Convergence of bipartite functionals

F. Barthe & Ch. Bordenave

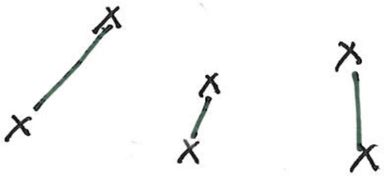
(work in progress)

Classical optimisation problems

Given $x_1, \dots, x_n \in (\mathbb{R}^d, |\cdot|)$

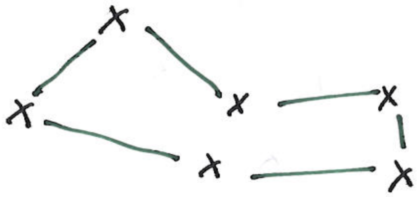
- Minimal matching $(n=2m)$.

  $H(x_1, \dots, x_n) = \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{m} |x_{\sigma(2i)} - x_{\sigma(2i-1)}|$



- Travelling salesperson problem

  $T(x_1, \dots x_n) = \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n-1} |x_{\sigma(i+1)} - x_{\sigma(i)}| + |x_{\sigma(1)} - x_{\sigma(n)}|$



- Minimal spanning tree

  $ST(x_1, \dots x_n)$.

- . .

Probabilistic theory of these problems

- Theorem of Beardwood, Halton, Hammersley (1959):

  $X_i, i \geq 1$ iid random var. in $[0,1]^d$, $d \geq 2$

  $X_i \sim f(x)dx$ + singular part.

  then a.s.: $\displaystyle \lim_{n\to\infty} \frac{T(X_1,\dots,X_n)}{n^{\frac{d-1}{d}}} = \alpha(d) \int f^{\frac{d-1}{d}}$

  Remark: $d(x_1, \{X_2, \dots X_n\}) \approx (f(x_1)n)^{-1/d}$

- Similar results for other functionals:

  Papadimitriou (Π), Steele (ST), Avis Davis Steele, Talagrand, Goemans Bertsimas, Rhee, Yukich....

- Unified approach:

  eg books of Steele (1997), Yukich (1998).

An "umbrella theorem" $p<d$ [Yukich]

Let $L : \{\text{finite multisets of } \mathbb{R}^d\} \longrightarrow \mathbb{R}^+$ be

- p-homogeneous: $L(\lambda \mathcal{X}+a) = \lambda^p L(\mathcal{X})$.
- p-subadditive: for all $\mathcal{X}, \mathcal{Y} \subset Q$
$$L(\mathcal{X} \cup \mathcal{Y}) \leq L(\mathcal{X}) + L(\mathcal{Y}) + c(\operatorname{diam} Q)^p.$$
(implies: $L(\mathcal{X}) \leq c(\operatorname{diam} Q)^p (\operatorname{card} \mathcal{X})^{\frac{d-p}{d}}$ [Rhee])
- p-smooth : $\forall \mathcal{X}, \mathcal{Y} \subset Q$
$$|L(\mathcal{X} \cup \mathcal{Y}) - L(\mathcal{X})| \leq c(\operatorname{diam} Q)^p (\operatorname{card} \mathcal{Y})^{\frac{d-p}{d}}.$$

Then, if a boundary functional L_∂ is close to L, $X_i \sim f(x)dx + \mu_s$

$$\lim_{n\to\infty} \frac{L(X_1, \dots, X_n)}{n^{\frac{d-p}{d}}} = \alpha_L \int f^{\frac{d-p}{d}} \quad \text{a.s.}$$

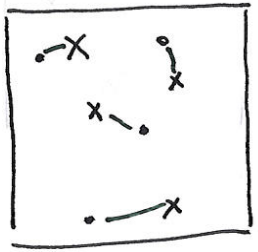
Remark: L_∂ not used for:
- limsup $\leq$.
- $X_i \sim \mathbb{1}_K(x)\, dx / \operatorname{vol}(K)$.

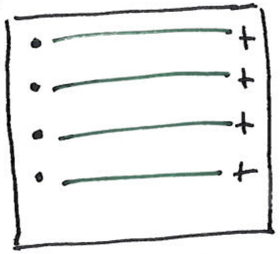
Bipartite matching

Given $X_1, \dots, X_n$, $Y_1, \dots Y_n$ iid $\sim \mu$ on $[0,1]^d$.

$$L(\{X_1, \dots X_n\}, \{Y_1, \dots Y_n\}) := \inf_{\sigma \in \mathcal{S}_n} \sum_i |X_i - Y_{\sigma(i)}| = n W_1\left(\frac{\sum \delta_{X_i}}{n}, \frac{\sum \delta_{Y_i}}{n}\right).$$

- 1-homogeneous.
- $L(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n) \leq cn$ (not $cn^{\frac{d-1}{d}}$).





