Optimal martingale transport in general dimensions

Young-Heon Kim

University of British Columbia

Based on joint work with

Nassif Ghoussoub (UBC) and Tongseok Lim (Oxford)

May 1, 2017

Optimal Transport meets Probability, Statistics and Machine Learning
April 30 – May 5, 2017
Oaxaca

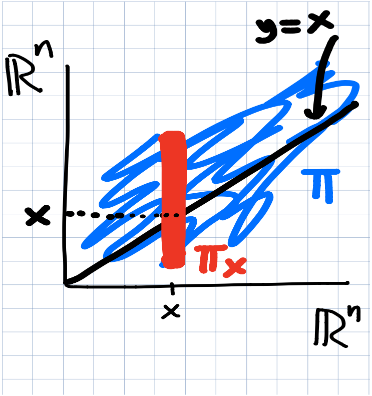
The point of this talk:

> Optimal martingale transport has rich but hidden structures, especially in multi-dimensions.

Optimal Martingale Transport Problem

- cost function $c : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$,
- Probability measures μ, ν on $\mathbf{R}^n$.
- $MT(\mu, \nu)$: probability measures π on $\mathbf{R}^n \times \mathbf{R}^n$ with the marginals μ, ν, and its disintegration $(\pi_x)_{x \in \mathbf{R}^n}$ has barycenter at x (martingale constraint):

$$\int y d\pi_x(y) = x.$$



Study the optimal solutions of

$$\max / \min_{\pi \in MT(\mu,\nu)} \int_{\mathbf{R}^n \times \mathbf{R}^n} c(x, y) d\pi(x, y).$$

Remark: [Strassen '65]

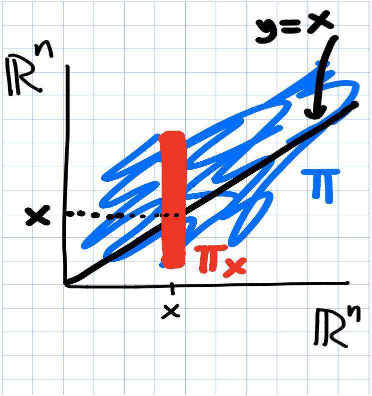
- $MT(\mu, \nu) \neq \emptyset$

$\Leftrightarrow$ μ and ν are in convex order;
$\mu \leq_c \nu$, i.e. $\int \xi d\mu \leq \int \xi d\nu$, $\forall$ convex $\xi : \mathbf{R}^n \to \mathbf{R}$.

Optimal Martingale Transport Problem

- cost function $c : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$,
- Probability measures μ, ν on $\mathbf{R}^n$.
- $MT(\mu, \nu)$: probability measures π on $\mathbf{R}^n \times \mathbf{R}^n$ with the marginals μ, ν, and its disintegration $(\pi_x)_{x \in \mathbf{R}^n}$ has barycenter at x (martingale constraint):

$$\int y d\pi_x(y) = x.$$



Study the optimal solutions of

$$\max / \min_{\pi \in MT(\mu,\nu)} \int_{\mathbf{R}^n \times \mathbf{R}^n} c(x, y) d\pi(x, y).$$

Remark: [Strassen '65]

- $MT(\mu, \nu) \neq \emptyset$

$\Leftrightarrow$ μ and ν are in convex order;
$\mu \leq_c \nu$, i.e. $\int \xi d\mu \leq \int \xi d\nu$, $\forall$ convex $\xi : \mathbf{R}^n \to \mathbf{R}$.

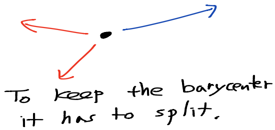
Some references:

- Discrete-time : **Beiglböck, Davis, De March, Ghoussoub, Griessler, Henry-Labordère, Hobson, Kim, Klimmek, Lim, Neuberger, Nutz, Penkner, Juillet, Schachermayer, Touzi.....**
- Continuous-time : **Beiglböck, Bayraktar, Claisse, Cox, Davis, Dolinsky, Galichon, Guo, Hu, Henry-Labordère, Hobson, Huesmann, Perkowski, Proemel, Kallblad, Klimmek, Oblój, Siorpaes, Soner, Spoida, Stebegg, Tan, Touzi, Zaev....**

Optimal Martingale Transport Problem

Existence of optimal π again follows from **weak compactness.**

[***Graphical solution (mapping solution) not available***]
π is martingale $\int y d\pi_x(y) = x$
$\Rightarrow$
for π to move a unit mass, it has to **split the mass**!



