Optimal Transport and Point Distributions on the Torus

Stefan Steinerberger

Point Distribution Webinar, September 2020

UNIVERSITY of WASHINGTON

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$?

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

There are many different ways of measuring this regularity. Certainly a very popular one is **discrepancy**.

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

There are many different ways of measuring this regularity. Certainly a very popular one is **discrepancy**. It is (1) geometrically meaningful

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

There are many different ways of measuring this regularity. Certainly a very popular one is **discrepancy**. It is (1) geometrically meaningful and (2) connected to practical applications via the Koksma-Hlawka inequality

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

There are many different ways of measuring this regularity. Certainly a very popular one is **discrepancy**. It is (1) geometrically meaningful and (2) connected to practical applications via the Koksma-Hlawka inequality

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq D_N(\mathbf{x}) \cdot \mathrm{Var}(f),$$

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

There are many different ways of measuring this regularity. Certainly a very popular one is **discrepancy**. It is (1) geometrically meaningful and (2) connected to practical applications via the Koksma-Hlawka inequality

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq D_N(\mathbf{x}) \cdot \mathrm{Var}(f),$$

where D_N is the discrepancy and Var denotes Hardy-Krause variation.

The Story in 2 Slides

Here's a classical problem: how do distribute sequences of points in the most regular way in $[0,1]^d$? Or how do we distribute sets of points most regularly?

There are many different ways of measuring this regularity. Certainly a very popular one is **discrepancy**. It is (1) geometrically meaningful and (2) connected to practical applications via the Koksma-Hlawka inequality

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq D_N(\mathbf{x}) \cdot \mathrm{Var}(f),$$

where D_N is the discrepancy and Var denotes Hardy-Krause variation. Hardy-Krause is tricky: it tends to grow exponentially in the dimension.

The Story in 2 Slides

The point of this talk is to discuss a new type of notion. I propose we look at something called the Wasserstein distance

$$W_1 = W_1\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \qquad \text{as a measure of regularity.}$$

The Story in 2 Slides

The point of this talk is to discuss a new type of notion. I propose we look at something called the Wasserstein distance

$$W_1 = W_1\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \qquad \text{as a measure of regularity.}$$

It is (1) geometrically meaningful

The Story in 2 Slides

The point of this talk is to discuss a new type of notion. I propose we look at something called the Wasserstein distance

$$W_1 = W_1\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \qquad \text{as a measure of regularity.}$$

It is (1) geometrically meaningful and (2) connected to practical applications via what is known as the Kantorovich-Rubinstein duality

The Story in 2 Slides

The point of this talk is to discuss a new type of notion. I propose we look at something called the Wasserstein distance

$$W_1 = W_1\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \qquad \text{as a measure of regularity.}$$

It is (1) geometrically meaningful and (2) connected to practical applications via what is known as the Kantorovich-Rubinstein duality

$$\left|\int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k)\right| \leq W_1 \cdot \|\nabla f\|_{L^\infty}.$$

The Story in 2 Slides

The point of this talk is to discuss a new type of notion. I propose we look at something called the Wasserstein distance

$$W_1 = W_1\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \qquad \text{as a measure of regularity.}$$

It is (1) geometrically meaningful and (2) connected to practical applications via what is known as the Kantorovich-Rubinstein duality

$$\left|\int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k)\right| \leq W_1 \cdot \|\nabla f\|_{L^\infty}.$$

Moreover, this inequality is sharp.

The Story in 2 Slides

The point of this talk is to discuss a new type of notion. I propose we look at something called the Wasserstein distance

$$W_1 = W_1\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \qquad \text{as a measure of regularity.}$$

It is (1) geometrically meaningful and (2) connected to practical applications via what is known as the Kantorovich-Rubinstein duality

$$\left|\int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k)\right| \leq W_1 \cdot \|\nabla f\|_{L^\infty}.$$

Moreover, this inequality is sharp. $\|\nabla f\|_{L^\infty}$ is, I would argue, a lot more natural than Hardy-Krause.

Suppose I want to sample a function $f : [0,1]^{100} \to \mathbb{R}$. It is known that there are point sets for which

$$D_N \sim \frac{(\log N)^{d-1}}{N}.$$

Suppose I want to sample a function $f : [0,1]^{100} \to \mathbb{R}$. It is known that there are point sets for which

$$D_N \sim \frac{(\log N)^{d-1}}{N}.$$

This function is actually increasing until $N \sim e^d$. Moreover, Hardy-Krause variation also tends to grow quite quickly.

Suppose I want to sample a function $f : [0,1]^{100} \to \mathbb{R}$. It is known that there are point sets for which

$$D_N \sim \frac{(\log N)^{d-1}}{N}.$$

This function is actually increasing until $N \sim e^d$. Moreover, Hardy-Krause variation also tends to grow quite quickly.

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} \right| \leq D_N(\mathbf{x}) \cdot \mathrm{Var}(f)$$

is not really useful until $N \gg d^d$.