However, Dobrić - Yukich show (1995).

If $d \geq 3$, $X_i \sim f(x)dx + \mu_s$, then a.s.

$$\lim_{n \to \infty} \frac{L(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n)}{n^{\frac{d-1}{d}}} = \beta(d) \int f^{\frac{d-1}{d}}.$$

Our goal.

- [DY] uses Kantorovich duality.

$$L(\{X_i\},\{Y_i\}) = \sup_{f \text{ 1-Lipschitz}} \sum_{i=1}^{n} f(X_i) - f(Y_i).$$

- We propose a direct approach, following. Boutet de Monvel, Martin (2002).

- it extends to an umbrella theorem for bipartite functionals

ex: bipartite TSP: shortest bipartite cycle on $\{X_i\}$ $\{Y_j\}$.

bipartite spanning tree . . .

Bipartite umbrella theorem, $d > 2p$

Notation: $\mathcal{X}(Q) := \operatorname{card}(\mathcal{X} \cap Q)$.

Let $F : \{\text{Finite multisets of } \mathbb{R}^d\}^2 \longrightarrow \mathbb{R}^+$ be

- p-homogeneous: $L(\lambda\mathcal{X}+a, \lambda\mathcal{Y}+a) = \lambda^p L(\mathcal{X},\mathcal{Y})$.
- p subadditive: $\forall k \geq 1$, $\forall \mathcal{X}_i, \mathcal{Y}_i \subset Q$

$$L\left(\bigcup_{i=1}^{k} \mathcal{X}_i, \bigcup_{i=1}^{k} \mathcal{Y}_i\right) \leq \sum_{i=1}^{k} \left[L(\mathcal{X}_i, \mathcal{Y}_i) + C(\operatorname{diam} Q)^p \left(1 + |\mathcal{X}_i(Q) - \mathcal{Y}_i(Q)|\right)\right]$$

- p-smooth: $\forall \mathcal{X}_i, \mathcal{Y}_i \subset Q$

$$L(\mathcal{X}_0 \cup \mathcal{X}_1, \mathcal{Y}_0 \cup \mathcal{Y}_1) \leq L(\mathcal{X}_0 \cup \mathcal{X}_2, \mathcal{Y}_0 \cup \mathcal{Y}_2) + C(\operatorname{diam} Q)^p \left(\mathcal{X}_1(Q) + \mathcal{Y}_1(Q) + \mathcal{X}_2(Q) + \mathcal{Y}_2(Q)\right).$$

If X_i, Y_i are iid $\sim f(x)\,dx + \mu_s$ (bounded support).

Then $\limsup \dfrac{L(\{X_i\}_{i\leq n}, \{Y_i\}_{i \leq n})}{n^{\frac{d-p}{d}}} \leq \alpha_L \int f^{\frac{d-p}{d}}$ a.s.

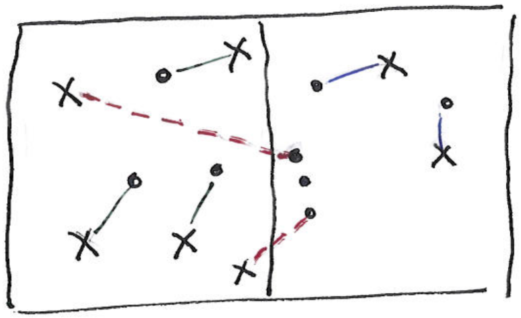
and. $\lim$ " " $= \alpha_L \int f^{\frac{d-p}{d}}$ if $X_i \sim \dfrac{\mathbb{1}_K}{\operatorname{Vol} K}\,dx$

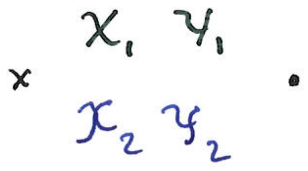
or L has a good boundary functional. ($\mu_s = 0$ or "convexity")

