The (nonconvex) uniform optimal quantization problem.

Quentin Mérigot[1] Filippo Santambrogio[2] Clément Sarrazin[1]

[1]Université Paris-Saclay [2]Université Lyon 1

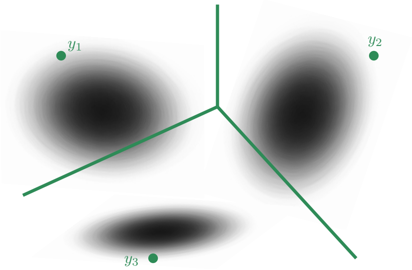
Dynamics and Discretization: PDEs, Sampling, and Optimization
Simons Institute, October 2021

Optimal quantization and Lloyd's algorithm

- Opt. quant. of a density $\rho \in \mathcal{P}(\Omega)$: $\boxed{\min_{y_1,\ldots,y_N} \min_{\alpha \in \Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}$

Optimal quantization and Lloyd's algorithm

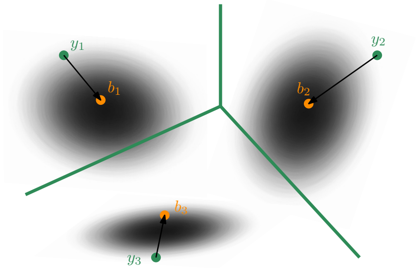
- Opt. quant. of a density $\rho \in \mathcal{P}(\Omega)$: $\boxed{\min_{y_1,\dots,y_N} \min_{\alpha \in \Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}$



- **Lloyd's algorithm**: given $Y = (y_1, \dots, y_N) \in \Omega^N$
  1. Compute the **Voronoi cells** of Y

Optimal quantization and Lloyd's algorithm

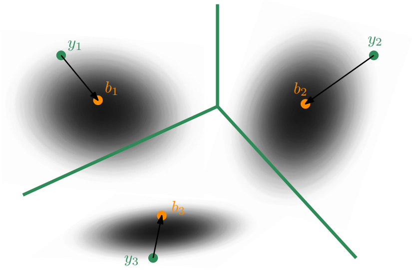
- Opt. quant. of a density $\rho \in \mathcal{P}(\Omega)$: $\boxed{\min_{y_1,\dots,y_N} \min_{\alpha \in \Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}$



- **Lloyd's algorithm**: given $Y = (y_1, \dots, y_N) \in \Omega^N$
  1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
  2. Set $y_i := b_i$ and repeat.

Optimal quantization and Lloyd's algorithm

- Opt. quant. of a density $\rho \in \mathcal{P}(\Omega)$: $\boxed{\min_{y_1,\dots,y_N} \min_{\alpha \in \Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}$

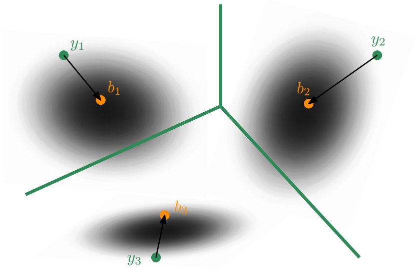


- **Lloyd's algorithm**: given $Y = (y_1, \dots, y_N) \in \Omega^N$
  1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
  2. Set $y_i := b_i$ and repeat.

- Used/studied intensively in computer science (e.g. clustering: N is small).

Optimal quantization and Lloyd's algorithm

- Opt. quant. of a density $\rho \in \mathcal{P}(\Omega)$: $\boxed{\min_{y_1,\dots,y_N} \min_{\alpha \in \Delta_N} \mathrm{W}_2^2 \left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}$

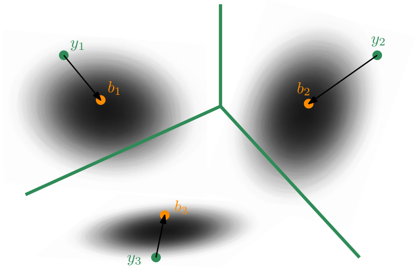


- **Lloyd's algorithm**: given $Y = (y_1, \dots, y_N) \in \Omega^N$
  1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
  2. Set $y_i := b_i$ and repeat.

- Used/studied intensively in computer science (e.g. clustering: N is small).
- Limit $N \to +\infty$: behaviour of minimizers ([Fejes Tóth], [Grüber], $d = 2$)

Optimal quantization and Lloyd's algorithm

- Opt. quant. of a density $\rho \in \mathcal{P}(\Omega)$: $\boxed{\min_{y_1,\dots,y_N} \min_{\alpha\in\Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}$



- **Lloyd's algorithm**: given $Y = (y_1, \dots, y_N) \in \Omega^N$
  1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
  2. Set $y_i := b_i$ and repeat.

- Used/studied intensively in computer science (e.g. clustering: N is small).
- Limit $N \to +\infty$: behaviour of minimizers ([Fejes Tóth], [Grüber], $d = 2$)
- Limit $N \to +\infty$: behaviour of Lloyd's algorithm [Iacobelli].

Optimal uniform quantization

- Approximate an image using a point cloud: [de Goes *et al.* 2012]

ρ

$y_1, \ldots, y_N$

Optimal uniform quantization

- Approximate an image using a point cloud: [de Goes *et al.* 2012]

ρ

$y_1, \dots, y_N$

- Optimal location problems [Bourne, Schmitzer, Wirth, 2018]

Optimal uniform quantization

- Approximate an image using a point cloud: [de Goes *et al.* 2012]

ρ

$y_1, \ldots, y_N$

- Optimal location problems [Bourne, Schmitzer, Wirth, 2018]
- Generation of polycrystalline microstructures [Bourne *et al.* 2020]

Motivation

- Let $\Omega \subseteq \mathbb{R}^d$ compact convex, and $\mathcal{P}(\Omega)$ = probability measures on Ω.

Motivation

- Let $\Omega \subseteq \mathbb{R}^d$ compact convex, and $\mathcal{P}(\Omega) =$ probability measures on Ω.
- Wasserstein distances W_p ($p \geq 1$) are geometric distances on $\mathcal{P}(\Omega)$, defined through optimal transport.

Motivation

- Let $\Omega \subseteq \mathbb{R}^d$ compact convex, and $\mathcal{P}(\Omega) =$ probability measures on Ω.
- Wasserstein distances W_p ($p \geq 1$) are geometric distances on $\mathcal{P}(\Omega)$, defined through optimal transport.
- W_p is used in *generative modeling*: if $\rho =$ real data and $\mu_\theta =$ generated data,

$$\min_\theta \mathrm{W}_1(\rho, \mu_\theta) \qquad \text{[Arjovsky \textit{et. al}, 2017,][Genevay \textit{et al.} 2018]}$$

Difficult **non-convex** optimization problem when the support of μ_θ depends on θ, e.g. when $\mu_\theta = \sum_{1 \leq i \leq N} \alpha_i(\theta) \delta_{x_i(\theta)}$

Motivation

- Let $\Omega \subseteq \mathbb{R}^d$ compact convex, and $\mathcal{P}(\Omega) =$ probability measures on Ω.
- Wasserstein distances W_p ($p \geq 1$) are geometric distances on $\mathcal{P}(\Omega)$, defined through optimal transport.
- W_p is used in *generative modeling*: if $\rho =$ real data and $\mu_\theta =$ generated data,

$$\min_\theta \mathrm{W}_1(\rho, \mu_\theta)$$ [Arjovsky *et. al*, 2017,][Genevay *et al.* 2018]

  Difficult **non-convex** optimization problem when the support of μ_θ depends on θ, e.g. when $\mu_\theta = \sum_{1 \leq i \leq N} \alpha_i(\theta)\delta_{x_i(\theta)}$
- Uniform optimal quantization = simple variant of this problem:

$$\min_{Y=(y_1,\ldots,y_N)\in\Omega^d} \mathrm{W}_2^2\left(\rho, \frac{1}{N}\sum_i \delta_{y_i}\right)$$

  Efficiently solved using gradient descent despite the non-convexity.

