Fourier-Based Bounds for Wasserstein Distances and Their Implications for Data-Driven Problems

Vlad Kobzar[1]
joint work with Wanli Hong[2] and Kui Ren[1]

[1]Columbia, [2]NYU

Outline

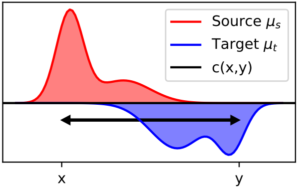
- Background and motivation
  - Optimal transport
  - W_p distances
  - Frequency domain/Fourier basis
- Fourier-based bounds on W_p
- Resolution of frequencies in W_p mismatch minimization
- Future directions

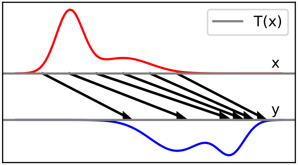
What is optimal transport?

- A geometric framework for comparing probability measures
- Rich history of applications
  - Construction of fortifications under Napoleon (Monge, 1781)
  - Optimization of production in the USSR (Kantorovich, 1942)
  - Promising recent applications in ML and inverse problems
- Led to many discoveries in geometry, analysis and optimization

Monge (1781)

Monge





- Optimal *transport map* T is the minimizer of

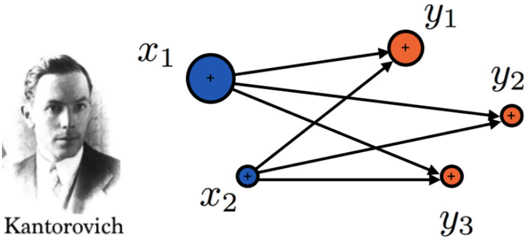
$$MP(\mu_s, \mu_t) := \min_{\mu_t = T_\sharp \mu_s} \int c(x, T(x)) d\mu_s$$

- μ_s, μ_t are a.c. $\Rightarrow$ Δ of vars. for assoc. densities f, g

$$\mu_t = T_\sharp \mu_s \Leftrightarrow g(x)dx = f(T(x)) |\det \partial_x T(x)| dx$$

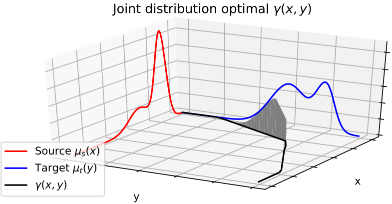
- Nonconvex problem even if c is convex
- Original Monge's cost $c(x, y) = |x - y|$

Kantorovich (1942)



- Look for an optimal assignment (coupling) of masses
- Allows to split mass, in contrast to transport map T

Kantorovich (1942)



- The optimal *transport plan* γ is a minimizer of

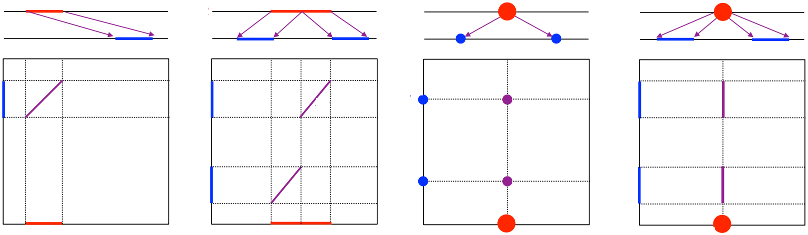
$$KP(\mu_s, \mu_t) := \min_{\gamma \in \mathcal{P}} \mathbb{E}_\gamma c(x, y)$$

- $\mathcal{P}$ are all joint probability measures γ with

$$\mu_s = \int \gamma(\cdot, y) dy \text{ and } \mu_t = \int \gamma(x, \cdot) dx$$

- This problem is convex (for discrete measures it's an LP)