Suppose I want to sample a function $f : [0,1]^{100} \to \mathbb{R}$. It is known that there are point sets for which

$$D_N \sim \frac{(\log N)^{d-1}}{N}.$$

This function is actually increasing until $N \sim e^d$. Moreover, Hardy-Krause variation also tends to grow quite quickly.

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} \delta_{x_k} \right| \leq D_N(\mathbf{x}) \cdot \mathrm{Var}(f)$$

is not really useful until $N \gg d^d$. In contrast,

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} \delta_{x_k} \right| \leq W_1 \cdot \|\nabla f\|_{L^\infty}$$

has no such hidden costs. The price: $W_1 \gtrsim N^{-1/d}$.

The Overall Goal

- What is Optimal Transport?

The Overall Goal

- What is Optimal Transport? More precisely, what is the Wasserstein Distance W_1?

The Overall Goal

- What is Optimal Transport? More precisely, what is the Wasserstein Distance W_1?
- Computing the Wasserstein Distance for some classical sequences

The Overall Goal

- What is Optimal Transport? More precisely, what is the Wasserstein Distance W_1?
- Computing the Wasserstein Distance for some classical sequences (which is a very nice thing: it's not some abstract quantity, it can actually be computed)

The Overall Goal

- What is Optimal Transport? More precisely, what is the Wasserstein Distance W_1?
- Computing the Wasserstein Distance for some classical sequences (which is a very nice thing: it's not some abstract quantity, it can actually be computed)
- What does this mean for Numerical Integration?

Gaspard Monge (1746 – 1818)

1781: 'Sur la théorie des déblais et des remblais'

Roughly: 'On the Theory of Rubble and Embankments'

Leonid Kantorovich (1912 – 1986)

Leonid Kantorovich (1912 – 1986)

The CIA File on Kantorovich (stolen from US Embassy in Tehran,

Leonid Kantorovich (1912 – 1986)

The CIA File on Kantorovich (stolen from US Embassy in Tehran, now on wikipedia)

Leonid Kantorovich (1912 – 1986)

The CIA File on Kantorovich (stolen from US Embassy in Tehran, now on wikipedia)

USSR **Leonid Vital'yevich KANTOROVICH**

Head, Problems Laboratory of Economic-Mathematical Methods and Operations Research, Institute of Management of the National Economy

(1975)

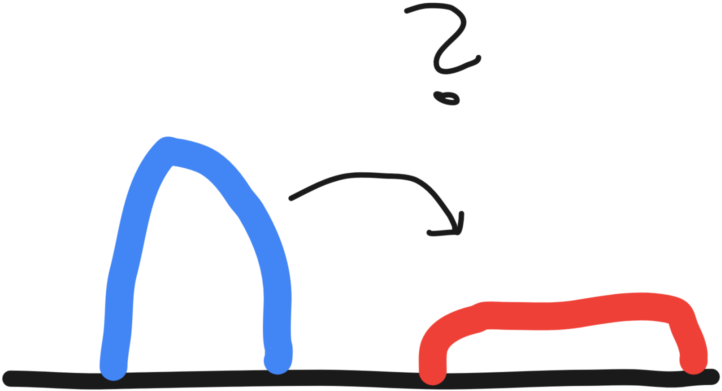
An internationally recognized creative genius in the fields of mathematics and the application of electronic computers to economic affairs, Academician Leonid Kantorovich (pronounced kahntuhROHvich) has worked at the Institute of Management of the National Economy since 1971. He has been involved in advanced mathematical research since the age of 15; in 1939 he invented linear programming, one of the most significant contributions to economic management in the twentieth century. Kantorovich has spent most of his adult life battling to win acceptance for his revolutionary concept from Soviet

Optimal Transport

Suppose we are given two measure μ and ν having same total mass and want to transport one to the other. (In all our applications, μ will be the measure induced by the points and $\nu = dx$.)

Optimal Transport

Suppose we are given two measure μ and ν having same total mass and want to transport one to the other. (In all our applications, μ will be the measure induced by the points and $\nu = dx$.)

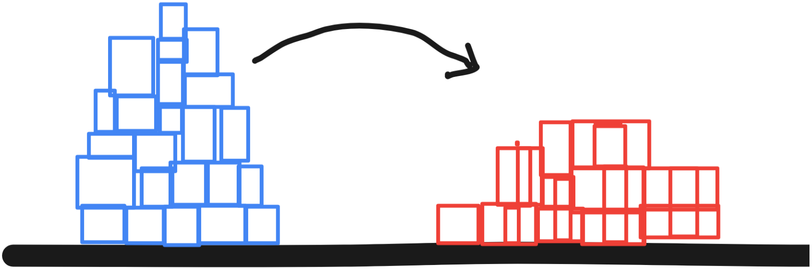


Optimal Transport

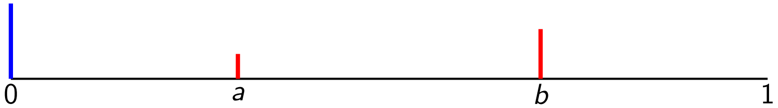
Think of both measures as being a collection of little boxes. Suppose it costs $\delta \cdot \varepsilon$ to move a box of weight ε distance δ. What is the cheapest way to move the boxes to the desired goal?