Motivation

- Let $\Omega \subseteq \mathbb{R}^d$ compact convex, and $\mathcal{P}(\Omega)$ = probability measures on Ω.
- Wasserstein distances W_p ($p \geq 1$) are geometric distances on $\mathcal{P}(\Omega)$, defined through optimal transport.
- W_p is used in *generative modeling*: if ρ = real data and μ_θ = generated data,

$$\min_\theta \mathrm{W}_1(\rho, \mu_\theta)$$

[Arjovsky *et. al*, 2017,][Genevay *et al.* 2018]

Difficult **non-convex** optimization problem when the support of μ_θ depends on θ, e.g. when $\mu_\theta = \sum_{1 \leq i \leq N} \alpha_i(\theta)\delta_{x_i(\theta)}$

- Uniform optimal quantization = simple variant of this problem:

$$\min_{Y=(y_1,\ldots,y_N)\in\Omega^d} \mathrm{W}_2^2\left(\rho, \frac{1}{N}\sum_i \delta_{y_i}\right)$$

Efficiently solved using gradient descent despite the non-convexity.

Does gradient descent lead to low-energy configurations?

Wasserstein distance

- 2-Wasserstein distance: $\mathrm{W}_2^2(\mu, \rho) = \min_{X \sim \mu, Y \sim \rho} \mathbb{E}(\|X - Y\|^2)$.

Wasserstein distance

- 2-Wasserstein distance: $\mathrm{W}_2^2(\mu, \rho) = \min_{X\sim\mu, Y\sim\rho} \mathbb{E}(\|X - Y\|^2)$.
- The uniform quantization energy involves a **semi-discrete** problem:

$$F_N : Y = (y_1, \ldots, y_N) \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$$

Wasserstein distance

- 2-Wasserstein distance: $\mathrm{W}_2^2(\mu, \rho) = \min_{X \sim \mu, Y \sim \rho} \mathbb{E}(\|X - Y\|^2)$.
- The uniform quantization energy involves a **semi-discrete** problem:

$$F_N : Y = (y_1, \ldots, y_N) \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$$

- By Kantorovich duality [Kantorovich 1942][Aurenhammer *et al.* 1998],

$$F_N(Y) = \max_{\Phi} \sum_i \alpha_i \phi_i + \int_{\Omega} \min_i \|x - y_i\|^2 - \phi_i \mathrm{d}\rho(x)$$

Wasserstein distance

- 2-Wasserstein distance: $\mathrm{W}_2^2(\mu, \rho) = \min_{X\sim\mu, Y\sim\rho} \mathbb{E}(\|X - Y\|^2)$.
- The uniform quantization energy involves a **semi-discrete** problem:

$$F_N : Y = (y_1, \ldots, y_N) \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$$

- By Kantorovich duality [Kantorovich 1942][Aurenhammer *et al.* 1998],

$$\begin{aligned} F_N(Y) &= \max_{\Phi} \sum_i \alpha_i \phi_i + \int_\Omega \min_i \|x - y_i\|^2 - \phi_i \mathrm{d}\rho(x) \\ &= \max_{\Phi} \sum_i \left[\alpha_i \phi_i + \int_{\mathrm{Lag}_i(Y,\Phi)} \|x - y_i\|^2 - \phi_i \mathrm{d}\rho(x)\right], \end{aligned}$$

where *Laguerre cells* are defined for $Y \in \Omega^N$ and $\Phi \in \mathbb{R}^N$:

$$\mathrm{Lag}_i(Y, \Phi) \stackrel{\text{def}}{=} \left\{x \in \mathbb{R}^d \;\middle|\; \forall j,\; \|x - y_i\|^2 - \phi_i \leq \|x - y_j\|^2 - \phi_j\right\}$$

Wasserstein distance

- 2-Wasserstein distance: $\mathrm{W}_2^2(\mu, \rho) = \min_{X\sim\mu, Y\sim\rho} \mathbb{E}(\|X - Y\|^2)$.
- The uniform quantization energy involves a **semi-discrete** problem:

$$F_N : Y = (y_1, \ldots, y_N) \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$$

- By Kantorovich duality [Kantorovich 1942][Aurenhammer *et al.* 1998],

$$F_N(Y) = \max_\Phi \sum_i \alpha_i\phi_i + \int_\Omega \min_i \|x - y_i\|^2 - \phi_i \mathrm{d}\rho(x)$$
$$= \max_\Phi \sum_i \left[\alpha_i\phi_i + \int_{\mathrm{Lag}_i(Y,\Phi)} \|x - y_i\|^2 - \phi_i \mathrm{d}\rho(x)\right],$$

where *Laguerre cells* are defined for $Y \in \Omega^N$ and $\Phi \in \mathbb{R}^N$:

$$\mathrm{Lag}_i(Y, \Phi) \stackrel{\text{def}}{=} \left\{x \in \mathbb{R}^d \;\middle|\; \forall j,\; \|x - y_i\|^2 - \phi_i \leq \|x - y_j\|^2 - \phi_j\right\}$$

- Given pairwise distinct points $Y \in \Omega^N$, the maximizer $\Phi_Y \in \mathbb{R}^N$ is unique and characterized by $\rho(\mathrm{Lag}_i(Y, \Phi_Y)) = \frac{1}{N}$: all cells have mass $\frac{1}{N}$.

Optimal quantization energy

- We minimize $F_N : Y \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$.
- W_2^2 is convex on $\mathcal{P}(\Omega)$, yet F_N is **not convex** on Ω^N.

Optimal quantization energy

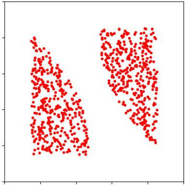
- We minimize $F_N : Y \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$.
- W_2^2 is convex on $\mathcal{P}(\Omega)$, yet F_N is **not convex** on Ω^N.

Proposition

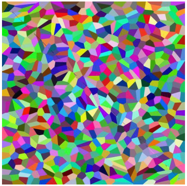
F_N is semi-concave on Ω^N, it is $\mathcal{C}^1$ on a dense open set and

$$F_N(Y) = \sum_i \int_{\mathrm{Lag}_i(Y,\Phi_Y)} \|x - y_i\|^2 \,\mathrm{d}\rho(x),$$

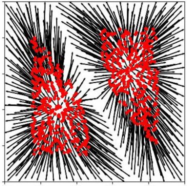
where $b_i(Y) = N \int_{\mathrm{Lag}_i(Y,\Phi_Y)} x d\rho(x)$ is the barycenter of the ith cell.



Point cloud Y



$\mathrm{Lag}_i(Y, \Phi_Y)$



$-\frac{N}{2}\nabla_{y_i} F_N(Y)$

Optimal quantization energy

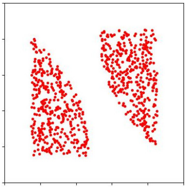
- We minimize $F_N : Y \in \Omega^N \mapsto \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)$.
- W_2^2 is convex on $\mathcal{P}(\Omega)$, yet F_N is **not convex** on Ω^N.

Proposition

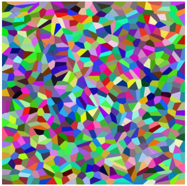
F_N is semi-concave on Ω^N, it is $\mathcal{C}^1$ on a dense open set and

$$F_N(Y) = \sum_i \int_{\mathrm{Lag}_i(Y,\Phi_Y)} \|x - y_i\|^2 \,\mathrm{d}\rho(x), \quad \nabla_{y_i} F_N(Y) = \frac{2}{N}(y_i - b_i(Y))$$

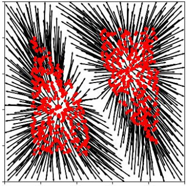
where $b_i(Y) = N \int_{\mathrm{Lag}_i(Y,\Phi_Y)} x d\rho(x)$ is the barycenter of the ith cell.