Transport plans



- A few examples of transport plans between uniform and discrete measures in 1D

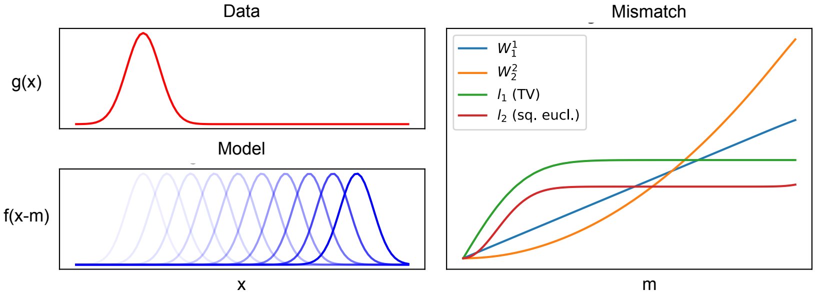
Wasserstein distances

- When c is a p-th moment of a distance, KP is also a distance
- Focus on the Euclidean case: $c(x, y) = |x - y|^p$

$$W_p^p(\mu_t, \mu_s) := \min_{\gamma \in \mathcal{P}} \mathbb{E}_\gamma |x - y|^p$$

- We'll consider only $p = 1$ vs 2 for simplicity

W_p mismatch

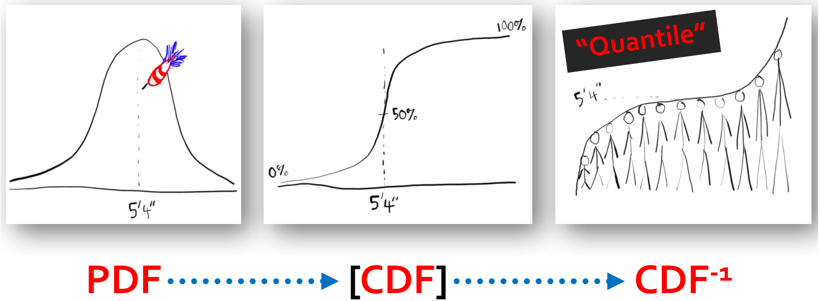


- Using a mismatch functional Φ, fit model $f(m)$ to data g

$$\min_m \Phi(f(m), g)$$

- W_p mismatch meaningful even if
  - $f(m)$ and g have non-overlapping support, or
  - g is in a low-dim. manifold or shifted relative to $f(m)$
- W_1 is used in ML, stats inference and inverse problems

W_p in 1D



- Explicit form

$$W_1(\mu, \nu) := \int_{\mathbb{R}} |\mathrm{CDF}(\mu) - \mathrm{CDF}(\nu)| dx$$

$$W_p^p(\mu, \nu) := \int_0^1 |\mathrm{CDF}^{-1}(\mu) - \mathrm{CDF}^{-1}(\nu)|^p dx$$

- Does not extend to higher dimensions

Kantorovich-Rubinstein Theorem

- W_1 has a dual formulation

$$W_1(\mu, \nu) = \max_{\varphi \in \text{Lip}-1} \mathbb{E}_\mu \varphi - \mathbb{E}_\nu \varphi$$

  where

$$\text{Lip} - K = \{\varphi : \ |\varphi(x) - \varphi(y)| \leq K|x - y|\}$$

- Used in ML applications, like WGANs

$$\max_w \frac{1}{m} \sum_{i=1}^{m} \varphi_w(x_i) - \varphi_w(g_\theta(z_i))$$

  where
  - φ_w is a $\text{Lip} - K$ neural network (critic),
  - $g_\theta(z_i)$ is a generative neural network (actor)
  - $x_1, \ldots, x_m$ are real data
  - z_i's are sampled from a fixed distribution

Kantorovich-Rubinstein Norm

- Generalizes W_1 to unbalanced signed measures

$$KR(\mu,\nu) := \max_{\substack{\varphi \in \mathrm{Lip}-1 \\ \|\varphi\|_\infty \leq 1}} \mathbb{E}_\mu \varphi - \mathbb{E}_\nu \varphi$$

- Used in inverse problems, like seismic imaging/FWI Lellmann et al. (2014); Métivier et al. (2016a,b); Métivier et al. (2022)
- W_1 and KR are challenging to analyze because the cost $|x-y|$ is not smooth or strictly convex

W_p mismatch minimization

- When f is highly nonconvex and m is high-dimensional, it is not computationally tractable to accurately solve

$$\min_m \Phi(f(\cdot, m), g)$$

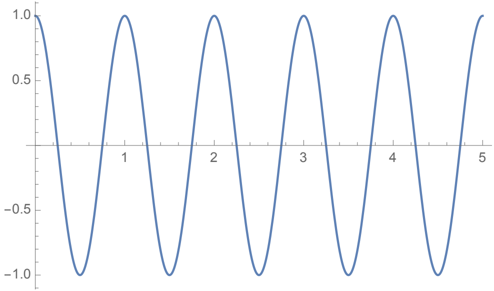
even if g is in the range of f.