Optimal Transport

Think of both measures as being a collection of little boxes. Suppose it costs $\delta \cdot \varepsilon$ to move a box of weight ε distance δ. What is the cheapest way to move the boxes to the desired goal?

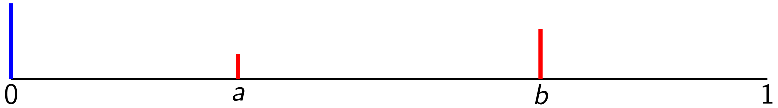


Wasserstein Distance



One unit of mass in 0 (blue), 1/3 unit of mass in a, 2/3 mass in b.

Wasserstein Distance

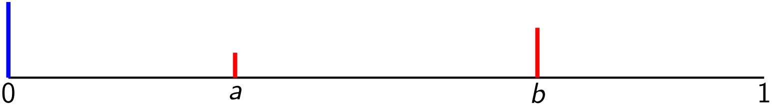


One unit of mass in 0 (blue), 1/3 unit of mass in a, 2/3 mass in b.

$$W_1(\mu, \nu) = \frac{a}{3} + \frac{2b}{3}$$

This is the *Earth Mover Distance*, the physical cost.

Wasserstein Distance



One unit of mass in 0 (blue), $1/3$ unit of mass in a, $2/3$ mass in b.

$$W_1(\mu,\nu) = \frac{a}{3} + \frac{2b}{3}$$

This is the *Earth Mover Distance*, the physical cost. There also exists an L^p-version of this, where $p > 1$, which leads to the $p-$Wasserstein distance

Wasserstein Distance

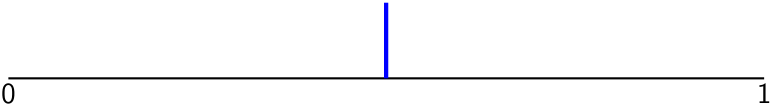
One unit of mass in 0 (blue), 1/3 unit of mass in a, 2/3 mass in b.

$$W_1(\mu, \nu) = \frac{a}{3} + \frac{2b}{3}$$

This is the *Earth Mover Distance*, the physical cost. There also exists an L^p-version of this, where $p > 1$, which leads to the $p-$Wasserstein distance

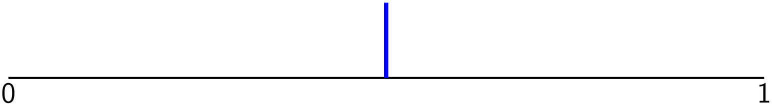
$$W_p(\mu, \nu) = \left(\frac{1}{3} a^p + \frac{2}{3} b^p \right)^{1/p}$$

Wasserstein Distance



One unit of mass in $1/2$ (blue). How much do I pay for the transport to dx?

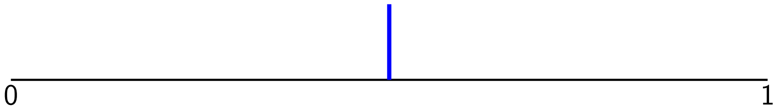
Wasserstein Distance



One unit of mass in 1/2 (blue). How much do I pay for the transport to dx?

$$W_1(\mu, dx) = \int_0^1 \left| x - \frac{1}{2} \right| dx = \frac{1}{4}.$$

Wasserstein Distance

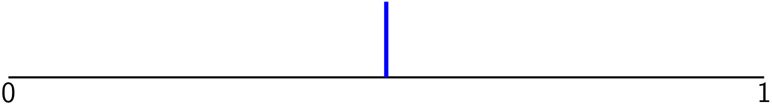


One unit of mass in 1/2 (blue). How much do I pay for the transport to dx?

$$W_1(\mu, dx) = \int_0^1 \left| x - \frac{1}{2} \right| dx = \frac{1}{4}.$$

$$W_p(\mu, \nu) = \left(\int_0^1 \left| x - \frac{1}{2} \right|^p dx \right)^{1/p} = \frac{1}{2} \frac{1}{(1+p)^{1/p}}.$$

Wasserstein Distance



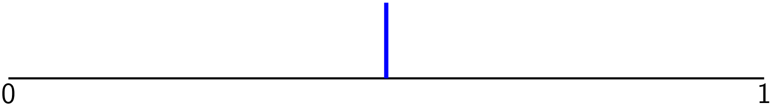
One unit of mass in 1/2 (blue). How much do I pay for the transport to dx?

$$W_1(\mu, dx) = \int_0^1 \left| x - \frac{1}{2} \right| dx = \frac{1}{4}.$$

$$W_p(\mu, \nu) = \left(\int_0^1 \left| x - \frac{1}{2} \right|^p dx \right)^{1/p} = \frac{1}{2} \frac{1}{(1+p)^{1/p}}.$$

Hölder's inequality implies that $W_p \geq W_1$.