Point cloud Y



$\mathrm{Lag}_i(Y, \Phi_Y)$



$-\frac{N}{2}\nabla_{y_i} F_N(Y)$

Lloyd's (uniform) algorithm

Lloyd's algorithm

Optimal quant. of a density $\rho \in \mathcal{P}(\Omega)$,

$$\min_{Y} \underbrace{\min_{\alpha \in \Delta_N} \mathrm{W}_2^2 \left(\sum_i \alpha_i \delta_{y_i}, \rho \right)}_{G_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
2. Set $y_i := b_i$ and repeat.

Lloyd's (uniform) algorithm

Lloyd's algorithm

Optimal quant. of a density $\rho \in \mathcal{P}(\Omega)$,

$$\min_Y \underbrace{\min_{\alpha \in \Delta_N} \mathrm{W}_2^2 \left(\sum_i \alpha_i \delta_{y_i}, \rho \right)}_{G_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
2. Set $y_i := b_i$ and repeat.

Lloyd's "uniform" algorithm

Optimal uniform quantization of ρ,

$$\min_Y \underbrace{\mathrm{W}_2^2 \left(\frac{1}{N} \sum_i \delta_{y_i}, \rho \right)}_{F_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Laguerre cells** $\mathrm{Lag}_i(Y, \Phi_Y)$ solving the OT problem between ρ and $\mu = \frac{1}{N} \sum_i \delta_{y_i}$ and their **barycenters** $b_i(Y)$.
2. Set $y_i := b_i(Y)$ and repeat.

Lloyd's (uniform) algorithm

Lloyd's algorithm

Optimal quant. of a density $\rho \in \mathcal{P}(\Omega)$,

$$\min_Y \underbrace{\min_{\alpha \in \Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}_{G_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
2. Set $y_i := b_i$ and repeat.

Lloyd's "uniform" algorithm

Optimal uniform quantization of ρ,

$$\min_Y \underbrace{\mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)}_{F_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Laguerre cells** $\mathrm{Lag}_i(Y, \Phi_Y)$ solving the OT problem between ρ and $\mu = \frac{1}{N}\sum_i \delta_{y_i}$ and their **barycenters** $b_i(Y)$.
2. Set $y_i := b_i(Y)$ and repeat.

- Lloyd's algorithms = fixed point algorithms for cancelling ∇G_N or ∇F_N.

Lloyd's (uniform) algorithm

Lloyd's algorithm

Optimal quant. of a density $\rho \in \mathcal{P}(\Omega)$,

$$\min_Y \underbrace{\min_{\alpha \in \Delta_N} \mathrm{W}_2^2\left(\sum_i \alpha_i \delta_{y_i}, \rho\right)}_{G_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
2. Set $y_i := b_i$ and repeat.

Lloyd's "uniform" algorithm

Optimal uniform quantization of ρ,

$$\min_Y \underbrace{\mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right)}_{F_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Laguerre cells** $\mathrm{Lag}_i(Y, \Phi_Y)$ solving the OT problem between ρ and $\mu = \frac{1}{N}\sum_i \delta_{y_i}$ and their **barycenters** $b_i(Y)$.
2. Set $y_i := b_i(Y)$ and repeat.

- Lloyd's algorithms = fixed point algorithms for cancelling ∇G_N or ∇F_N.
- The iterates converge (up to subseq.) to a critical point of F_N or G_N.

Lloyd's (uniform) algorithm

Lloyd's algorithm

Optimal quant. of a density $\rho \in \mathcal{P}(\Omega)$,

$$\min_Y \underbrace{\min_{\alpha \in \Delta_N} \mathrm{W}_2^2 \left(\sum_i \alpha_i \delta_{y_i}, \rho \right)}_{G_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Voronoi cells** of Y and their **barycenters** b_i w.r.t. to ρ.
2. Set $y_i := b_i$ and repeat.

Lloyd's "uniform" algorithm

Optimal uniform quantization of ρ,

$$\min_Y \underbrace{\mathrm{W}_2^2 \left(\frac{1}{N} \sum_i \delta_{y_i}, \rho \right)}_{F_N(Y)}$$

Algorithm: given $Y \in \Omega^N$
1. Compute the **Laguerre cells** $\mathrm{Lag}_i(Y, \Phi_Y)$ solving the OT problem between ρ and $\mu = \frac{1}{N} \sum_i \delta_{y_i}$ and their **barycenters** $b_i(Y)$.
2. Set $y_i := b_i(Y)$ and repeat.

- Lloyd's algorithms = fixed point algorithms for cancelling ∇G_N or ∇F_N.
- The iterates converge (up to subseq.) to a critical point of F_N or G_N.
- In both cases, there may exist critical points with high energy

Low- and high-energy critical points of F_N

- Given ρ bounded from above and below on a bounded convex set $\Omega \in \mathbb{R}^d$,

$$\min_{\Omega^N} F_N = \Theta\left(\left(\frac{1}{N}\right)^{2/d}\right).$$

Low- and high-energy critical points of F_N

- Given ρ bounded from above and below on a bounded convex set $\Omega \in \mathbb{R}^d$,

$$\min_{\Omega^N} F_N = \Theta\left(\left(\frac{1}{N}\right)^{2/d}\right).$$

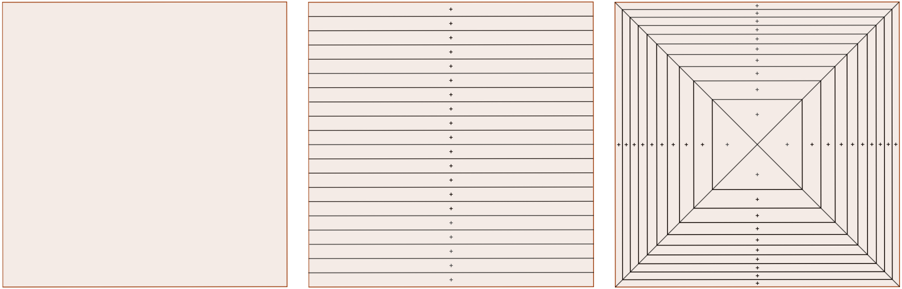
- Minimizers for F_N are **critical**, i.e. they satisfy $\boxed{\forall i,\ y_i = b_i(Y)}$.

Low- and high-energy critical points of F_N

- Given ρ bounded from above and below on a bounded convex set $\Omega \in \mathbb{R}^d$,

$$\min_{\Omega^N} F_N = \Theta\left(\left(\frac{1}{N}\right)^{2/d}\right).$$

- Minimizers for F_N are **critical**, i.e. they satisfy $\boxed{\forall i,\ y_i = b_i(Y)}$.
- Due to the non-convexity of F_N, some critical points are NOT minimizers:



ρ $\qquad$ $N = 20$ $\qquad$ $N = 40$

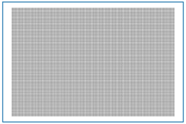
Figure: Two high-energy critical point for $\rho \equiv 1$ uniform on $\Omega = [0,1]^2$: $F_N(Y) = \Theta(1)$.

Convergence under a dimensionality condition

- Experimentally, when the point cloud $Y = (y_1, \ldots, y_N)$ is not chosen adversely, one observes that

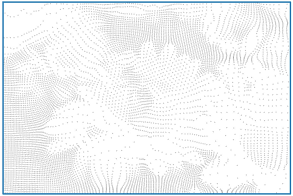
$$\mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{b_i(Y)}, \rho\right) \ll 1.$$

ρ



Y^N

$B_N = (b_1(Y^N), \ldots, b_N(Y^N))$



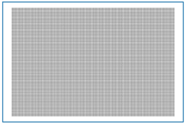
$B^N, N = 7280$

Convergence under a dimensionality condition

- Experimentally, when the point cloud $Y = (y_1, \ldots, y_N)$ is not chosen adversely, one observes that

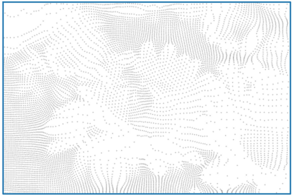
$$\mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{b_i(Y)}, \rho\right) \ll 1.$$

ρ



Y^N

$B_N = (b_1(Y^N), \ldots, b_N(Y^N))$



$B^N, N = 7280$

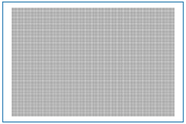
I.e., a *single step* Lloyd algorithm yields a good approximation of ρ.