- How does using W_p impact an approximate solution?

$$W_p(f(\cdot, m), g) < \delta$$

- What is the impact of different values of p?
- To study this, we bound $W_p(f, g)$ in terms of the Fourier coefficients of f and g

Fourier series



- Represent periodic signals with Fourier series

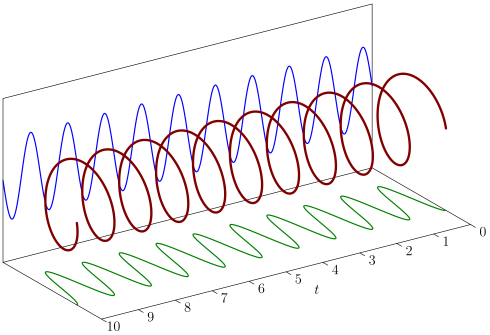
$$f(x) = \sum_k a_k \psi_{k,\theta_k}(x)$$

where

$$\psi_{k,\theta_k}(x) := \cos(2\pi kx + \theta_k)$$

- Not convenient because $\psi_{k,\theta_k}(x)$ depends nonlinearly on θ_k

Fourier basis



- We define the complex sinusoid with frequency $k \in \mathbb{Z}$ as

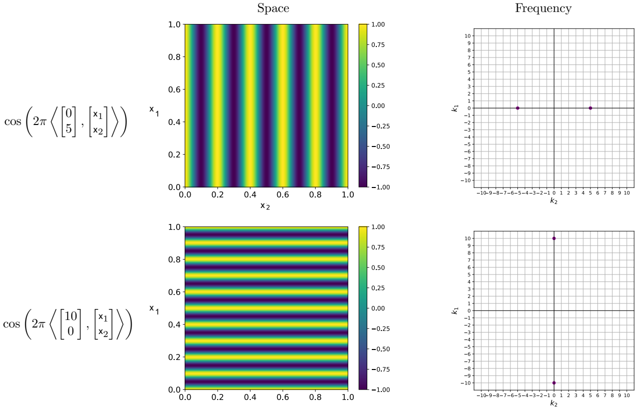
$$\psi_k(x) := \exp(i2\pi kx) = \cos(2\pi kx) + i\sin(2\pi kx)$$

- A real sinusoid with frequency k can be represented as

$$\cos(2\pi ft + \theta) = \frac{\exp(i\theta)}{2}\psi_k(x) + \frac{\exp(-i\theta)}{2}\psi_{-k}(x)$$

- Now the phase is encoded in the complex amplitudes.

Multivariate Fourier series



For $f : [0,1)^d \to \mathbb{R}$

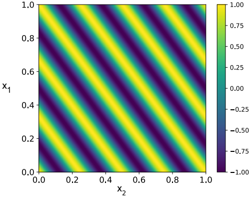
$$f(x) = \sum_k \hat{f}_k \psi_k(x)$$

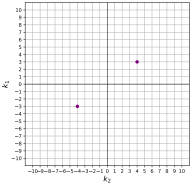
where

$$\psi_k(x) = \exp\left(i2\pi\langle k, x\rangle\right)$$

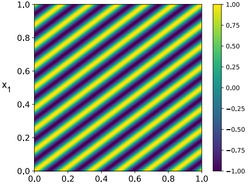
Multivariate Fourier series

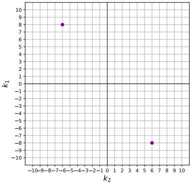
$$\cos\left(2\pi\left\langle\begin{bmatrix}3\\4\end{bmatrix},\begin{bmatrix}x_1\\x_2\end{bmatrix}\right\rangle\right)$$





$$\cos\left(2\pi\left\langle\begin{bmatrix}8\\6\end{bmatrix},\begin{bmatrix}x_1\\x_2\end{bmatrix}\right\rangle\right)$$