Illustrations for subadditivity

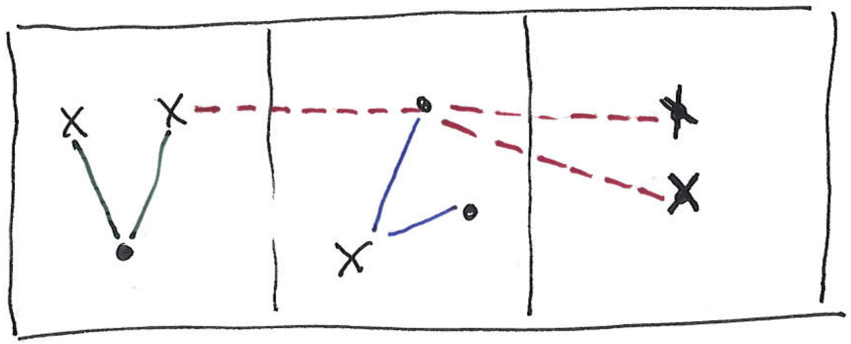
- Bipartite matching:

  extension: if $m>n$, $L(\{x_i\}_{i=1}^{m}, \{y_j\}_{j=1}^{m}) = \inf\limits_{\text{card} I = n} L(\{x_i\}_{i\in I}, \{y_j\}_{j=1}^{n})$.





- Bipartite spanning tree.



Consequences of smoothness

- convergence of $\frac{EL}{n^{\frac{d-p}{d}}}$ is enough (concentration, Rhee, Talagrand).
- poissonized version: $\left.\begin{array}{l}(X_i)_{i\geq 1} \quad (Y_j)_{j\geq 1} \sim \mu \\ N_1, N_2 \sim \mathcal{P}(n)\end{array}\right\}$ independent

$\mathcal{X} = \{X_1, \dots, X_{N_1}\}$ $\qquad$ $\mathcal{Y} = \{Y_1, \dots, Y_{N_2}\}$

are independent Poisson point processes $PPP(n\mu)$.

$$\left| L(\{X_1 \dots X_n\}, \{Y_1 \dots Y_n\}) - L\underbrace{(\{X_1, \dots X_{N_1}\}, \{Y_1, \dots Y_{N_2}\})}_{L(n\mu) \text{ for short}} \right| \leq C(\operatorname{diam} Q)^p \left(|N_1 - n| + |N_2 - n|\right)$$

in expectation $\leq 2c \, (\operatorname{diam} Q)^p \; E\,|N_1 - n|$

$\leq$ $\qquad$ " $\qquad$ $\sqrt{Var(N_1)} = c' n^{1/2} \overset{d>2p}{\ll} n^{\frac{d-p}{d}}$

- Coupling

$$|EL(n\mu) - EL(n\mu')| \leq c\,(\operatorname{diam} Q)^p \times n \, d_{TV}(\mu, \mu').$$

Dyadic partitions: downwards

- $\mathcal{X}, \mathcal{Y}$ $\quad PPP(\nu)$, $\operatorname{supp} \nu \subset Q$.

Given a partition $Q = \bigsqcup_{q \in \mathcal{P}} q$.

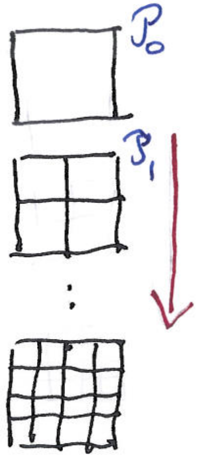
$$L(\mathcal{X} \cap Q, \mathcal{Y} \cap Q) \leq \sum_{q \in \mathcal{P}} L(\underset{\uparrow}{\mathcal{X} \cap q}, \underset{\uparrow}{\mathcal{Y} \cap q}) + c(\operatorname{diam} Q)^p \sum_{q \in \mathcal{P}} \left(1 + |\underset{\uparrow}{\mathcal{X}(q)} - \underset{\uparrow}{\mathcal{Y}(q)}|\right)$$

$PPP(1\!\!1_q \cdot \nu)$ $\qquad \mathcal{P}(\nu(q))$.

For $\nu = n\mu$

$$E\,L(n\mu) \leq \sum_{p \in \mathcal{P}} E\,L(n\mu 1\!\!1_q) + c\,(\operatorname{diam} Q)^p \sum_{q \in \mathcal{P}} 1 + \sqrt{n\mu(q)}$$

- $\mathcal{P}_k$ partition of $[0,1]^d$ in 2^{kd} cubes of size 2^{-k}.