Convergence under a dimensionality condition

- Experimentally, when the point cloud $Y = (y_1, \dots, y_N)$ is not chosen adversely, one observes that

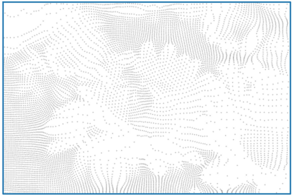
$$\mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{b_i(Y)}, \rho\right) \ll 1.$$

ρ



Y^N

$B_N = (b_1(Y^N), \dots, b_N(Y^N))$



$B^N, N = 7280$

I.e., a *single step* Lloyd algorithm yields a good approximation of ρ.

- Our main theorem explains this behaviour.

Convergence under a dimensionality condition

Theorem (Santambrogio, Sarrazin, M. (2021))

Let $\Omega \subseteq \mathbb{R}^d$ be convex and let $\rho \in \mathcal{P}(\Omega)$. Consider a point cloud Y in Ω^N s.t.

$$\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}, \text{ with } \beta \text{ and } C_0 > 0$$

Then,
$$\boxed{\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \mathrm{cst}(d, \Omega, C_0) \cdot N^{\frac{d-1}{\beta}-1}.}$$

Convergence under a dimensionality condition

Theorem (Santambrogio, Sarrazin, M. (2021))

Let $\Omega \subseteq \mathbb{R}^d$ be convex and let $\rho \in \mathcal{P}(\Omega)$. Consider a point cloud Y in Ω^N s.t.

$$\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}, \text{ with } \beta \text{ and } C_0 > 0$$

Then,

$$\boxed{\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \mathrm{cst}(d, \Omega, C_0) \cdot N^{\frac{d-1}{\beta}-1}.}$$

- The upper bound goes to zero as $N \to +\infty$ provided that $\beta > d - 1$.

Convergence under a dimensionality condition

Theorem (Santambrogio, Sarrazin, M. (2021))

Let $\Omega \subseteq \mathbb{R}^d$ *be convex and let* $\rho \in \mathcal{P}(\Omega)$. *Consider a point cloud* Y *in* Ω^N *s.t.*

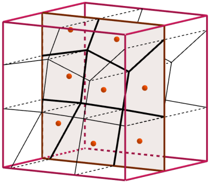
$$\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}, \text{ with } \beta \text{ and } C_0 > 0$$

Then,

$$\boxed{\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \mathrm{cst}(d, \Omega, C_0) \cdot N^{\frac{d-1}{\beta}-1}.}$$

- The upper bound goes to zero as $N \to +\infty$ provided that $\beta > d-1$. This is **tight**: If $(y_i)_{1 \leq i \leq N}$ lie on the $(d-1)$ hypercube $[0,1]^{d-1} \times \left\{\frac{1}{2}\right\}$ and if $\rho \equiv 1$ on $\Omega = [0,1]^d$:

$$\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \geq \frac{1}{12}$$



Convergence under a dimensionality condition

Theorem (Santambrogio, Sarrazin, M. (2021))

Let $\Omega \subseteq \mathbb{R}^d$ *be convex and let* $\rho \in \mathcal{P}(\Omega)$. *Consider a point cloud* Y *in* Ω^N *s.t.*

$$\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}, \text{ with } \beta \text{ and } C_0 > 0$$

Then, $$\boxed{\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \mathrm{cst}(d, \Omega, C_0) \cdot N^{\frac{d-1}{\beta}-1}.}$$

- The upper bound goes to zero as $N \to +\infty$ provided that $\beta > d-1$.
- When $\beta = d$, the upper bound of the theorem is

$$F_N(B_N) = \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right) \lesssim \left(\frac{1}{N}\right)^{1/d}.$$

Convergence under a dimensionality condition

Theorem (Santambrogio, Sarrazin, M. (2021))

Let $\Omega \subseteq \mathbb{R}^d$ *be convex and let* $\rho \in \mathcal{P}(\Omega)$*. Consider a point cloud* Y *in* Ω^N *s.t.*

$$\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}, \text{ with } \beta \text{ and } C_0 > 0$$

Then,

$$\boxed{\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \mathrm{cst}(d, \Omega, C_0) \cdot N^{\frac{d-1}{\beta}-1}.}$$

- The upper bound goes to zero as $N \to +\infty$ provided that $\beta > d - 1$.
- When $\beta = d$, the upper bound of the theorem is

$$F_N(B_N) = \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right) \lesssim \left(\frac{1}{N}\right)^{1/d}.$$

This does not match the upper bound on $\min_{\Omega^N} F_N$: $\min_{\Omega^N} F_N \lesssim \left(\frac{1}{N}\right)^{2/d}$.

Convergence under a dimensionality condition

Theorem (Santambrogio, Sarrazin, M. (2021))

Let $\Omega \subseteq \mathbb{R}^d$ be convex and let $\rho \in \mathcal{P}(\Omega)$. Consider a point cloud Y in Ω^N s.t.

$$\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}, \text{ with } \beta \text{ and } C_0 > 0$$

Then,
$$\boxed{\mathrm{W}_2^2\left(\tfrac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \mathrm{cst}(d, \Omega, C_0) \cdot N^{\frac{d-1}{\beta}-1}.}$$

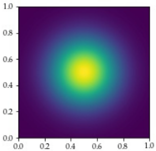
- The upper bound goes to zero as $N \to +\infty$ provided that $\beta > d - 1$.
- When $\beta = d$, the upper bound of the theorem is

$$F_N(B_N) = \mathrm{W}_2^2\left(\frac{1}{N}\sum_i \delta_{y_i}, \rho\right) \lesssim \left(\frac{1}{N}\right)^{1/d}.$$

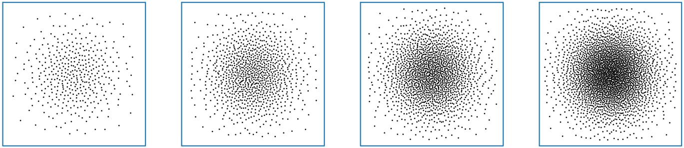
This does not match the upper bound on $\min_{\Omega^N} F_N$: $\min_{\Omega^N} F_N \lesssim \left(\frac{1}{N}\right)^{2/d}$.
However, the exponent in the upper bound **cannot be improved** !