For $f : [0,1)^d \to \mathbb{R}$

$$f(x) = \sum_k \hat{f}_k \psi_k(x)$$

where

$$\psi_k(x) = \exp\left(i2\pi\langle k, x\rangle\right)$$

Assumptions

1. μ and ν are a.c., and therefore are associated with their densities f and g

$$W_p(f, g) := W_p(\mu, \nu)$$

2. The domain Ω is flat torus $\mathbb{T}^d := [0, 1)^d$
   - Our analysis also works on hypercube $\mathbb{H}^d := [0, 1]^d$ where f and g must satisfy symmetric boundary condition
3. For $k \in \mathbb{Z}^d$, the Fourier series

$$f = \sum_k \hat{f}_k \psi_k \text{ and } g = \sum_k \hat{g}_k \psi_k$$

converge in $L^2(\Omega)$ where

$$\psi_k(x) := e^{2\pi i \langle k, x \rangle}$$

Existing Fourier-based bounds on W_2

- Peyre (2018):

$$W_2(f,g) \asymp \|f-g\|_{\dot{\mathcal{H}}^{-1}}$$

  where the Sobolev norm

$$\|f-g\|_{\dot{\mathcal{H}}^{-1}}^2 = \sum_k \left(\frac{\hat{f}_k - \hat{g}_k}{|k|}\right)^2$$

- Led to frequency resolution analysis in W_2 mismatch minimization (Engquist et al., 2020)

Existing W_p bounds

- **No previously known Fourier-based bounds for W_p when $p \neq 2$, except in special cases:**
  - Measures on a circle (Steinerberger, 2021)
  - Measures on a finite discrete grid (Auricchio et al., 2020)
  - Trivial bound $W_p \leq W_2 \lesssim \|f - g\|_{\dot{\mathcal{H}}^{-1}}$
- Open problem (Steinerberger, 2021)
  - To establish a Fourier-based l.b. on W_p that $\uparrow$ as $p \uparrow$
  - Applications in measure theoretic discrepancy
- Our work resolves this open problem (Hong et al., 2023)
  - We adapt arguments used to establish wavelet-based bounds on W_p (Niles-Weed and Berthet, 2022)

Lower bounds

$p = 1$ and 2	$d^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{\infty, w^1}$
$p = 2$, $\|f\|_{L_s}, \|g\|_{L_s} \leq M$, $s \in (1, \infty)$	$d^{-\frac{1}{2s}} M^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{q, w^{q'}}$
$p = 2$, $\|f\|_{L_\infty}, \|g\|_{L_\infty} \leq M$	$M^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{2, w^2}$

where

$$q = \frac{2s}{s-1} \in (\infty, 2) \text{ and } q' = \frac{2s}{s+1} \in (1, 2)$$

and

$$\|\hat{f} - \hat{g}\|_{q, w^r}^q = \sum_k \left| \frac{\hat{f}_k - \hat{g}_k}{2\pi |k|_r} \right|^q$$

over $k \in \mathbb{Z}^d \setminus 0$

Lower bounds

$p = 1$ and 2	$d^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{\infty, w^1}$
$p = 2,\ \ \|f\|_{L_s}, \|g\|_{L_s} \leq M,\ s \in (1, \infty)$	$d^{-\frac{1}{2s}} M^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{q, w^{q'}}$
$p = 2,\ \ \|f\|_{L_\infty}, \|g\|_{L_\infty} \leq M$	$M^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{2, w^2}$

- Tradeoff between
  - $M^{-\frac{1}{2}}$, decreasing as s increases, and
  - $d^{-\frac{1}{2s}}$, increasing with as s increases
- $\|\hat{f} - \hat{g}\|_{q, w^{q'}}$ increases as s increases
- The first bound works for unbounded densities

Upper bounds on $W_p(f, g)$

$p = 1$	$\|\hat{f} - \hat{g}\|_{2,w^2}$
$p = 1$, $\|f - g\|_{\dot{\mathcal{H}}^\beta} \leq z$	$O\left(\sqrt{z}\|\hat{f} - \hat{g}\|_{\infty,w^2}^{\frac{1}{2}}\right)$
$p = 2$, $f \wedge g \geq \xi > 0$	$O\left(\xi^{-\frac{1}{2}}\|\hat{f} - \hat{g}\|_{2,w^2}\right)$

- We require $\beta > \frac{d}{2} - 1$ for

$$\|f - g\|_{\dot{\mathcal{H}}^\beta}^2 := \sum_k (|k|^\beta(\hat{f}_k - \hat{g}_k))^2$$

- The $p = 1$ bounds works for densities that are not uniformly bounded from below.