Wasserstein Distance



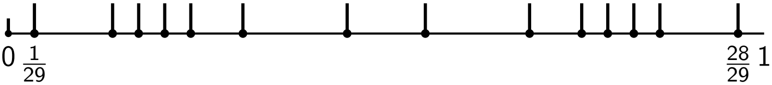
One unit of mass in 1/2 (blue). How much do I pay for the transport to dx?

$$W_1(\mu, dx) = \int_0^1 \left| x - \frac{1}{2} \right| dx = \frac{1}{4}.$$

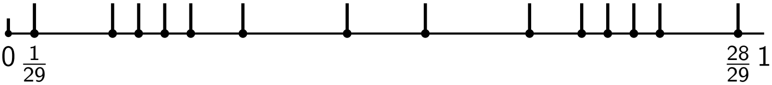
$$W_p(\mu, \nu) = \left(\int_0^1 \left| x - \frac{1}{2} \right|^p dx \right)^{1/p} = \frac{1}{2} \frac{1}{(1+p)^{1/p}}.$$

Hölder's inequality implies that $W_p \geq W_1$. For this talk: feel free to replace everything by W_1 (in fact, I assume that for most of the talk the W_1 and the W_2 behave similarly).

The Quadratic Residues in $\mathbb{F}_{29}$

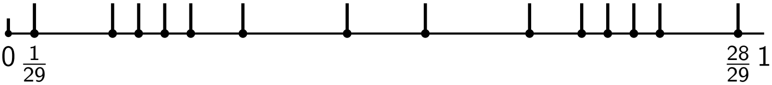


The Quadratic Residues in $\mathbb{F}_{29}$



$$0, 1, 1, 4, 4, 5, 5, 6, 6, 7, 7, 9, 9, 13, 13, \ldots$$

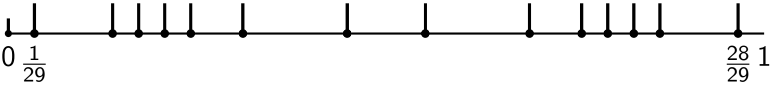
The Quadratic Residues in $\mathbb{F}_{29}$



$$0, 1, 1, 4, 4, 5, 5, 6, 6, 7, 7, 9, 9, 13, 13, \ldots$$

$$W_p\left(\frac{1}{29}\sum_{k=0}^{28}\delta_{\frac{k^2 \bmod 29}{29}},\ dx\right) \leq ?$$

The Quadratic Residues in $\mathbb{F}_{29}$



$$0, 1, 1, 4, 4, 5, 5, 6, 6, 7, 7, 9, 9, 13, 13, \ldots$$

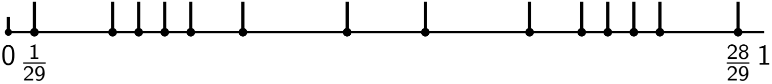
$$W_p\left(\frac{1}{29}\sum_{k=0}^{28}\delta_{\frac{k^2 \mod 29}{29}},\ dx\right) \leq ?$$

Theorem (S. 2018)

For primes p

$$W_2\left(\frac{1}{p}\sum_{k=0}^{p-1}\delta_{\frac{k^2 \mod p}{p}},\ dx\right) \lesssim \frac{1}{\sqrt{p}}$$

The Quadratic Residues in $\mathbb{F}_{29}$



$$0, 1, 1, 4, 4, 5, 5, 6, 6, 7, 7, 9, 9, 13, 13, \ldots$$

$$W_p \left(\frac{1}{29} \sum_{k=0}^{28} \delta_{\frac{k^2 \mod 29}{29}}, \ dx \right) \leq ?$$

Theorem (S. 2018)

For primes p

$$W_2 \left(\frac{1}{p} \sum_{k=0}^{p-1} \delta_{\frac{k^2 \mod p}{p}}, \ dx \right) \lesssim \frac{1}{\sqrt{p}}$$

This tells us that we have to move most particles roughly distance $\sim p^{-1/2}$. This is in line with the heuristic that these are ‘random’.

The Quadratic Residues in $\mathbb{F}_p$

Theorem (S. 2018)

For primes p

$$W_2\left(\frac{1}{p}\sum_{k=0}^{p-1}\delta_{\frac{k^2 \mod p}{p}},\ dx\right) \lesssim \frac{1}{\sqrt{p}}$$

It is natural to compare this to the **discrepancy**

$$\text{disc} = \sup_{0<a<b<1}\left|\frac{\#\left\{0 \le i \le p-1 : a \le \frac{i^2 \mod p}{p} \le b\right\}}{p} - (b-a)\right|$$

The Quadratic Residues in $\mathbb{F}_p$

Theorem (S. 2018)

For primes p

$$W_2\left(\frac{1}{p}\sum_{k=0}^{p-1}\delta_{\frac{k^2 \mod p}{p}},\ dx\right) \lesssim \frac{1}{\sqrt{p}}$$

It is natural to compare this to the **discrepancy**

$$\text{disc} = \sup_{0<a<b<1}\left|\frac{\#\left\{0 \le i \le p-1 : a \le \frac{i^2 \mod p}{p} \le b\right\}}{p} - (b-a)\right|$$