Tightness in the case $\beta = d$

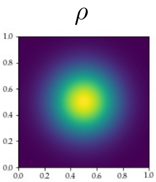
ρ



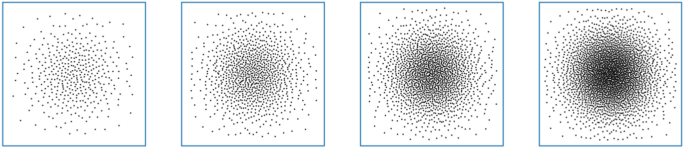
Random:



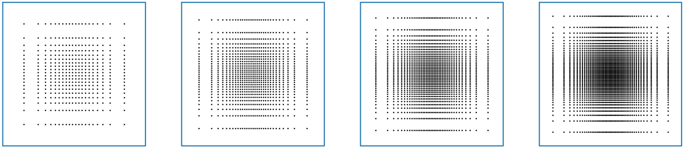
Tightness in the case $\beta = d$



Random:

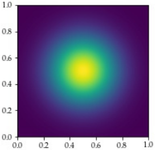


Grid-like:

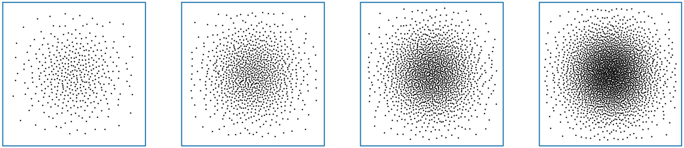


Tightness in the case $\beta = d$

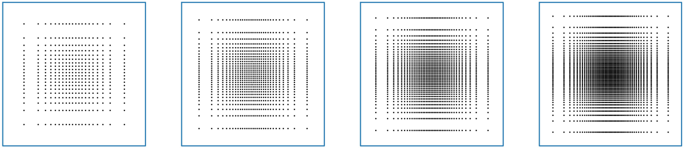
ρ



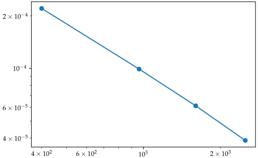
Random:



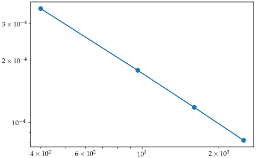
Grid-like:



$N = 400$ $N = 961$ $N = 1600$ $N = 2500$



Random $\sim N^{-0.99}$



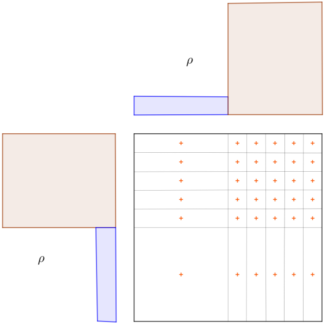
Grid-like $\sim N^{-0.78}$

Tightness in the case $\beta = d$

- On $\Omega = [-1,1]^d$, let $n \in \mathbb{N}^*$, $N = n^d$

$$\rho(x) = \frac{1}{n}\mathbb{1}_{[-1;0]} + \left(1 - \frac{1}{n}\right)\mathbb{1}_{[0;1]}, \qquad \rho_N = \underbrace{\rho \otimes \ldots \otimes \rho}_{d \text{ times}}$$

and Y^N a "grid-like" cloud of N points as on the picture.



Tightness in the case $\beta = d$

- On $\Omega = [-1,1]^d$, let $n \in \mathbb{N}^*$, $N = n^d$

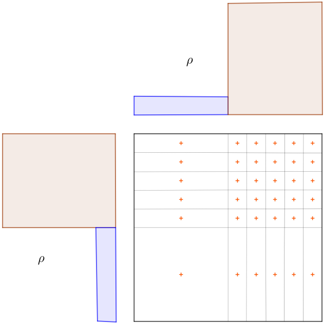
$$\rho(x) = \frac{1}{n}\mathbb{1}_{[-1;0]} + \left(1 - \frac{1}{n}\right)\mathbb{1}_{[0;1]}, \qquad \rho_N = \underbrace{\rho \otimes \ldots \otimes \rho}_{d \text{ times}}$$

and Y^N a "grid-like" cloud of N points as on the picture.

- Then, $\nabla_Y F_N(Y^N) = 0$, so that Y^N is a fixed point of Lloyd's algorithm, and

$$\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y^N)}, \rho_N\right) \geq CN^{-1/d}$$

with C independent of N.



Tightness in the case $\beta = d$

- On $\Omega = [-1,1]^d$, let $n \in \mathbb{N}^*$, $N = n^d$

$$\rho(x) = \frac{1}{n}\mathbb{1}_{[-1;0]} + \left(1 - \frac{1}{n}\right)\mathbb{1}_{[0;1]}, \qquad \rho_N = \underbrace{\rho \otimes \ldots \otimes \rho}_{d \text{ times}}$$

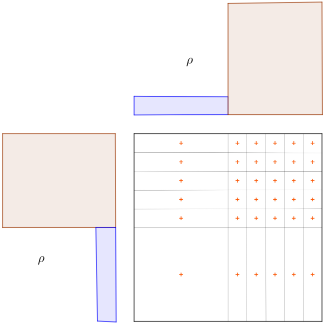
  and Y^N a "grid-like" cloud of N points as on the picture.

- Then, $\nabla_Y F_N(Y^N) = 0$, so that Y^N is a fixed point of Lloyd's algorithm, and

$$\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{b_i(Y^N)}, \rho_N\right) \geq CN^{-1/d}$$

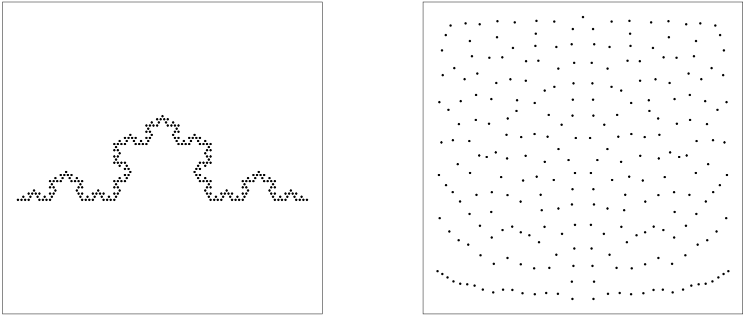
  with C independent of N.

- Thus, the exponent of the main theorem cannot be improved.



Numerical example with $d - 1 < \beta < d$

- Point are sampled from the Von Koch fractal (dimension $\beta = \frac{\ln 4}{\ln 3} \simeq 1.26$), $\rho \equiv 1$ on $\Omega = [0,1]^2$.

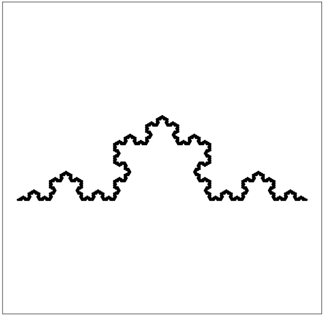


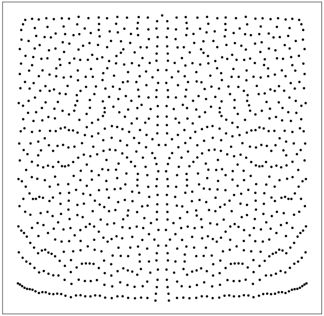
$N = 257$

- Numerically, it seems that $\mathrm{W}_2^2(\mu_N, \rho) \simeq N^{-1.01}$, while our upper bound would give an exponent of $\frac{d-1}{\beta} - 1 \simeq -0.207$.

Numerical example with $d - 1 < \beta < d$

- Point are sampled from the Von Koch fractal (dimension $\beta = \frac{\ln 4}{\ln 3} \simeq 1.26$), $\rho \equiv 1$ on $\Omega = [0, 1]^2$.



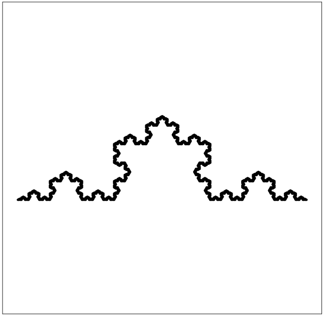


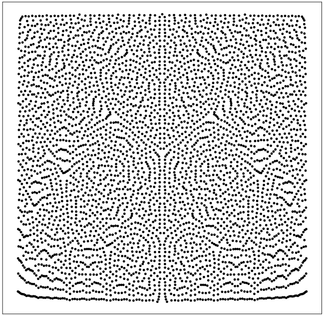
$N = 1025$

- Numerically, it seems that $\mathrm{W}_2^2(\mu_N, \rho) \simeq N^{-1.01}$, while our upper bound would give an exponent of $\frac{d-1}{\beta} - 1 \simeq -0.207$.

