(3):4096–4108, 2003. Nonlinear problems and function theory.
- [13] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *Int. J. Comput. Vision*, 60(3):225–240, 2004.
- [14] M. Knott and C. S. Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43(1):39–49, 1984.
- [15] R. G. Pratt. Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3):888–901, 1999.
- [16] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.
- [17] T. Van Leeuwen and W. A. Mulder. A correlation-based misfit criterion for wave-equation travelttime tomography. *Geophys. J. Int.*, 182(3):1383–1394, 2010.
- [18] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.