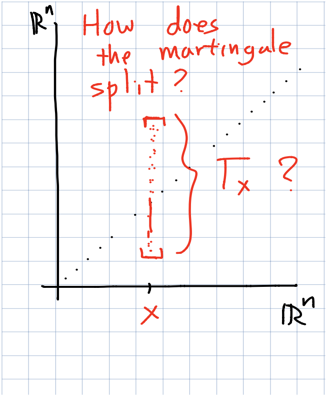
So, π **cannot** be supported on the graph $\{(x, T(x))\}$ of a map $T : \mathbf{R}^n \to \mathbf{R}^n$,
unless the trivial case $\mu = \nu$.

Optimal Martingale Transport Problem

Question: How does it split?

Let

- $\pi \in MT(\mu, \nu)$ optimal solution
- $\Gamma \subset \mathbf{R}^n \times \mathbf{R}^n$: concentration set of π, i.e. $\pi(\Gamma) = 1$
- $\Gamma_x = \Gamma \cap (\{x\} \times \mathbf{R}^n)$ the vertical slice at x (the **"Splitting set"**)



Question :

- What is the structure of π, or the set Γ, especially Γ_x?
- When is π unique?

From now on, we will focus on the case:

- $\mu \ll$ *Lebesgue*.
-

$$c(x, y) = |x - y|$$

1-dimensional results

Theorem (Hobson-Neuberger '13, Beiglböck-Juillet '13)

Suppose $n = 1$ *and*

- $c(x, y) = |x - y|$
- $\mu \leq_C \nu$ *on* $\mathbf{R}$ *and* $\mu << \mathcal{L}^1$.
- $\pi \in MT(\mu, \nu)$ *optimal solution (for max / min).*
- *Assume* $\mu \wedge \nu = 0$ *for the minimization problem.*

Then

- *There exists* $\Gamma \subset \mathbf{R} \times \mathbf{R}$*: concentration set of* π*, i.e.* $\pi(\Gamma) = 1$*, such that for a.e.* $x \in \mathbf{R}$*,*

$$\#(\Gamma_x) \leq 2, \quad \text{for } \Gamma_x = \Gamma \cap (\{x\} \times \mathbf{R}),$$

that is, the disintegration (conditional probability) π_x *is concentrated on at most two points.*

- *In particular, the optimal solution* π *is unique.*

1-dimensional results

Theorem (Hobson-Neuberger '13, Beiglböck-Juillet '13)

Suppose $n = 1$ *and*

- $c(x, y) = |x - y|$
- $\mu \leq_C \nu$ *on* **R** *and* $\mu << \mathcal{L}^1$.
- $\pi \in MT(\mu, \nu)$ *optimal solution (for max / min).*
- *Assume* $\mu \wedge \nu = 0$ *for the minimization problem.*

Then

- *There exists* $\Gamma \subset \mathbf{R} \times \mathbf{R}$*: concentration set of* π*, i.e.* $\pi(\Gamma) = 1$*, such that for a.e.* $x \in \mathbf{R}$*,*

$$\#(\Gamma_x) \leq 2, \quad \text{for } \Gamma_x = \Gamma \cap (\{x\} \times \mathbf{R}),$$

  that is, the disintegration (conditional probability) π_x *is concentrated on at most two points.*
- *In particular, the optimal solution* π *is unique.*

Higher dimensions?

Theorem (Dimension reduction. Ghoussoub, K. & Lim)

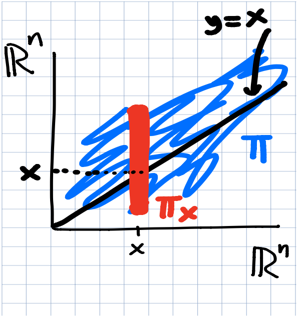
Assume:

- $c(x, y) = |x - y|$
- $\mu << \mathcal{L}^n$
- $\pi \in MT(\mu, \nu)$ *be optimal.*

Then the following holds:

- *There is concentration set of* π, $\Gamma \subset \mathbf{R}^n \times \mathbf{R}^n$ *such that*

$$\dim(\Gamma_x) \leq n - 1 \quad \text{for } \mu\text{-almost every } x,$$



Higher dimensions?