Early stopping in computational inversion

- Assume the forward model

$$f : \mathcal{M} \mapsto \mathcal{P}(\Omega)$$

  is invertible as a function of its parameters
- Approximate the exact solution of

$$f(m) = g$$

  by an approximation m_δ s.t.

$$W_p(f(m_\delta), g) \leq \delta$$

Low-pass approximation of the inverse

- Denote the noisy data and the noise by

$$g^{\delta} := f(m_{\delta}) \quad \text{and} \quad n := g^{\delta} - g$$

- We cut-off frequencies $|k| > k_c$ of the approximate solution

$$m_{\delta} = f^{-1}(g^{\delta})$$

- The approximate bandwidth-limited solution is

$$m_{\delta}^{c} := \beta(g^{\delta})$$

  where $\beta = L \circ f^{-1}$ and $L : \mathcal{M} \to \mathcal{M}$ is the corresponding low-pass filter
- **What is he optimal cut-off threshold k_c?**
- **How does it depend on p and δ?**

Approximation error

- If $m \in \dot{\mathcal{H}}^r$ for $r > 0$,

$$\begin{aligned}\|m - m_\delta^c\|_{L^2} &\leq \|m - L \circ f^{-1} \circ f(m)\|_{L^2} + \|\beta \circ f(m) - m_\delta^c\|_{L^2} \\ &\leq \|m - L(m)\|_{L^2} + \|\beta(g) - \beta(g_\delta)\|_{L^2} \\ &\leq (2\pi k_c)^{-r}\|m\|_{\dot{\mathcal{H}}^r} + \|\beta(g) - \beta(g_\delta)\|_{L^2}\end{aligned}$$

- Assume for some $\alpha > 0$,

$$\|\beta(x) - \beta(y)\|_{L^2} \asymp k_c^\alpha \|\hat{x} - \hat{y}\|_{\ell_{q,w^2}} \tag{1}$$

  i.e., β is a *de-smoothing operator*
- Then the bound is minimized by

$$k_c^{\alpha+r} \asymp \frac{(2\pi)^{-r} r}{\alpha} \cdot \text{SNR}$$

  where

$$\text{SNR} := \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta_q}$$

De-smoothing inversion

- Assume $\|n\|_{\dot{\mathcal{H}}^\beta} \leq z$, and for some constants $\alpha', \epsilon > 0$ and function $h > 0$.

$$\|\beta(x) - \beta(y)\|_{L^2} \asymp k_c^{\alpha'} h(z) \|\hat{x} - \hat{y}\|_{\ell_{q,w^2}}^{\epsilon}$$

- If $r > 0$,

$$k_c^{\alpha'+r} \asymp \frac{(2\pi)^{-r} r}{\alpha'} \cdot \mathrm{SNR}_z$$

where

$$\mathrm{SNR}_z := \frac{\|m\|_{\dot{\mathcal{H}}^r}}{h(z)\delta_q^{\epsilon}}$$

Lower bounds on $k_c^{\alpha+r}$ or $k_c^{\alpha'+r}$,

$p = 1$	$\Omega\left(\frac{(2\pi)^{-r} r}{\alpha} \cdot \mathsf{SNR}\right)$
$p = 1,\ \|f(m) - g\|_{\dot{\mathcal{H}}^\beta} \leq z$	$\Omega\left(\frac{(2\pi)^{-r} r}{\alpha'} \cdot \mathsf{SNR}_z\right)$
$p = 2,\ f(m) \wedge g \geq \xi > 0$	$\Omega_\xi\left(\frac{(2\pi)^{-r} r}{\alpha} \cdot \mathsf{SNR}\right)$

- α and α' may depend on d, p, s
- α' may also depend on β.

Upper bounds on $k_c^{\alpha+r}$

$p = 1$	$O\left(d^{\frac{1}{2}} \frac{(2\pi)^{-r} r}{\alpha} \cdot \text{SNR}\right)$
$p = 2$, $\|f(m)\|_{L_s}, \|g\|_{L_s} \leq M$, $s \in (1, \infty)$	$O_M\left(d^{-\frac{1}{2s}} \frac{(2\pi)^{-r} r}{\alpha} \cdot \text{SNR}\right)$
$p = 2$, $\|f(m)\|_{L_\infty}, \|g\|_{L_\infty} \leq M$	$O_M\left(\frac{(2\pi)^{-r} r}{\alpha} \cdot \text{SNR}\right)$