Numerical example with $d - 1 < \beta < d$

- Point are sampled from the Von Koch fractal (dimension $\beta = \frac{\ln 4}{\ln 3} \simeq 1.26$), $\rho \equiv 1$ on $\Omega = [0,1]^2$.



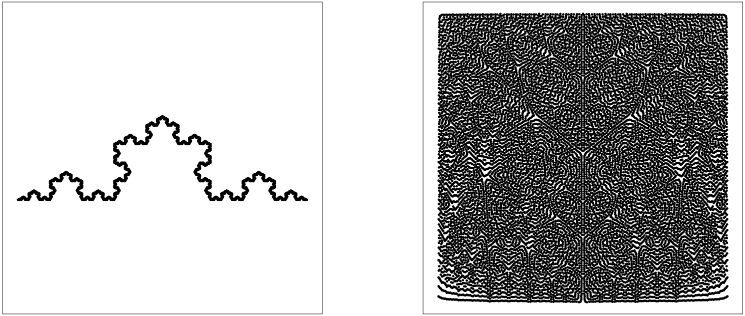


$N = 4097$

- Numerically, it seems that $\mathrm{W}_2^2(\mu_N, \rho) \simeq N^{-1.01}$, while our upper bound would give an exponent of $\frac{d-1}{\beta} - 1 \simeq -0.207$.

Numerical example with $d - 1 < \beta < d$

- Point are sampled from the Von Koch fractal (dimension $\beta = \frac{\ln 4}{\ln 3} \simeq 1.26$), $\rho \equiv 1$ on $\Omega = [0, 1]^2$.



$N = 16385$

- Numerically, it seems that $\mathrm{W}_2^2(\mu_N, \rho) \simeq N^{-1.01}$, while our upper bound would give an exponent of $\frac{d-1}{\beta} - 1 \simeq -0.207$.

Main theorem: sketch of proof

We assume: $\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}$, with β and $C_0 > 0$

Main idea: There cannot be "too many" Laguerre cells that are "elongated"

Main theorem: sketch of proof

We assume: $\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}$, with β and $C_0 > 0$

Main idea: There cannot be "too many" Laguerre cells that are "elongated"

- We use the concavity of the Laguerre cells w.r.t the weights Φ:

$$\tfrac{1}{2}\mathrm{Lag}_i(Y,0) \oplus \tfrac{1}{2}\mathrm{Lag}_i(Y,\Phi) \subset \mathrm{Lag}_i(Y,\Phi/2)$$

Main theorem: sketch of proof

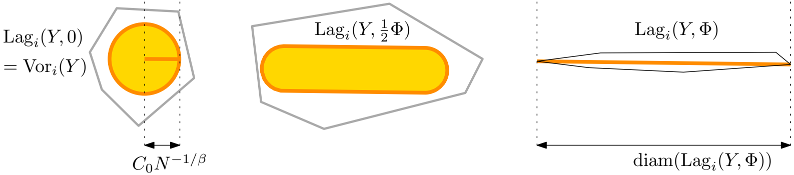
We assume: $\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}$, with β and $C_0 > 0$

Main idea: There cannot be "too many" Laguerre cells that are "elongated"

- We use the concavity of the Laguerre cells w.r.t the weights Φ:

$$\tfrac{1}{2}\mathrm{Lag}_i(Y,0) \oplus \tfrac{1}{2}\mathrm{Lag}_i(Y,\Phi) \subset \mathrm{Lag}_i(Y,\Phi/2)$$

$\Longrightarrow$ if $\mathrm{Lag}_i(Y,\Phi)$ is "elongated", then $|\mathrm{Lag}_i(Y,\tfrac{1}{2}\Phi)|$ is "large":



Main theorem: sketch of proof

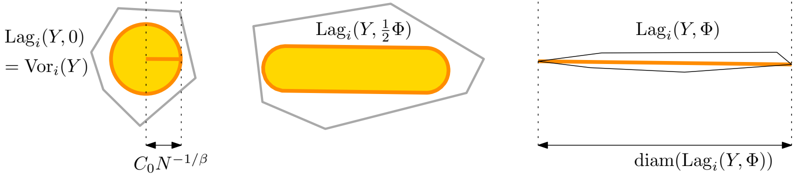
We assume: $\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}$, with β and $C_0 > 0$

Main idea: There cannot be "too many" Laguerre cells that are "elongated"

- We use the concavity of the Laguerre cells w.r.t the weights Φ:

$$\tfrac{1}{2}\mathrm{Lag}_i(Y,0) \oplus \tfrac{1}{2}\mathrm{Lag}_i(Y,\Phi) \subset \mathrm{Lag}_i(Y,\Phi/2)$$

$\Longrightarrow$ if $\mathrm{Lag}_i(Y,\Phi)$ is "elongated", then $|\mathrm{Lag}_i(Y,\tfrac{1}{2}\Phi)|$ is "large":



- The $(\mathrm{Lag}_i(Y,\tfrac{1}{2}\Phi))_i$ do not overlap: $\boxed{\sum_{i=1}^{N} \mathrm{diam}(\mathrm{Lag}_i(Y,\Phi)) \leq \frac{|\Omega|}{C_0 N^{-\frac{d-1}{\beta}}}}$

Main theorem: sketch of proof

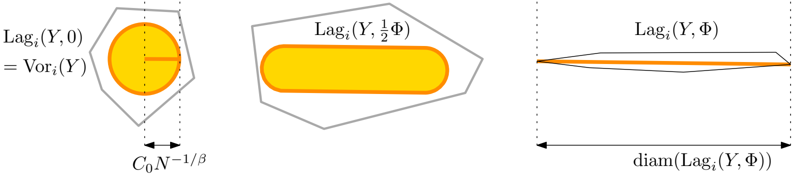
We assume: $\forall i \neq j, \quad \|y_i - y_j\| \geq C_0 N^{-\frac{1}{\beta}}$, with β and $C_0 > 0$

Main idea: There cannot be "too many" Laguerre cells that are "elongated"

- We use the concavity of the Laguerre cells w.r.t the weights Φ:

$$\tfrac{1}{2}\mathrm{Lag}_i(Y,0) \oplus \tfrac{1}{2}\mathrm{Lag}_i(Y,\Phi) \subset \mathrm{Lag}_i(Y,\Phi/2)$$

$\Longrightarrow$ if $\mathrm{Lag}_i(Y,\Phi)$ is "elongated", then $|\mathrm{Lag}_i(Y,\frac{1}{2}\Phi)|$ is "large":



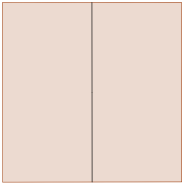
- The $(\mathrm{Lag}_i(Y,\frac{1}{2}\Phi))_i$ do not overlap: $\boxed{\sum_{i=1}^N \mathrm{diam}(\mathrm{Lag}_i(Y,\Phi)) \leq \frac{|\Omega|}{C_0 N^{-\frac{d-1}{\beta}}}}$
- $\mathrm{W}_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{b_i(Y)}, \rho\right) \leq \sum_{i=1}^N \int_{\mathrm{Lag}_i(Y,\Phi)} \|b_i(Y) - x\|^2 \,\mathrm{d}\rho(x) \lesssim N^{\frac{d-1}{\beta}-1}$

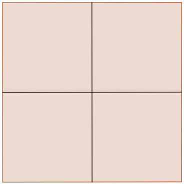
Limit of critical points as $N \to \infty$:

- If $Y^N \in \Omega^N$ is a critical point of $F_N(Y) = \mathrm{W}_2^2(\rho, \delta_Y)$, what are the possible narrow limit of the "discrete critical measure" $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{y_i^N}$?

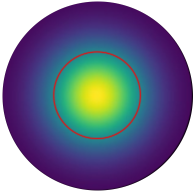
Limit of critical points as $N \to \infty$:

- If $Y^N \in \Omega^N$ is a critical point of $F_N(Y) = \mathrm{W}_2^2(\rho, \delta_Y)$, what are the possible narrow limit of the "discrete critical measure" $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{y_i^N}$?
- Discrete critical measures seem to converge to ρ or towards measures supported on union of regular sets as $N \to +\infty$.