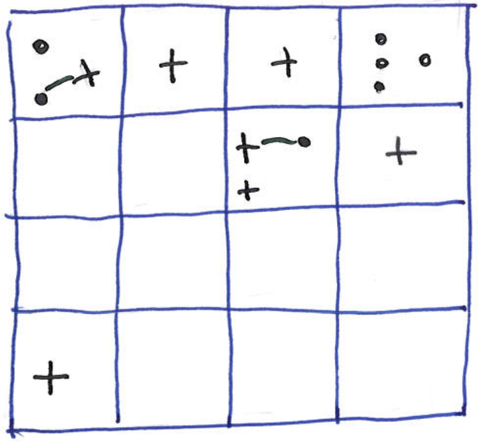


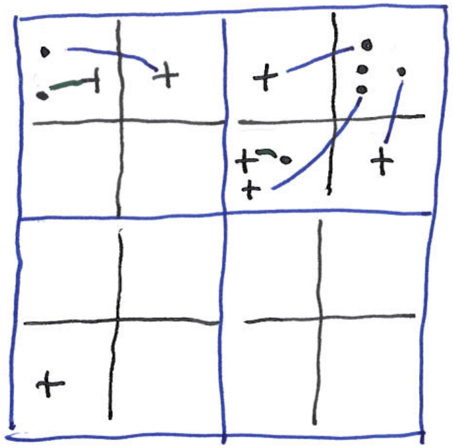
$$EL(n\mu) \leq \sum_{p \in \mathcal{P}_1} EL(n\mu 1\!\!1_p) + c \sum_{p \in \mathcal{P}} 1 + \sqrt{n\mu(p)}$$

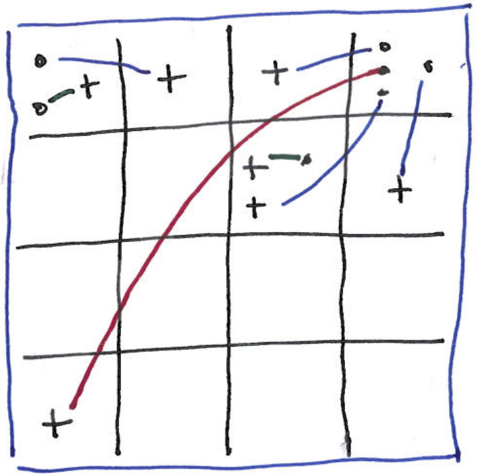
iterate $\vdots$

$$\leq \sum_{p \in \mathcal{P}_k} EL(n 1\!\!1_p \cdot \mu) + c' \sum_{\ell=1}^{k} 2^{-\ell p} \sum_{p \in \mathcal{P}_\ell} \left(1 + \sqrt{n\mu(p)}\right).$$

Dyadic partitions: upwards





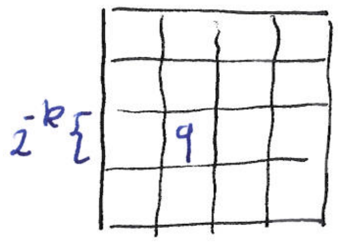


CV for uniform samples on $[0,1]^d$ adapting [BdM-M]

$(m=2^k)$ dyadic partition $\mathcal{P}_1, \mathcal{P}_2 \ldots \mathcal{P}_k$.

$$E\,L(n \mathbb{1}_{[0,1]^d}) \le \sum_{\underset{\nearrow\, 2^{kd} \text{ terms}}{q \in \mathcal{P}_k}} E L(n \mathbb{1}_{\overset{\uparrow}{q}}) + \text{remainder}.$$

$q = a_q + 2^{-k}[0,1]^d$.



p-homogeneity $\Rightarrow E\,L(n \mathbb{1}_q) = (2^{-k})^p E\,L(n 2^{-kd} \mathbb{1}_{[0,1]^d})$.

Setting $f(t) = \dfrac{E\,L(t^d \mathbb{1}_{[0,1]^d})}{t^{d-p}}$.

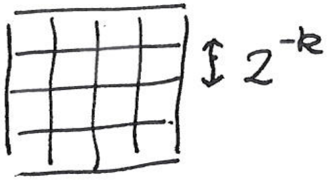
one gets $f(tm) \le f(t) + O(\underbrace{t^{p-d/2}}_{\to 0 \text{ if } d>2p})$

so $\exists \lim\limits_{t\to\infty} f(t) =: \alpha_L$

Upper limits for densities $f\,dx$ on $[0,1]^d$.

$$EL(nf) \leq \sum_{q\in\mathcal{P}_k} EL(nf\mathbb{1}_q) + c\sum_{\ell=1}^{k} 2^{-\ell p} \sum_{q\in\mathcal{P}_\ell} 1+\sqrt{n\int_q f}$$

let $f_k = \sum_{q\in\mathcal{P}_k} \frac{\int_q f}{\mathrm{vol}(q)} \mathbb{1}_q$ piecewise constant