Theorem

$$\text{disc} \lesssim \frac{\log p}{\sqrt{p}} \qquad \text{(Polya-Vinogradov)}$$

$$\text{disc} \lesssim \frac{\log\log p}{\sqrt{p}} \qquad \text{(Vaughan-Montgomery (GRH))}$$

The Quadratic Residues in $\mathbb{F}_p$

Theorem (Cole Graham 2020)

For primes p and $2 < q < \infty$

$$W_q\left(\frac{1}{p}\sum_{k=0}^{p-1}\delta_{\frac{k^2 \mod p}{p}}\right) \lesssim \frac{1}{\sqrt{p}}$$

The Quadratic Residues in $\mathbb{F}_p$

Theorem (Cole Graham 2020)

For primes p and $2 < q < \infty$

$$W_q\left(\frac{1}{p}\sum_{k=0}^{p-1}\delta_{\frac{k^2 \mod p}{p}}\right) \lesssim \frac{1}{\sqrt{p}}$$

He also pointed out that

$$W_2\left(\frac{1}{p}\sum_{k=0}^{p-1}\delta_{\frac{k^2 \mod p}{p}}\right) \geq \frac{1}{\sqrt{12p}}$$

which shows that this result is sharp.

Irrational Rotations: Kronecker sequences

Theorem (S 2018)

$$W_2 \left(\frac{1}{N} \sum_{n=1}^{N} \delta_{\sqrt{2} \cdot n \mod 1}, \ dx \right) \lesssim \frac{\sqrt{\log N}}{N}$$

Irrational Rotations: Kronecker sequences

Theorem (S 2018)

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{\sqrt{2}\cdot n \mod 1},\ dx\right) \lesssim \frac{\sqrt{\log N}}{N}$$

We also have the classical result (Bohr? Weyl?)

$$D_N \lesssim \frac{\log N}{N}.$$

Irrational Rotations: Kronecker sequences

Theorem (S 2018)

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{\sqrt{2}\cdot n \mod 1},\ dx\right) \lesssim \frac{\sqrt{\log N}}{N}$$

We also have the classical result (Bohr? Weyl?)

$$D_N \lesssim \frac{\log N}{N}.$$

Theorem (Cole Graham 2020)

For every $(x_n)_{n=1}^{\infty}$ in $[0,1]$, there are infinitely many N such that

$$W_1\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}, dx\right) \geq c\frac{\sqrt{\log N}}{N}$$

Something Very Nice

Theorem (R. Peyre, 2018)

$$W_2(\mu, dx) \lesssim \|\mu\|_{\dot{H}^{-1}}$$

Something Very Nice

Theorem (R. Peyre, 2018)

$$W_2(\mu, dx) \lesssim \|\mu\|_{\dot{H}^{-1}}$$

If

$$\mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_k},$$

Something Very Nice

Theorem (R. Peyre, 2018)

$$W_2(\mu, dx) \lesssim \|\mu\|_{\dot{H}^{-1}}$$

If

$$\mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_k},$$

then

$$W_2(\mu, dx) \lesssim \left(\sum_{\ell \neq 0} \frac{1}{\ell^2} \left| \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i \ell x_l} \right|^2 \right)^{1/2}.$$

Something Very Nice

Theorem (R. Peyre, 2018)

$$W_2(\mu, dx) \lesssim \|\mu\|_{\dot{H}^{-1}}$$

If

$$\mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_k},$$

then

$$W_2(\mu, dx) \lesssim \left(\sum_{\ell \neq 0} \frac{1}{\ell^2} \left| \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i \ell x_l} \right|^2 \right)^{1/2}.$$

This is reminiscent of the Erdős-Turan inequality.

Summary

- Wasserstein Distance gives us yet another perspective on the (ir-)regularity of distributions...

Summary

- Wasserstein Distance gives us yet another perspective on the (ir-)regularity of distributions...
- ... and it is cheap to compute! It's classical exponential sum estimates

$$W_2(\mu, dx) \lesssim \left(\sum_{\ell \neq 0} \frac{1}{\ell^2} \left| \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i \ell x_l} \right|^2 \right)^{1/2}.$$

Summary

- Wasserstein Distance gives us yet another perspective on the (ir-)regularity of distributions...
- ... and it is cheap to compute! It's classical exponential sum estimates

$$W_2(\mu, dx) \lesssim \left(\sum_{\ell \neq 0} \frac{1}{\ell^2} \left| \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i \ell x_l} \right|^2 \right)^{1/2}.$$

Cole Graham (arXiv:1910.14181) has similar results on the torus, Bence Borda (arXiv:2005.04925) on compact Lie groups.

$$W_2(\mu, dx) \lesssim \left(\sum_{\ell \neq 0} \frac{1}{\ell^2} \left| \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i \ell x_l} \right|^2 \right)^{1/2}.$$

The upper bound is also known as **Zinterhof's diaphony**.

$$W_2(\mu, dx) \lesssim \left(\sum_{\ell \neq 0} \frac{1}{\ell^2} \left| \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i \ell x_l} \right|^2 \right)^{1/2}.$$

The upper bound is also known as **Zinterhof's diaphony**. This allows us to easily deal with the van der Corput sequence

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \dots$$

Theorem (Proinov)

For the van der Corput sequence

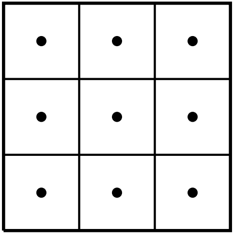
$$W_2\left(\frac{1}{N}\sum_{n=1}^{N} \delta_{x_n}, dx\right) \lesssim \frac{\sqrt{\log N}}{N}$$

The Coffee Shop Problem: Irregularities of Distributions

How to place your coffee shops?