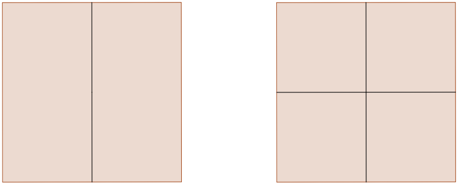
$\rho =$ Lebesgue on $[0;1]^2$



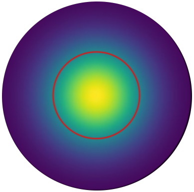
$\rho =$ Gaussian on $B(0;1)$

Limit of critical points as $N \to \infty$:

- If $Y^N \in \Omega^N$ is a critical point of $F_N(Y) = \mathrm{W}_2^2(\rho, \delta_Y)$, what are the possible narrow limit of the "discrete critical measure" $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{y_i^N}$?
- Discrete critical measures seem to converge to ρ or towards measures supported on union of regular sets as $N \to +\infty$.



$\rho =$ Lebesgue on $[0;1]^2$



$\rho =$ Gaussian on $B(0;1)$

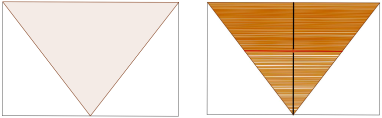
- Can this behavior be proven ? More generally, what can we say about limits of discrete critical measures?

Lagrangian critical measures

- **Disintegration of OT plan** Let $\rho, \mu \in \mathcal{P}(\Omega)$ and γ be the quadratic optimal transport plan, which we disintegrate into $\gamma = \int_\Omega \rho_y \mathrm{d}\mu(y)$.

Lagrangian critical measures

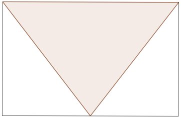
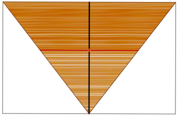
- **Disintegration of OT plan** Let $\rho, \mu \in \mathcal{P}(\Omega)$ and γ be the quadratic optimal transport plan, which we disintegrate into $\gamma = \int_\Omega \rho_y \mathrm{d}\mu(y)$.



$(\rho_y)_y \simeq$ Laguerre cells.

Lagrangian critical measures

- **Disintegration of OT plan** Let $\rho, \mu \in \mathcal{P}(\Omega)$ and γ be the quadratic optimal transport plan, which we disintegrate into $\gamma = \int_\Omega \rho_y \mathrm{d}\mu(y)$.

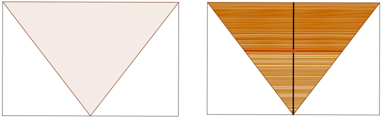
$(\rho_y)_y \simeq$ Laguerre cells.

- $\mu \in \mathcal{P}(\Omega)$ is *Lagrangian critical* for ρ if $\boxed{\text{for } \mu\text{-a.e.} y \in \Omega, y = \int_\Omega x \mathrm{d}\rho_y(x)}$

$$\iff \boxed{\forall \xi \in \mathcal{C}_c^0(\Omega, \mathbb{R}^d), \quad \frac{\mathrm{d}}{\mathrm{dt}} \mathrm{W}_2^2((\mathrm{id} + t\xi)_{\#}\mu, \rho)\Big|_{t=0} = 0.}$$

Lagrangian critical measures

- **Disintegration of OT plan** Let $\rho, \mu \in \mathcal{P}(\Omega)$ and γ be the quadratic optimal transport plan, which we disintegrate into $\gamma = \int_\Omega \rho_y \mathrm{d}\mu(y)$.



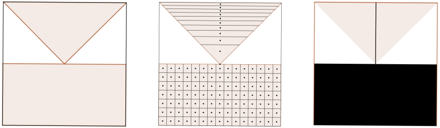
$(\rho_y)_y \simeq$ Laguerre cells.

- $\mu \in \mathcal{P}(\Omega)$ is *Lagrangian critical* for ρ if $\boxed{\text{for } \mu\text{-a.e.} y \in \Omega, y = \int_\Omega x \mathrm{d}\rho_y(x)}$

$$\iff \boxed{\forall \xi \in \mathcal{C}^0_c(\Omega, \mathbb{R}^d), \quad \frac{\mathrm{d}}{\mathrm{dt}} \mathrm{W}_2^2((\mathrm{id} + t\xi)_\# \mu, \rho)\Big|_{t=0} = 0.}$$

- **Narrow limit of critical points of F_N are Lagrangian critical:**

Assume that $\nabla_Y F_N(Y^N) = 0$ and $\lim_{N \to +\infty} \frac{1}{N} \sum_i \delta_{y_i^N} = \mu$. Then μ is Lagrangian critical.



ρ $\quad$ $\frac{1}{N} \sum_i \delta_{y_i^N}$ $\quad$ μ

Classification of Lagrangian critical measures

- Given $\mu \in \mathcal{P}(\Omega)$, let $\mathcal{E}_k$ the points of $\mathrm{spt}(\mu)$ whose "Laguerre cell" ρ_y has dimension $d - k$:

$$\boxed{\mathcal{E}_k = \{y \in \mathrm{spt}(\mu) \mid \dim(\mathrm{span}(\mathrm{spt}(\rho_y))) = d - k\}\,.}$$

Classification of Lagrangian critical measures

- Given $\mu \in \mathcal{P}(\Omega)$, let $\mathcal{E}_k$ the points of $\mathrm{spt}(\mu)$ whose "Laguerre cell" ρ_y has dimension $d - k$:

$$\boxed{\mathcal{E}_k = \{y \in \mathrm{spt}(\mu) \mid \dim(\mathrm{span}(\mathrm{spt}(\rho_y))) = d - k\}\,.}$$

- [Alberti 1994] $\Longrightarrow$ $\mathcal{E}_k$ is included in a countable union of $\mathcal{C}^2$ k-dimensional submanifolds, up to a $\mathcal{H}^k$-negligible set.

Classification of Lagrangian critical measures

- Given $\mu \in \mathcal{P}(\Omega)$, let $\mathcal{E}_k$ the points of $\mathrm{spt}(\mu)$ whose "Laguerre cell" ρ_y has dimension $d - k$:

$$\boxed{\mathcal{E}_k = \{y \in \mathrm{spt}(\mu) \mid \dim(\mathrm{span}(\mathrm{spt}(\rho_y))) = d - k\}\,.}$$

- [Alberti 1994] $\Longrightarrow$ $\mathcal{E}_k$ is included in a countable union of $\mathcal{C}^2$ k-dimensional submanifolds, up to a $\mathcal{H}^k$-negligible set.
- Is μ absolutely continuous with respect to $\mathcal{H}^k$ on $\mathcal{E}_k$?

Classification of Lagrangian critical measures

- Given $\mu \in \mathcal{P}(\Omega)$, let $\mathcal{E}_k$ the points of $\mathrm{spt}(\mu)$ whose "Laguerre cell" ρ_y has dimension $d - k$:

$$\boxed{\mathcal{E}_k = \{y \in \mathrm{spt}(\mu) \mid \dim(\mathrm{span}(\mathrm{spt}(\rho_y))) = d - k\}\,.}$$

- [Alberti 1994] $\implies \mathcal{E}_k$ is included in a countable union of $\mathcal{C}^2$ k-dimensional submanifolds, up to a $\mathcal{H}^k$-negligible set.
- Is μ absolutely continuous with respect to $\mathcal{H}^k$ on $\mathcal{E}_k$?

Proposition (Santambrogio, Sarrazin, M. 2021)

If μ is a Lagrangian critical point and $k \in \{0, 1, d\}$, then μ is absolutely continuous with respect to $\mathcal{H}^k$ on $\mathcal{E}_k$.