Smoothness $\Rightarrow EL(nf\mathbb{1}_q) \leq EL(nf_k\mathbb{1}_q) + c(2^{-k})^p n \int_q |f-f_k|$

$$\rightsquigarrow EL(nf) \leq \int g(n2^{-kd}f_k) f_k^{\frac{d-p}{d}} + c\,(n^{1/d}2^{-k})^p \int|f-f_k| + (2^k n^{-1/d})^{\frac{d}{2}-p}$$

where $g(t) = \frac{EL(t\mathbb{1}_{[0,1]^d})}{t^{\frac{d-p}{d}}} \xrightarrow[t\to\infty]{} \alpha_L$.

Tuning parameters: $k = k(n) \to \infty$.

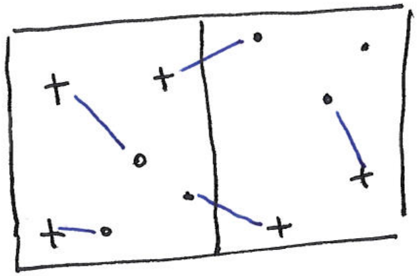
$n^{1/d}\, 2^{-k(n)} \longrightarrow \infty$.

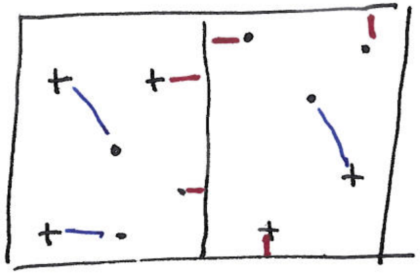
but $n^{1/d}\, 2^{-k(n)} \int|f-f_{k(n)}| \longrightarrow 0$

Lower limit, boundary functional : Matching $p=1$, $d \geq 3$

$Q = \bigcup_{q \in \mathcal{P}} q$ partition in cubes.

need: $E\,L(n 1\!\!1_Q) \geq \sum E\,L(n 1\!\!1_q)$ − error term.





If $X_1, \dots X_n, Y_1, \dots Y_m \in S$

$$L_{\partial S, \varepsilon}\left(\{X_i\}_{i \leq n}, \{Y_j\}_{j \leq m}\right) := \min_{A,B} \sum_{i \in A} |X_i - Y_{\sigma(i)}| + \sum_{i \notin A} \left(\varepsilon + d(X_i, \partial S)\right) + \sum_{j \notin B} \left(\varepsilon + d(Y_j, \partial S)\right)$$

$\sigma : A \to B$ bijection.

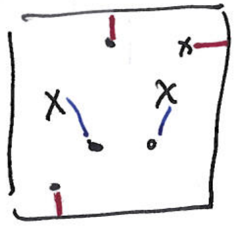
Taking expectations

$$(\operatorname{diam} Q) \sqrt{\nu(Q)} + E\,L(1\!\!1_Q \nu) \geq \sum_{q \in \mathcal{P}} E\,L_{\partial q}(1\!\!1_q \cdot \nu).$$

Main step: $Q=[0,1]^d$.

show: $$\liminf_n \frac{E L_{\partial Q}(n 1\!\!1_Q)}{n^{\frac{d-1}{d}}} \geq \lim \frac{E L(n 1\!\!1_Q)}{n^{\frac{d-1}{d}}} = \alpha_L$$

- Start from optimal boundary matching on Q.
- build a usual matching (reallocate boundary-matched points).



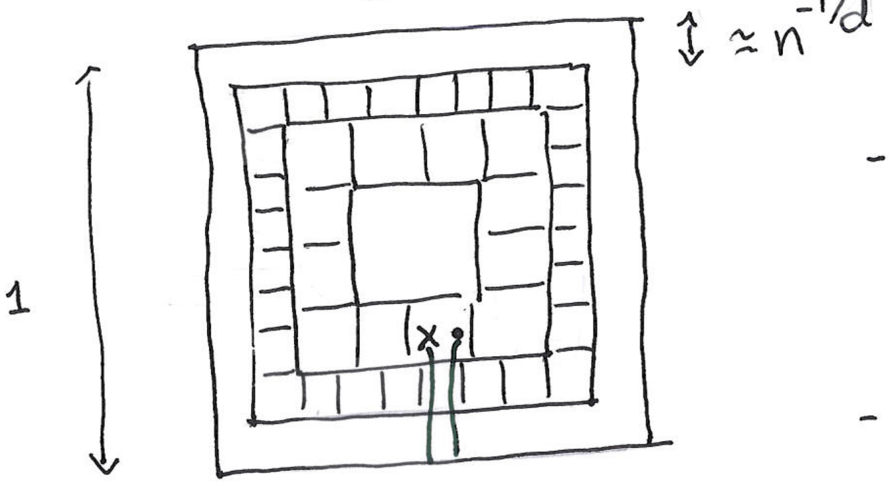
- Lemma $X, Y \sim PPP(n 1\!\!1_{[0,1]^d})$.