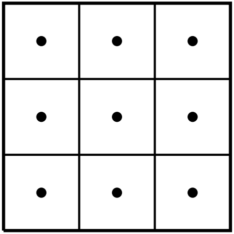
The Coffee Shop Problem: Irregularities of Distributions

How to place your coffee shops?



The Coffee Shop Problem: Irregularities of Distributions

How to place your coffee shops?

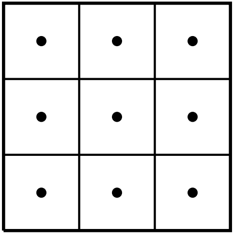


Question. Is there a sequence $(x_n)_{n=1}^{\infty}$ on $[0,1]^2$ such that

$$W_2\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \lesssim N^{-1/2} \ ?$$

The Coffee Shop Problem: Irregularities of Distributions

How to place your coffee shops?

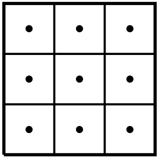


Question. Is there a sequence $(x_n)_{n=1}^{\infty}$ on $[0,1]^2$ such that

$$W_2\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \lesssim N^{-1/2} \ ?$$

(Recall, Cole Graham: on $[0,1]$, no sequence has $\lesssim N^{-1}$.)

The Coffee Shop Problem

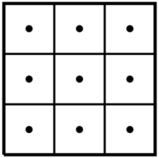


Theorem (Louis Brown and S, 2019)

Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be badly approximable. Then the Kronecker sequence $x_k = k\alpha \mod 1$ satisfies

$$W_2\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \lesssim_{c_\alpha, d} N^{-1/d}$$

The Coffee Shop Problem



Theorem (Louis Brown and S, 2019)

Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be badly approximable. Then the Kronecker sequence $x_k = k\alpha \mod 1$ satisfies

$$W_2\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \lesssim_{c_\alpha, d} N^{-1/d}$$

In $d \geq 3$, this seems to be fairly easy to do. **Open Problem.** But $d = 2$ appears subtle, are there other constructions?

Something Quite Nice

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

Something Quite Nice

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

Peyre's estimate works but Dirac measures are no longer in $\dot{H}^{-1}$.

Something Quite Nice

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

Peyre's estimate works but Dirac measures are no longer in $\dot{H}^{-1}$.

This has an interesting analogue in Analytic Number Theory: **Zinterhof's Diaphony**.

Something Quite Nice

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

Peyre's estimate works but Dirac measures are no longer in $\dot{H}^{-1}$.

This has an interesting analogue in Analytic Number Theory: **Zinterhof's Diaphony**. For $\{x_1, \dots, x_N\} \subset [0,1]$, Zinterhof's diaphony F_N is given by

$$F_N = \left(\sum_{\ell \neq 0}\frac{1}{\ell^2}\left|\frac{1}{N}\sum_{k=1}^{N} e^{2\pi i \ell x_l}\right|^2\right)^{1/2}.$$

It has never been generalized to higher dimensions.

Again Exponential Sums!

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

Again Exponential Sums!

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

We use the triangle inequality

$$W_2\left(\mu, dx\right) \leq W_2\left(\mu, \mu_{\text{nice}}\right) + W_2(\mu_{\text{nice}}, dx).$$

Again Exponential Sums!

How does one get good estimates on

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim ?$$

We use the triangle inequality

$$W_2(\mu, dx) \leq W_2(\mu, \mu_{\text{nice}}) + W_2(\mu_{\text{nice}}, dx).$$

Theorem (Louis Brown and S, 2019)

For each $t > 0$,

$$\boxed{W_2(\mu, dx)^2 \lesssim_d \inf_{t>0}\left[t + \sum_{\substack{k\in\mathbb{Z}^d \\ k\neq 0}} \frac{e^{-\|k\|^2 t}}{\|k\|^2}\left|\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i\langle k, x_n\rangle}\right|^2\right]}$$

Open Problems

I think it could be interesting to revisit classical objects!
What about

Open Problems

I think it could be interesting to revisit classical objects!
What about

- the Halton sequence?

Open Problems

I think it could be interesting to revisit classical objects!
What about

- the Halton sequence?
- the Hammersley set?

Open Problems

I think it could be interesting to revisit classical objects!
What about

- the Halton sequence?
- the Hammersley set?
- Sobol?

Open Problems

I think it could be interesting to revisit classical objects!
What about

- the Halton sequence?
- the Hammersley set?
- Sobol?
- $(t, m, s)-$nets?