- As before, $q = \frac{2s}{s-1}$
- Here too α may depend on d, p, s

Diagonal operators in Fourier domain

- If $\widehat{f_k} = \|k\|_2^{-\gamma}$, then

$$\alpha = \begin{cases} 1 + \gamma + \frac{d}{2} - \frac{d}{q} & \text{if } \gamma > -1 \\ \frac{d}{2} - \frac{d}{q} & \text{if } \gamma \leq -1 \end{cases}$$

- For $p = 1$, $q = \infty$

$$\alpha = \begin{cases} 1 + \gamma + \frac{d}{2} & \text{if } \gamma > -1 \\ \frac{d}{2} & \text{if } \gamma \leq -1 \end{cases}$$

- For $p = 2$, $q = 2$

$$\alpha = \begin{cases} 1 + \gamma & \text{if } \gamma > -1 \\ 0 & \text{if } \gamma \leq -1 \end{cases}$$

Diagonal operators in Fourier domain

- If $\|n\|_{\dot{\mathcal{H}}_\beta} \leq z$, then

$$\|Bn\|_{\ell_{1,w^2} \to L^2} \lesssim k_c^{\alpha'} \sqrt{z\delta_q}$$

where

$$\alpha' = \begin{cases} 1 + 2\gamma - \beta + \frac{d}{2} - \frac{d}{q} & \text{if } \gamma > \beta/2 - 1/2 \\ \frac{d}{2} - \frac{d}{q} & \text{if } \gamma \leq \beta/2 - 1/2 \end{cases} \tag{2}$$

- For $p = 1$

$$\alpha' = \begin{cases} 1 + 2\gamma - \beta + \frac{d}{2} & \text{if } \gamma > \beta/2 - 1/2 \\ \frac{d}{2} & \text{if } \gamma \leq \beta/2 - 1/2 \end{cases}$$

- For $p = 2$

$$\alpha' = \begin{cases} 1 + 2\gamma - \beta & \text{if } \gamma > \beta/2 - 1/2 \\ 0 & \text{if } \gamma \leq \beta/2 - 1/2 \end{cases}$$

- We assume that γ, d and q and if applicable β are such that $\alpha > 0$ or $\alpha' > 0$

Diagonal operators in Fourier domain - W_p vs $\mathcal{H}^\beta$

- Engquist et al. (2020) showed that in the context of using the $\mathcal{H}^\beta$ matching
$$k_c \asymp SNR^{\frac{1}{1+r+\gamma-\beta}}.$$
- Taking $\beta = 0$, corresponds to the L_2 norm matching
  - Leads to lower resolution (smoother reconstruction) than the W_p metric minimization
- Using the negative $\beta = \frac{d}{2} - \frac{d}{p}$ leads to the same resolution for $\mathcal{H}^\beta$ and W_p matching for $p \in [1, 2]$ (in each case, assuming that the early stopping thresholds are the same).

Qualitative results for diagonal operators in Fourier domain

Holding the early stopping threshold δ constant,

- all p: bounds on k_c constant or increase if $p \uparrow$
  - depends on regularity of the data g
- $p \in [1, 2]$: bounds on k_c constant or decrease if $d \uparrow$
  - also depends on regularity of g
- $p \in (2, \infty)$: bounds on k_c increase if $d \uparrow$

More qualitative results

Now holding the noise n constant,

- As we $\downarrow p$, by the monotonicity of ℓ_{q,w^2}, $\|\hat{n}\|_{p',w^2} \downarrow$
- We show $k_c \uparrow$ when $p \downarrow$ by constructing specific high-frequency noise

Generalizations

- Upper bounds on $\mathbb{H}^d$ have additional terms due to boundary condition
- We also provide similar Fourier based bounds for generalized W_p for unbalanced measures Piccoli and Rossi (2000); Piccoli et al. (2023)
  - The only difference is appearance of constant frequency $k = 0$
  - for $p = 1$, this is the KR norm used in FWI experiments mentioned earlier

Lower bounds proof

1. Construct a test function $h \in W^{1,q}$ from $\hat{f} - \hat{g}$ and Fourier basis functions
   - We adapt this idea from Niles-Weed and Berthet (2022) establishing wavelet based bounds on W_p
2. Bound $\|\nabla h\|_{L_q(\Omega)}$ using Hausdorff-Young inequality
3. Responsible for ℓ_p' vs ℓ_p for wavelets