  $X_{\partial Q}$:= number of points matched to boundary ∂Q.

  then $E X_{\partial Q} \leq c n^{\frac{d-1}{d}}$

  too large ...

Counting boundary-matched points.



- each cube verifies

$$\text{diam} q < \text{dist}\left(q, \partial\left([0,1]^d\right)\right).$$

- $\Rightarrow$ in each q at most $|\mathcal{X}(q) - \mathcal{Y}(q)|$ points matched to the boundary.

Penalized problem: $\varepsilon + d(x, \partial Q)$.

$$E\, \mathcal{X}_{\partial Q, \varepsilon}(n 1\!\!1_Q) \leq c\, \varepsilon^{1-\frac{d}{2}} \sqrt{n}$$

$$\ll n^{\frac{d-1}{a}} \text{ if } \varepsilon \gg n^{-1/d}$$

Final step : $Q = [0,1]^d \quad d \geqslant 3$

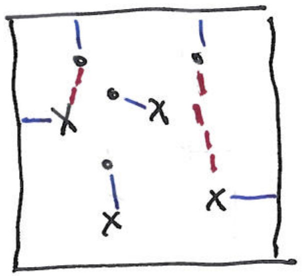
$$L_{\partial Q}(\mathcal{X},\mathcal{Y}) = \sum_{i \in A} |X_i - Y_{\sigma(i)}| + \sum_{i \notin A} d(X_i, \partial Q) + \sum_{j \notin B} d(Y_j, \partial Q).$$

$$= \sum_{i \in A} |X_i - Y_{\sigma(i)}| + \sum_{i \notin A} (\varepsilon + d(X_i, \partial Q)) + \sum_{j \notin B} (\varepsilon + d(Y_j, \partial Q)) - \varepsilon \mathcal{X}_{\partial Q} - \varepsilon \mathcal{Y}_{\partial Q}.$$

$$\geqslant L_{\partial Q, \varepsilon}(\mathcal{X},\mathcal{Y}) - \varepsilon (\mathcal{X}_{\partial Q} + \mathcal{Y}_{\partial Q}).$$

$$\Rightarrow \quad E\, L_{\partial Q}(n 1\!\!1_Q) \geqslant E\, L_{\partial Q, \varepsilon}(n 1\!\!1_Q) - 2c\varepsilon n^{\frac{d-1}{d}}$$

$$\Rightarrow \quad \liminf_n \frac{E\, L_{\partial Q}(n 1\!\!1_Q)}{n^{\frac{d-1}{d}}} \geqslant \liminf \frac{E\, L_{\partial Q, \varepsilon}(n 1\!\!1_Q)}{n^{\frac{d-1}{d}}} - 2c\varepsilon.$$



$\varepsilon^{1-\frac{d}{2}} \sqrt{n} \ll n^{\frac{d-1}{d}}.$

points to reallocate.

Final remarks

- Method adapts to
  - unbounded distributions
  - $W_p\left(\frac{\sum \delta_{X_i}}{n}, \mu\right)$ X_i iid $\sim \mu$.
  - bipartite graphs with random degrees (X_i, a_i) (Y_j, b_j).
  - stationary versions:

  Ex: $d \geq 3$ [Holroyd.] $\mathcal{X}, \mathcal{Y}$. PPP(Lebesgue $_{\mathbb{R}^d}$).

  $\exists$ stationnary, locally minimal, bipartite matching.

- $d > 2p$ $\rightsquigarrow$ points are linked to close neighbors.

  $d \leq 2p$ hard:

  Ex: Bipartite 1 matching in $[0,1]^2$.

  [Ajtai-Komlos-Tusnady] [Talagrand].

  distance to matched point $\approx \sqrt{\frac{\log n}{n}}$

  " " closest point $\approx \frac{1}{\sqrt{n}}$.