Theorem (Dimension reduction. Ghoussoub, K. & Lim)

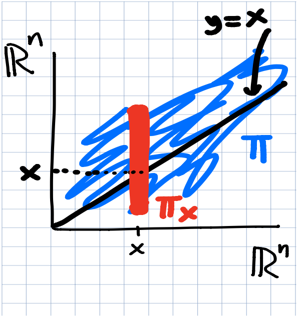
Assume:

- $c(x, y) = |x - y|$
- $\mu << \mathcal{L}^n$
- $\pi \in MT(\mu, \nu)$ *be optimal.*

Then the following holds:

- *There is concentration set of* π, $\Gamma \subset \mathbf{R}^n \times \mathbf{R}^n$ *such that*

$$\dim(\Gamma_x) \le n - 1 \quad \textit{for } \mu\textit{-almost every } x,$$



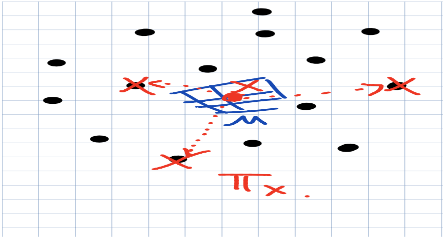
Theorem (Discrete target. Ghousshoub, K. & Lim)

If *furthermore,* ν *is discrete* $\nu = \sum_{k=1}^{\infty} q_i \delta_{y_i}$,

then *for* μ *a.e.* x*, under the optimal martingale transport,*

$$x \mapsto \; n+1 \text{ vertices of a } n\text{-dimensional simplex in } \mathbf{R}^n.$$

Moreover. the optimal solution is unique.



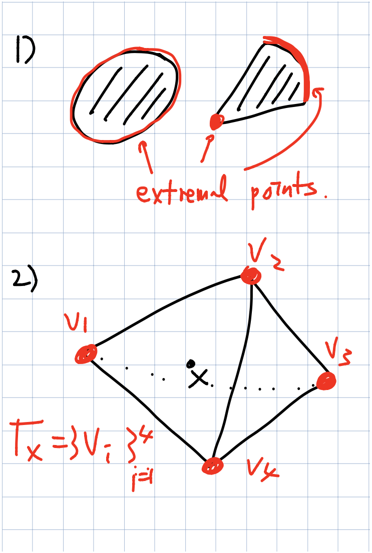
Conjectures in higher dimensions. [Ghousshoub, K. & Lim]

Assume:

- $c(x, y) = |x - y|$
- $\mu << \mathcal{L}^n$
- $\pi \in MT(\mu, \nu)$ be optimal.

Conjecture: Then, $\exists$ concentration set Γ, such that for μ almost every x,

$$\overline{\Gamma}_x = \mathrm{Ext}\left(\mathrm{conv}(\overline{\Gamma}_x)\right).$$



Progress towards the conjecture

Assume:

- $c(x, y) = |x - y|$
- $\mu << \mathcal{L}^n$
- $\pi \in MT(\mu, \nu)$ be optimal.

Conjecture: Then, $\exists$ concentration set Γ, such that for μ almost every x,

$$\overline{\Gamma}_x = \mathrm{Ext}\left(\mathrm{conv}(\overline{\Gamma}_x)\right).$$

Theorem (Ghoussuob, K. & Lim)

Conjecture 1 holds in the following cases:

- *$n = 2$, or*
- *ν is obtained from μ by diffusion with respect to a time-dependent elliptic operator. More generally, if there is a stopping time $T > 0$ of a Brownian motion with $B_0 \sim \mu$ and $B_T \sim \nu$.*

Key principle

- Duality

Duality

- Duality (e.g, [Beiglböck-Juillet '13])
  - $$\inf_{\pi \in MT(\mu,\nu)} \int c(x,y) d\pi(x,y)$$
    $$= \sup\{\int \beta(y) d\nu(x) - \int \alpha(x) d\mu(x) \ :$$
    $$\beta(y) \leq c(x,y) + \alpha(x) + \gamma(x) \cdot (y-x), \ \ \forall x, y\}$$
  - $$\sup_{\pi \in MT(\mu,\nu)} \int c(x,y) d\pi(x,y)$$
    $$= \inf\{\int \beta(y) d\nu(x) - \int \alpha(x) d\mu(x) \ :$$
    $$\beta(y) \geq c(x,y) + \alpha(x) + \gamma(x) \cdot (y-x), \ \ \forall x, y\}$$
- If the maximizer/minimizer (α, β, γ) exists, then the set,

  saturation set: $\Gamma = \{(x,y) \mid \beta(y) = c(x,y) + \alpha(x) + \gamma(x) \cdot (y-x)\}$

  gives a concentration set of an optimal π.
  In this case, we say "π admits a dual".