Open Problems

I think it could be interesting to revisit classical objects!
What about

- the Halton sequence?
- the Hammersley set?
- Sobol?
- $(t, m, s)-$nets?

Surely many of these objects satisfy

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim N^{-1/d}?$$

Open Problems

I think it could be interesting to revisit classical objects!
What about

- the Halton sequence?
- the Hammersley set?
- Sobol?
- $(t, m, s)-$nets?

Surely many of these objects satisfy

$$W_2\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \lesssim N^{-1/d}?$$

Some of them can probably be attacked via Exponential Sums?
Others (nets?) via explicit constructions?

Open Problems

I think it could be interesting to revisit classical objects!

We recall that

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq W_1 \left(\frac{1}{N} \sum_{n=1}^{N} \delta_{x_k}, dx \right) \cdot \|\nabla f\|_{L^\infty}.$$

What if the function is twice-differentiable? Or in other smoothness classes?

Open Problems

This is another classical problem: it is known that

$$D_N \lesssim \frac{(\log N)^{d-1}}{N}$$

and the implicit constants are your enemy.

Open Problems

This is another classical problem: it is known that

$$D_N \lesssim \frac{(\log N)^{d-1}}{N}$$

and the implicit constants are your enemy.

Theorem (Heinrich, Novak, Wasilkowski, Wozniakowski, 2001)

There exist $\{x_1, \dots, x_N\} \subset [0,1]^d$ such that

$$D_N \le c\sqrt{\frac{d}{N}}.$$

Open Problems

This is another classical problem: it is known that

$$D_N \lesssim \frac{(\log N)^{d-1}}{N}$$

and the implicit constants are your enemy.

Theorem (Heinrich, Novak, Wasilkowski, Wozniakowski, 2001)

There exist $\{x_1, \dots, x_N\} \subset [0,1]^d$ such that

$$D_N \leq c\sqrt{\frac{d}{N}}.$$

Aistleitner: $c = 10$ works (since then other improvements).

Open Problems

Likewise, we have

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \leq \frac{\sqrt{d}}{N^{1/d}} \qquad \text{as} \quad N \to \infty$$

Open Problems

Likewise, we have

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \leq \frac{\sqrt{d}}{N^{1/d}} \qquad \text{as} \quad N \to \infty$$

But probably not for $N = 1000$?

Open Problems

Likewise, we have

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \leq \frac{\sqrt{d}}{N^{1/d}} \qquad \text{as} \quad N \to \infty$$

But probably not for $N = 1000$?

Question

Given N and d, how small can you make

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right)?$$

Open Problems

Likewise, we have

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \le \frac{\sqrt{d}}{N^{1/d}} \qquad \text{as} \quad N \to \infty$$

But probably not for $N = 1000$?

Question

Given N and d, how small can you make

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right)?$$

When N is large, some kind of lattice structure (sphere packing?) is presumably optimal (see also Hinrichs, Novak, Ullrich, Wozniakowski, 2016).

Open Problems

Likewise, we have

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right) \le \frac{\sqrt{d}}{N^{1/d}} \qquad \text{as} \quad N \to \infty$$

But probably not for $N = 1000$?

Question

Given N and d, how small can you make

$$W_p\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_k}, dx\right)?$$

When N is large, some kind of lattice structure (sphere packing?) is presumably optimal (see also Hinrichs, Novak, Ullrich, Wozniakowski, 2016).But $N = 1000$ in $d = 30$? ($2^{30} \gg 1000$)

A Final Application

The following is **very classical**. Let $f : [0,1]^d \to \mathbb{R}$. Then there are points $\{x_1, \dots, x_N\} \subset [0,1]^d$ such that

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

A Final Application

The following is **very classical**. Let $f : [0,1]^d \to \mathbb{R}$. Then there are points $\{x_1, \dots, x_N\} \subset [0,1]^d$ such that

$$\left| \int_{\mathbb{T}^d} f(x) dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

If you don't know anything about the function, this is clearly best possible. Take

$$f(x) = \min_{1 \leq i \leq n} \|x - x_i\|.$$

A Final Application

The following is **very classical**. Let $f : [0,1]^d \rightarrow \mathbb{R}$. Then there are points $\{x_1, \dots, x_N\} \subset [0,1]^d$ such that

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

If you don't know anything about the function, this is clearly best possible. Take

$$f(x) = \min_{1 \leq i \leq n} \|x - x_i\|.$$

The average distance from a point in $[0,1]^d$ to a point is $\sim N^{-1/d}$.

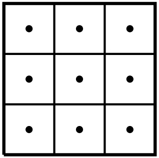
A Final Application

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

A Final Application

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

This suggests we take the points

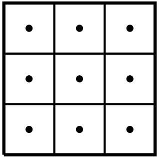


Sukharev (1979) showed that this leads to the smallest constant.

A Final Application

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \le c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

This suggests we take the points

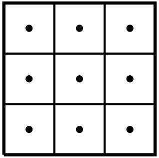


Sukharev (1979) showed that this leads to the smallest constant.
But what if we want to take a sequence?