Classification of Lagrangian critical measures

- Given $\mu \in \mathcal{P}(\Omega)$, let $\mathcal{E}_k$ the points of $\mathrm{spt}(\mu)$ whose "Laguerre cell" ρ_y has dimension $d - k$:

$$\boxed{\mathcal{E}_k = \{y \in \mathrm{spt}(\mu) \mid \dim(\mathrm{span}(\mathrm{spt}(\rho_y))) = d - k\}\,.}$$

- [Alberti 1994] $\Longrightarrow$ $\mathcal{E}_k$ is included in a countable union of $\mathcal{C}^2$ k-dimensional submanifolds, up to a $\mathcal{H}^k$-negligible set.
- Is μ absolutely continuous with respect to $\mathcal{H}^k$ on $\mathcal{E}_k$?

Proposition (Santambrogio, Sarrazin, M. 2021)

If μ is a Lagrangian critical point and $k \in \{0, 1, d\}$, then μ is absolutely continuous with respect to $\mathcal{H}^k$ on $\mathcal{E}_k$.

- Very preliminary result: most of the questions about Lagrangian critical measures are open.

Summary and Perspectives

Take-home message: Despite the non-convexity, gradient descent strategies for optimal uniform quantization problem, i.e.

$$\min_{Y \in \Omega^N} \mathrm{W}_2^2 \left(\frac{1}{N} \sum_i \delta_{y_i}, \rho \right)$$

lead to low energy configurations when the points in the initial point are far enough from each other, i.e. $\gtrsim \left(\frac{1}{N}\right)^{\frac{1}{\beta}}$ with $\beta > d - 1$ and $d = \dim(\Omega)$.

Summary and Perspectives

Take-home message: Despite the non-convexity, gradient descent strategies for optimal uniform quantization problem, i.e.

$$\min_{Y \in \Omega^N} \mathrm{W}_2^2 \left(\frac{1}{N} \sum_i \delta_{y_i}, \rho \right)$$

lead to low energy configurations when the points in the initial point are far enough from each other, i.e. $\gtrsim \left(\frac{1}{N}\right)^{\frac{1}{\beta}}$ with $\beta > d - 1$ and $d = \dim(\Omega)$.

(Some) open questions:

- Can the analysis be extended to Wasserstein linear regression ?
- Can the exponent be improved if ρ is bounded from above and below?
- Can the exponent be improved when ρ is supported on a submanifold of $\mathbb{R}^d$?

Summary and Perspectives

Take-home message: Despite the non-convexity, gradient descent strategies for optimal uniform quantization problem, i.e.

$$\min_{Y \in \Omega^N} \mathrm{W}_2^2 \left(\frac{1}{N} \sum_i \delta_{y_i}, \rho \right)$$

lead to low energy configurations when the points in the initial point are far enough from each other, i.e. $\gtrsim \left(\frac{1}{N}\right)^{\frac{1}{\beta}}$ with $\beta > d - 1$ and $d = \dim(\Omega)$.

(Some) open questions:

- Can the analysis be extended to Wasserstein linear regression ?
- Can the exponent be improved if ρ is bounded from above and below?
- Can the exponent be improved when ρ is supported on a submanifold of $\mathbb{R}^d$?
- What can we say about the limits of discrete critical points?
- What about *stable* critical points ?

Summary and Perspectives

Take-home message: Despite the non-convexity, gradient descent strategies for optimal uniform quantization problem, i.e.

$$\min_{Y \in \Omega^N} \mathrm{W}_2^2 \left(\frac{1}{N} \sum_i \delta_{y_i}, \rho \right)$$

lead to low energy configurations when the points in the initial point are far enough from each other, i.e. $\gtrsim \left(\frac{1}{N}\right)^{\frac{1}{\beta}}$ with $\beta > d - 1$ and $d = \dim(\Omega)$.

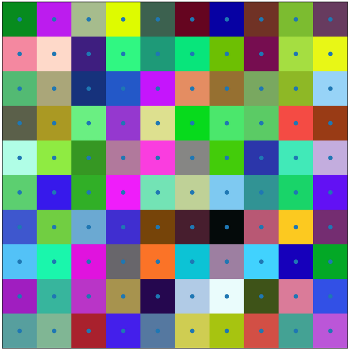
(Some) open questions:

- Can the analysis be extended to Wasserstein linear regression ?
- Can the exponent be improved if ρ is bounded from above and below?
- Can the exponent be improved when ρ is supported on a submanifold of $\mathbb{R}^d$?
- What can we say about the limits of discrete critical points?
- What about *stable* critical points ?

Thank you for your attention!

An unstable critical point

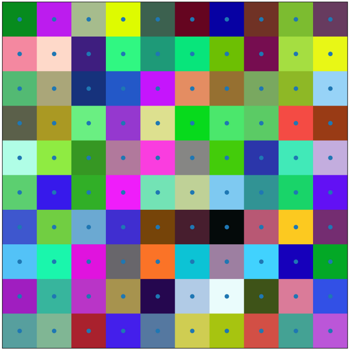
- $\Omega = [-\pi, \pi]^2, \rho \equiv 1/(4\pi^2), N = 10^2, Y^0 =$ uniform grid.
- Iterates follow Lloyd's algorithm: $Y^{k+1} = (b_1(Y^k), \ldots, b_N(Y^k))$.



k=1

An unstable critical point

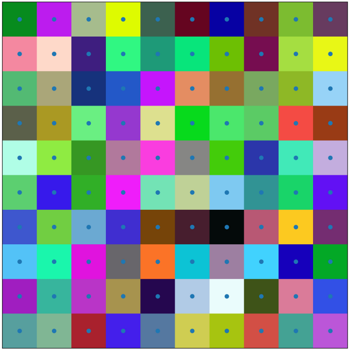
- $\Omega = [-\pi, \pi]^2, \rho \equiv 1/(4\pi^2)$, $N = 10^2$, Y^0 = uniform grid.
- Iterates follow Lloyd's algorithm: $Y^{k+1} = (b_1(Y^k), \ldots, b_N(Y^k))$.



k=101

An unstable critical point

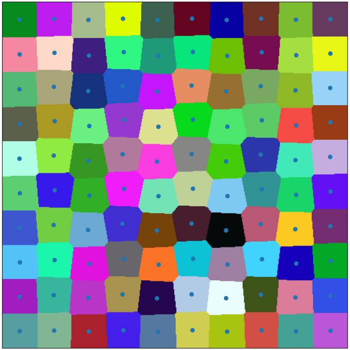
- $\Omega = [-\pi, \pi]^2, \rho \equiv 1/(4\pi^2), N = 10^2, Y^0 =$ uniform grid.
- Iterates follow Lloyd's algorithm: $Y^{k+1} = (b_1(Y^k), \ldots, b_N(Y^k))$.



k=121

An unstable critical point

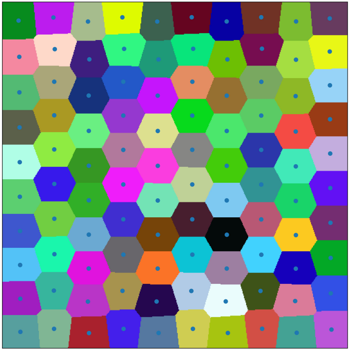
- $\Omega = [-\pi, \pi]^2, \rho \equiv 1/(4\pi^2), N = 10^2, Y^0 =$ uniform grid.
- Iterates follow Lloyd's algorithm: $Y^{k+1} = (b_1(Y^k), \ldots, b_N(Y^k))$.



k=141

An unstable critical point

- $\Omega = [-\pi, \pi]^2, \rho \equiv 1/(4\pi^2), N = 10^2, Y^0 =$ uniform grid.
- Iterates follow Lloyd's algorithm: $Y^{k+1} = (b_1(Y^k), \ldots, b_N(Y^k))$.



k=161

- Lloyd's iterate escape the critical point due to numerical error + instability.