Question Can one always have a dual (α, β, γ) for an optimal π?

Answer

No! [Beiglböck-Juillet '13]

Counterexample: For the maximization problem,

$\mu = \nu$ cannot attain dual (Exercise: Otherwise, γ must be $\pm\infty$ on $[0, 1]$.)

(The term $\gamma(x) \cdot (y - x)$ is the trouble maker.)

We do not know for the minimization problem in general.

Question Can one always have a dual (α, β, γ) for an optimal π?

Answer
No! [Beiglböck-Juillet '13]
Counterexample: For the maximization problem,
$\mu = \nu$ cannot attain dual (Exercise: Otherwise, γ must be $\pm\infty$ on $[0, 1]$.)
(The term $\gamma(x) \cdot (y - x)$ is the trouble maker.)

We do not know for the minimization problem in general.

There are cases where dual functions exist

Theorem (Ghoussoub, K., & Lim)

The dual functions (locally) exist for an optimal $\pi \in MT(\mu, \nu)$ if

- $\mu << Leb$*, compactly supported*
- ν *is obtained from* μ *by diffusion with respect to a time-dependent elliptic operator. More generally, if there is a stopping time* $T > 0$ *of a Brownian motion with* $B_0 \sim \mu$ *and* $B_T \sim \nu$.

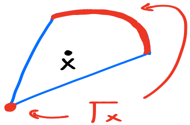
It is good to have dual functions.

If dual functions are attained.

Lemma (Ghoussoub, K. & Lim '15)

Let $c = |x - y|$. Suppose a dual (α, β, γ) is attained and Γ its saturation set.
Then *for a.e. x*

$$\overline{\Gamma}_x = \mathrm{Ext}\left(\mathrm{conv}(\overline{\Gamma}_x)\right).$$



Proof.
" Differentiate the duality relation to get information!"

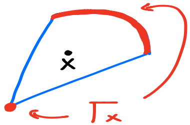
□

If dual functions are attained.

Lemma (Ghoussoub, K. & Lim '15)

Let $c = |x - y|$*. Suppose a dual* (α, β, γ) *is attained and* Γ *its saturation set.*
Then *for a.e.* x

$$\overline{\Gamma}_x = \mathrm{Ext}\left(\mathrm{conv}(\overline{\Gamma}_x)\right).$$



Proof.

" Differentiate the duality relation to get information!"

□

Proof continued.

duality relation (for the minimization problem)

$$\beta(y) \le c(x,y) + \alpha(x) + \gamma(x)\cdot(y-x)\ \forall x \in X_\Gamma, y \in Y_\Gamma,$$
$$\beta(y) = c(x,y) + \alpha(x) + \gamma(x)\cdot(y-x)\ \forall (x,y) \in \Gamma.$$

If $(x,y) \in \Gamma$,

$$|x-y| + \gamma(x)\cdot(y-x) + \alpha(x) \le |x'-y| + \gamma(x')\cdot(y-x') + \alpha(x') \quad \forall x'$$
$$\Rightarrow \nabla_x(|x-y| + \gamma(x)\cdot(y-x) + \alpha(x))$$
$$= \frac{x-y}{|x-y|} + \nabla\gamma(x)\cdot(y-x) - \gamma(x) + \nabla\alpha(x) = 0.$$

Now suppose that we can find $\{y, y_0, ..., y_s\} \subset \overline{\Gamma}_x$ with $y = \Sigma_{i=0}^s p_i y_i$, $\Sigma_{i=0}^s p_i = 1$, $p_i > 0$. Then we get

$$\frac{x-y}{|x-y|} = \sum_{i=0}^{s} p_i \frac{x-y_i}{|x-y_i|}.$$

But this can hold only if all y_i lie on the same ray emanated from x. Hence... □