A Final Application

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

This suggests we take the points

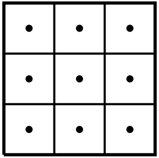


Sukharev (1979) showed that this leads to the smallest constant.
But what if we want to take a sequence? On-line sampling?

A Final Application

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \frac{\|\nabla f\|_{L^\infty}}{N^{1/d}}.$$

This suggests we take the points



Sukharev (1979) showed that this leads to the smallest constant.
But what if we want to take a sequence? On-line sampling? We do not know how many points we get?

Theorem (Louis Brown and S, 2019)

Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be a badly approximable vector. Then, for some universal $c_\alpha > 0$ and all differentiable $f : \mathbb{T}^d \to \mathbb{R}$

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(k\alpha) \right| \leq c_\alpha \|\nabla f\|_{L^\infty(\mathbb{T}^d)}^{(d-1)/d} \|\nabla f\|_{L^2(\mathbb{T}^d)}^{1/d} N^{-1/d}.$$

Theorem (Louis Brown and S, 2019)

Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be a badly approximable vector. Then, for some universal $c_\alpha > 0$ and all differentiable $f : \mathbb{T}^d \to \mathbb{R}$

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(k\alpha) \right| \leq c_\alpha \|\nabla f\|_{L^\infty(\mathbb{T}^d)}^{(d-1)/d} \|\nabla f\|_{L^2(\mathbb{T}^d)}^{1/d} N^{-1/d}.$$

- Uniformly for a sequence and

Theorem (Louis Brown and S, 2019)

Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be a badly approximable vector. Then, for some universal $c_\alpha > 0$ and all differentiable $f : \mathbb{T}^d \to \mathbb{R}$

$$\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(k\alpha) \right| \leq c_\alpha \|\nabla f\|_{L^\infty(\mathbb{T}^d)}^{(d-1)/d} \|\nabla f\|_{L^2(\mathbb{T}^d)}^{1/d} N^{-1/d}.$$

- Uniformly for a sequence and
- better L^p-spaces.

Theorem (Louis Brown and S, 2019)

Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be a badly approximable vector. Then, for some universal $c_\alpha > 0$ and all differentiable $f : \mathbb{T}^d \to \mathbb{R}$

$$\left| \int_{\mathbb{T}^d} f(x) dx - \frac{1}{N} \sum_{k=1}^{N} f(k\alpha) \right| \leq c_\alpha \| \nabla f \|_{L^\infty(\mathbb{T}^d)}^{(d-1)/d} \| \nabla f \|_{L^2(\mathbb{T}^d)}^{1/d} N^{-1/d}.$$

- Uniformly for a sequence and
- better L^p-spaces.

... this is **strange**. The grid should actually be the best....

Slight Improvement over a Classical Result

Theorem (Louis Brown and S, 2019)

We have, for some explicit constant c_d depending only on the dimension, for all differentiable $f : [0,1]^d \to \mathbb{R}$ sampled on the regular grid $(x_k)_{k=1}^N$

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \|\nabla f\|_{L^\infty(\mathbb{T}^d)}^{(d-1)/d} \|\nabla f\|_{L^1(\mathbb{T}^d)}^{1/d} N^{-1/d}.$$

Slight Improvement over a Classical Result

Theorem (Louis Brown and S, 2019)

We have, for some explicit constant c_d depending only on the dimension, for all differentiable $f : [0,1]^d \to \mathbb{R}$ sampled on the regular grid $(x_k)_{k=1}^N$

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq c_d \|\nabla f\|_{L^\infty(\mathbb{T}^d)}^{(d-1)/d} \|\nabla f\|_{L^1(\mathbb{T}^d)}^{1/d} N^{-1/d}.$$

This is sharp again (probably?): take $0 < \varepsilon \ll 1$ and

$$f(x) = \min\left\{\varepsilon, \min_{1 \leq i \leq N} \|x - x_i\|\right\}.$$

On Friday

One big issue with classical discrepancy is that it is adapted to the torus $\mathbb{T}^d$ (since we use axis-parallel rectangles). There are natural variations on the sphere (take spherical caps) but it's not clear what to do on a general manifold.

On Friday

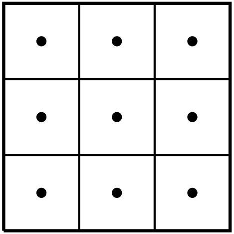
One big issue with classical discrepancy is that it is adapted to the torus $\mathbb{T}^d$ (since we use axis-parallel rectangles). There are natural variations on the sphere (take spherical caps) but it's not clear what to do on a general manifold.

In contrast, the Wasserstein distance does not care very much about the underlying background. This makes it a **stable** notion.

On Friday

One big issue with classical discrepancy is that it is adapted to the torus $\mathbb{T}^d$ (since we use axis-parallel rectangles). There are natural variations on the sphere (take spherical caps) but it's not clear what to do on a general manifold.

In contrast, the Wasserstein distance does not care very much about the underlying background. This makes it a **stable** notion. But there are lots of problems on, say, $\mathbb{S}^2$ as well, and we'll discuss some of them on Friday.



Thank you!