Maury et al. (2010) For all $h \in W^{1,q}$, if μ and $\nu \in \mathcal{P}(\Omega) \cap L^s(\Omega)$ and $\|\nu\|_{L^s}, \|\mu\|_{L^s} \leq M$ and

$$\frac{1}{q} + \frac{1}{p} + \frac{1}{s} = 1 + \frac{1}{ps}$$

then

$$\int_\Omega h \; d(\mu - \nu) \leq M^{1/p'} \|\nabla h\|_{L_q(\Omega)} W_p(\mu, \nu)$$

Upper bounds proof

1. Construct feasible (ρ, E) using $\hat{f} - \hat{g}$ and Fourier basis
   - Also adapt this idea from Niles-Weed and Berthet (2022)
2. Develop Sobolev-like embedding of sequences
   - Higher regularity of $f - g$ allows to embed ℓ_{2,w^2} in a space with a stronger norm ℓ_{q,w^2} for $q > 2$.

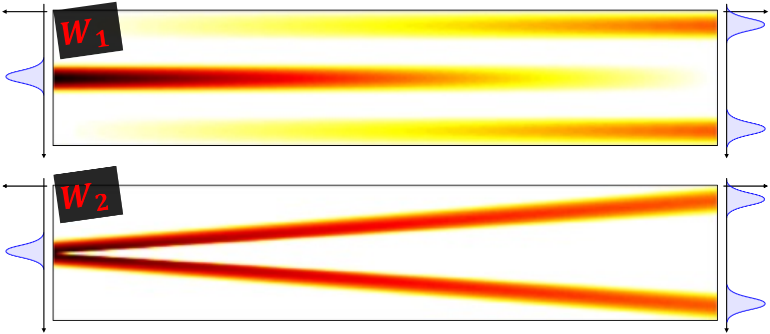
Fluid dynamics formulation Benamou and Brenier (2000); Brenier (2003): For $p \in [1, \infty)$,

$$W_p^p(\mu, \nu) = \inf_{(\rho, E)} \Big\{ \mathcal{B}_p(\rho, E) : \rho(\cdot, 1) = \mu, \rho(\cdot, 0) = \nu \\ \partial_t \rho + \nabla_x \cdot E = 0 \Big\}$$

where

$$\mathcal{B}_p(\rho, E) := \begin{cases} \int_{\Omega \times [0,1]} \| \frac{dE}{d\rho}(x,t) \|^p d\rho(x,t) & \text{if } E \ll \rho \\ +\infty & \text{otherwise.} \end{cases}$$

Dynamic formulation



Conclusion

- Offer a fresh Fourier-based perspective on Wasserstein-p distance resolving an open problem in analysis and probability
- Determine resolution of frequencies in computational inversion using W_p as the mismatch functional
  - Our analysis extends to nonlinear inverse problems
- Our bounds provide a leading order approximation of W_p
- Current work on understanding the higher order effects
  - convexity of W_p mismatch minimization
  - regularity of iterative solutions in discrete-time W_p minimization schemes
- Fundamental relationship between W_p and Fourier-based norms
- Expect many other connections in analysis, probability and applied fields

Acknowledgements

- We thank the anonymous referees for their useful comments (NeurIPS OTML 2023).
- This work is partially supported by the NSF grants DMS-1937254 and DMS-2309802.
- V.A.K. is grateful to Robert Kohn for his fruitful suggestion to explore wavelet analysis in this context.
- V.A.K. thanks Sinan Gunturk, James Scott and Stefan Steinerberger for helpful input at critical points in this project.

Image credits: Flamary and Courty (2019)

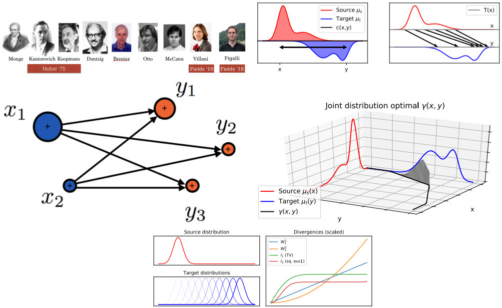
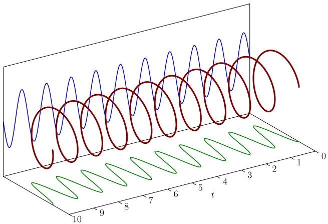
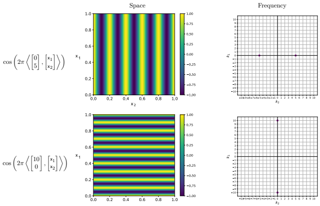