Proof continued.

duality relation (for the minimization problem)

$$\beta(y) \leq c(x,y) + \alpha(x) + \gamma(x)\cdot(y-x) \ \forall x \in X_\Gamma, y \in Y_\Gamma,$$
$$\beta(y) = c(x,y) + \alpha(x) + \gamma(x)\cdot(y-x) \ \forall (x,y) \in \Gamma.$$

If $(x,y) \in \Gamma$,

$$|x-y| + \gamma(x)\cdot(y-x) + \alpha(x) \leq |x'-y| + \gamma(x')\cdot(y-x') + \alpha(x') \quad \forall x'$$
$$\Rightarrow \nabla_x(|x-y| + \gamma(x)\cdot(y-x) + \alpha(x))$$
$$= \frac{x-y}{|x-y|} + \nabla\gamma(x)\cdot(y-x) - \gamma(x) + \nabla\alpha(x) = 0.$$

Now suppose that we can find $\{y, y_0, ..., y_s\} \subset \overline{\Gamma}_x$ with $y = \Sigma_{i=0}^s p_i y_i$, $\Sigma_{i=0}^s p_i = 1$, $p_i > 0$. Then we get

$$\frac{x-y}{|x-y|} = \sum_{i=0}^{s} p_i \frac{x-y_i}{|x-y_i|}.$$

But this can hold only if all y_i lie on the same ray emanated from x. Hence...

□

Summary: the conclusion under dual attainment

Theorem (Ghousshoub, K. & Lim '15)

Let

- $c(x, y) = |x - y|$,
- $\mu << \mathcal{L}^n$,
- $\pi \in MT(\mu, \nu)$: *optimal solution for martingale transport problem.*

Suppose that π *admits a dual* (α, β, γ). *Let*

$$\Gamma = \{(x, y) \in \mathbf{R}^d \times \mathbf{R}^d \,|\, \beta(y) = c(x, y) + \alpha(x) + \gamma(x) \cdot (y - x)\}.$$

Then Γ *is a concentration set of* π, *(i.e.* $\pi(\Gamma) = 1$*), and for* μ *a.e.* x,

$$\overline{\Gamma}_x = \mathrm{Ext}\left(\mathrm{conv}(\overline{\Gamma}_x)\right).$$

Summary: the conclusion under dual attainment

Theorem (Ghousshoub, K. & Lim '15)

Let

- $c(x,y) = |x-y|$,
- $\mu << \mathcal{L}^n$,
- $\pi \in MT(\mu,\nu)$: *optimal solution for martingale transport problem.*

Suppose that π *admits a dual* (α,β,γ). *Let*

$$\Gamma = \{(x,y) \in \mathbf{R}^d \times \mathbf{R}^d | \ \beta(y) = c(x,y) + \alpha(x) + \gamma(x)\cdot(y-x)\}.$$

Then Γ *is a concentration set of* π, *(i.e.* $\pi(\Gamma) = 1$*), and*
for μ *a.e.* x,

$$\overline{\Gamma}_x = \mathrm{Ext}\left(\mathrm{conv}(\overline{\Gamma}_x)\right).$$

For general cases where we do not have dual functions:

Partition:
Make partition into duality attainable components!

For general cases where we do not have dual functions:

Theorem (Beiglböck-Juillet '13)

Suppose

- $c : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ *continuous.*
- $\pi \in MT(\mu, \nu)$*: an optimal solution for martingale transport problem.*

Then *there exists a concentration set* $\Gamma \subset \mathbf{R}^n \times \mathbf{R}^n$*, (i.e.* $\pi(\Gamma) = 1$*)*
such that Γ *is* **monotone***,*
that is, any **finite subset** $H \subset \Gamma$ *admits a dual.*

Partition into dual attainable components.

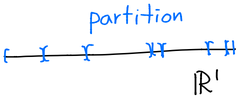
Theorem (Beiglböck-Juillet '13 for 1dim, Ghousshoub, K. & Lim '15 for general dim)

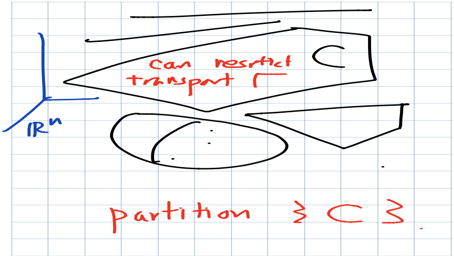
Suppose

- $c : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ *continuous.*
- $\pi \in MT(\mu, \nu)$*: an optimal solution for martingale transport problem.*

Then *there exists a concentration set* $\Gamma \subset \mathbf{R}^n \times \mathbf{R}^n$, *(i.e.* $\pi(\Gamma) = 1$*):*

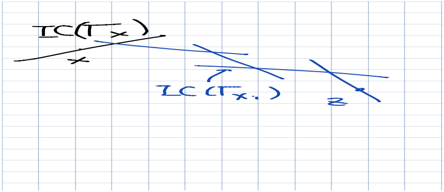
- *One can define mutually disjoint convex sets* $\{C\}$
- *such that "transport"* Γ *is partitioned on* C*'s,*
- *and on each such component* C*, the set* Γ *attains a dual.*





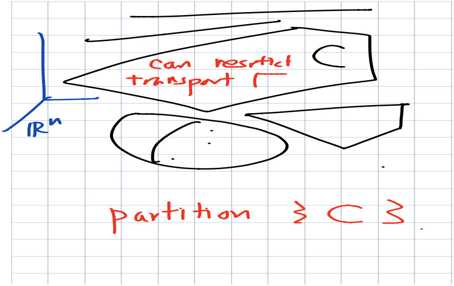
Convex Partition in n-dimensions

- $x \sim_1 z$ equivalence relation if there is a chain of $IC(\Gamma_x) := int(conv\Gamma_{x_i})$'s



- Get partition for $\sim_1$.
  Rmk: In 1-dim, we can stop here.
- Take convex hull for each component of $\sim_1$.
- Define equivalence relation $\sim_2$ using chains of those convex hulls
- Iterate this procedure on and on,
- to get equivalence relation $\sim$ and corresponding "convex" partition $\{C\}$ generated by Γ.
- It can be shown (highly nontrivial) that each such component C attains dual!

Now, for each such component, dual is attained.



The method of [Ghousoub, K. & Lim '15]:

Disintegrate μ and ν into partition $\{C\}$, each of which attains dual.

If the disintegration of μ on each C is **absolutely continuous**,

to use the dual functions and their a.e. differentiability to get the structural result for μ-a.e. x.

Partition can be useful **only if** we know good disintegration of μ along it.

But unfortunately, getting such a **good disintegration** is NOT clear in general.

Nikodym set [Ambrosio, Kirchheim, and Pratelli '04]
There is a Nikodym set in $\mathbf{R}^3$,

- having full measure in the unit cube,
- intersecting each element of a family of pairwise disjoint open lines
- only at one point.

This means, the point where we have differentiability of dual may not, in general, belong to the set we want.

Partition can be useful **only if** we know good disintegration of μ along it.

But unfortunately, getting such a **good disintegration** is NOT clear in general.

Nikodym set [Ambrosio, Kirchheim, and Pratelli '04]
There is a Nikodym set in $\mathbf{R}^3$,

- having full measure in the unit cube,
- intersecting each element of a family of pairwise disjoint open lines
- only at one point.

This means, the point where we have differentiability of dual may not, in general, belong to the set we want.

Partition can be useful **only if** we know good disintegration of μ along it.

But unfortunately, getting such a **good disintegration** is NOT clear in general.

Nikodym set [Ambrosio, Kirchheim, and Pratelli '04]

There is a Nikodym set in $\mathbf{R}^3$,

- having full measure in the unit cube,
- intersecting each element of a family of pairwise disjoint open lines
- only at one point.

This means, the point where we have differentiability of dual may not, in general, belong to the set we want.

Still can handle dimension question even without good disintegration:

Corollary (Ghoussoub, K. & Lim '15)

Suppose

- $c(x, y) = |x - y|$
- $\pi \in MT(\mu, \nu)$ *optimal*
- $\mu << \mathcal{L}^n$.

Then, *there is a concentration set* Γ *of* π*, such that for* μ*-almost every* x*,*

$$\dim \Gamma_x \leq n - 1.$$

Proof.

- If $\dim C = n$, then C is open, thus, μ can be restricted on C, so absolutely continuous on C! Apply previous results.
- For other components with $\dim C \leq n - 1$, but, in this case already the dimension is $\leq n - 1$.