Image credits Fernandez-Granda (2020)





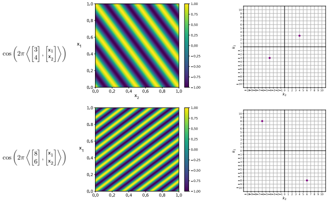
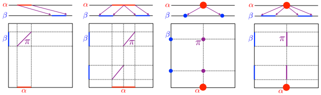
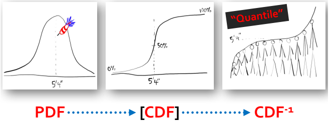


Image credits



Peyré and Cuturi (2018)



real gl (2013)

References I

Auricchio, G., CodeGoni, A., Gualandi, S., Toscani, G., and Veneroni, M. (2020). The equivalence of Fourier-based and Wasserstein metrics on imaging problems. *arXiv:2005.06530v1*.

Benamou, J.-D. and Brenier, Y. (2000). A computational fluid mechanics solution to the monge- kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393.

Brenier, Y. (2003). Extended Monge-Kantorovich theory. In Franca, M., editor, *Optimal transportation and applications*, volume 1813 of *Lecture Notes in Math.*, pages 91–121. Springer, Berlin.

Engquist, B., Ren, K., and Yang, Y. (2020). The quadratic Wasserstein metric for inverse data matching. *Inverse Problems*, 36:055001. arXiv:1911.06911.

Fernandez-Granda, C. (2020). Ds-ga 1013 / math-ga 2821 mathematical tools for data science, lecture notes. `https://cims.nyu.edu/~cfgranda/pages/MTDS_spring19/`.

References II

Flamary, R. and Courty, N. (2019). Optimal transport for machine learning: Introduction to optimal transport, tutorial isbi 2019, venice, italy. https://remi.flamary.com/cours/otml/OTML_ISBI_2019_OTintro.pdf.

Hong, W., Kobzar, V., and Ren, K. (2023). Fourier-based bounds for Wasserstein distances and their implications in computational inversion. In *NeurIPS 2023 Workshop Optimal Transport and Machine Learning*.

Kantorovich, L. V. (1942). On the translocation of masses. 37(7–8):227–229.

Lellmann, J., Lorenz, D., Schönlieb, C., and Valkonen, T. (2014). Imaging with Kantorovich-Rubinstein discrepancy. *SIAM J. Imag. Sci.*, 7:2833–2859.

Maury, B., Roudneff-Chupin, A., and Santambrogio, F. (2010). A macroscopic crowd motion model of gradient flow type. *Math. Models Methods Appl. Sci.*, 10(20):1787—-1821.

References III

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016a). Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion. *Geophys. J. Int.*, 205:345–377.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016b). An optimal transport approach for seismic tomography: application to 3D full waveform inversion. *Inverse Probl.*, 32. 115008.

Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année.

Métivier, L., Brossier, R., Kpadonou, F., Messud, J., and Pladys, A. (2022). A review of the use of optimal transport distances for high resolution seismic imaging based on the full waveform.

References IV

Niles-Weed, J. and Berthet, Q. (2022). Minimax estimation of smooth densities in Wasserstein distance. *Annals of Statistics*, (3):1519–1540.

Peyre, R. (2018). Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of wasserstein distance. *ESAIM: COCV*, 24(4):1489–1501.

Peyré, G. and Cuturi, M. (2018). *Computational Optimal Transport*. ArXiv:1803.00567.

Piccoli, B. and Rossi, F. (2000). On properties of the generalized Wasserstein distance. *Archive for Rational Mechanics and Analysis*, 222(3):1339– 1365.

Piccoli, B., Rossi, F., and Tournus, M. (2023). A Wasserstein norm for signed measures, with application to non-local transport equation with source term.

real gl (2013). pdf, cdf, inverse cdf, lecture notes. `http://realgl.blogspot.com/2013/01/pdf-cdf-inv-cdf.html`.

References V

Steinerberger, S. (2021). Wasserstein distance, Fourier series and applications. *Monatshefte für Mathematik*, 194:305–338.