□

A case with good disintegration: discrete target, thus countable partition components

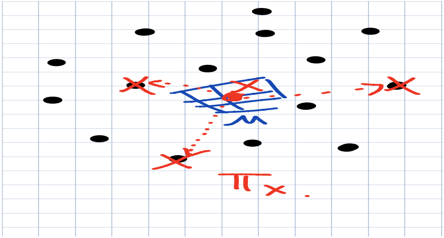
Theorem (Discrete target. Ghousshoub, K. & Lim '15)

If *furthermore,* ν *is discrete* $\nu = \sum_{k=1}^{\infty} q_i \delta_{y_i}$,

then *for* μ *a.e. x, under the optimal martingale transport,*

$$x \mapsto \; n+1 \text{ vertices of a } n\text{-dimensional simplex in } \mathbf{R}^n.$$

Moreover. the optimal solution is unique.



A case with good disintegration: two dimensions

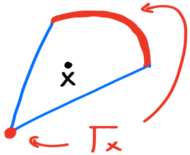
Theorem (Ghoussoub, K. & Lim '15 $n = 2$)

Suppose

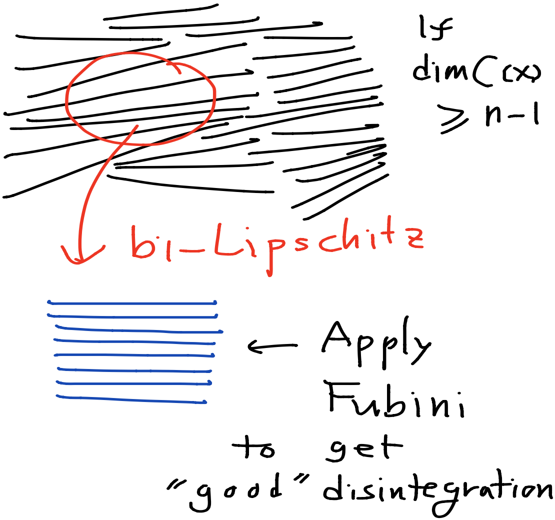
- $c(x, y) = |x - y|$,
- $\pi \in MT(\mu, \nu)$ *optimal,*
- $\mu << \mathcal{L}^n$,
- $n = 2$,

Then, *there is a concentration set* Γ *of* π*, such that for* μ*-almost every* x*,*

$$\overline{\Gamma}_x = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma}_x)).$$



Codimension ≤ 1 case. **Idea:** Flattening!



Summary:

To study the structure of optimal martingale transport in $MT(\mu, \nu)$ with $\mu << \mathcal{L}^n$ in **general dimensions** n:

- Find **optimal** martingale plan $\pi \in MT(\mu, \nu)$ using compactness.
- Get a suitable **monotone set** Γ.
- Apply the **partition** of Γ into **duality** attainable components C.
- Get **dual functions** α, β, γ for Γ in C.
- **Almost everywhere differentiability** of α, γ on C.

If μ disintegrates into an **absolutely continuous measure** μ_C on each component C,

- Get the **structure** of Γ (of Γ_x for μ_C a.e. x) in each C from **almost everywhere differentiability**,
- thus finally, get the **structure** of Γ (of Γ_x for μ-a.e. x)!

Summary:

To study the structure of optimal martingale transport in $MT(\mu,\nu)$ with $\mu << \mathcal{L}^n$ in **general dimensions** n:

- Find **optimal** martingale plan $\pi \in MT(\mu,\nu)$ using compactness.
- Get a suitable **monotone set** Γ.
- Apply the **partition** of Γ into **duality** attainable components C.
- Get **dual functions** α, β, γ for Γ in C.
- **Almost everywhere differentiability** of α, γ on C.

If μ disintegrates into an **absolutely continuous measure** μ_C on each component C,

- Get the **structure** of Γ (of Γ_x for μ_C a.e. x) in each C from **almost everywhere differentiability**,
- thus finally, get the **structure** of Γ (of Γ_x for μ-a.e. x)!

Some related work:

- **[Beiglböck, Nutz, & Touzi '15]** : quasi-sure duality.
- **[De March& Touzi '17] [Oblój & Siorpaes '17]**: canonical partition for martingale transport.

Thank You Very Much!