Limit Theorems for Optimal Mass Transportation and Applications to Networks

Gershon Wolansky

Department of Mathematics, Technion
32000 Haifa, ISRAEL

E-mail: gershonw@math.technion.ac.il

S. Petersburg, May 2010

- Review of Kantorovich metrics on the space of positive measures

- Review of Kantorovich metrics on the space of positive measures
- Conditioned Kantorovich metrics and relation to metrics on 1-D graphs

- Review of Kantorovich metrics on the space of positive measures
- Conditioned Kantorovich metrics and relation to metrics on 1-D graphs
- Cost function for transporting networks

- Review of Kantorovich metrics on the space of positive measures
- Conditioned Kantorovich metrics and relation to metrics on 1-D graphs
- Cost function for transporting networks
- Replacing optimal networks by points allocation?

- Review of Kantorovich metrics on the space of positive measures
- Conditioned Kantorovich metrics and relation to metrics on 1-D graphs
- Cost function for transporting networks
- Replacing optimal networks by points allocation?
- Generalization to Lagrangian action on compact manifolds

Definition

The Kantorovich metric for $\lambda^-, \lambda^+ \in \mathcal{B}_+$ satisfying $\int d\lambda^- = \int d\lambda^+$

$$W_p(\lambda^+, \lambda^-) = \left\{ \inf_\Lambda \int_\Omega \int_\Omega |x-y|^p d\Lambda \right\}^{1/p}$$

Where $\Lambda \in \mathcal{B}^+(\Omega \times \Omega)$, $\pi_{1,\#}\Lambda = \lambda^+$, $\pi_{2,\#}\Lambda = \lambda^-$.

Definition

The Kantorovich metric for $\lambda^-, \lambda^+ \in \mathcal{B}_+$ satisfying $\int d\lambda^- = \int d\lambda^+$

$$W_p(\lambda^+, \lambda^-) = \left\{ \inf_{\Lambda} \int_\Omega \int_\Omega |x-y|^p d\Lambda \right\}^{1/p}$$

Where $\Lambda \in \mathcal{B}^+(\Omega \times \Omega)$, $\pi_{1,\#}\Lambda = \lambda^+$, $\pi_{2,\#}\Lambda = \lambda^-$.

In case $p = 1$, $W_1(\lambda^+, \lambda^-)$ depends only on $\lambda = \lambda^+ - \lambda^- \in \mathcal{B}_0$.

Definition

The Kantorovich metric for $\lambda^-, \lambda^+ \in \mathcal{B}_+$ satisfying $\int d\lambda^- = \int d\lambda^+$

$$W_p(\lambda^+, \lambda^-) = \left\{ \inf_{\Lambda} \int_\Omega \int_\Omega |x-y|^p d\Lambda \right\}^{1/p}$$

Where $\Lambda \in \mathcal{B}^+(\Omega \times \Omega)$, $\pi_{1,\#}\Lambda = \lambda^+$, $\pi_{2,\#}\Lambda = \lambda^-$.

In case $p = 1$, $W_1(\lambda^+, \lambda^-)$ depends only on $\lambda = \lambda^+ - \lambda^- \in \mathcal{B}_0$. An equivalent definition

Definition

$$W_1(\lambda) = \sup_{\phi \in Lip_1(\Omega)} \int_\Omega \phi d\lambda$$

Where $Lip_1(\Omega) := \{\phi \in C(\Omega)\ ; \phi(x) - \phi(y) \le |x-y| \ \ \forall x, y \in \Omega\}$

Example:

If

$$\lambda^{+} = \sum_{1}^{N} m_i \delta_{x_i} \quad ; \quad \lambda^{-} = \sum_{1}^{N} m_i^* \delta_{y_i} \tag{1}$$

subjected to $\sum_1^N m_i = \sum_i^N m_i^* = 1$, then

Example:

If

$$\lambda^+ = \sum_1^N m_i \delta_{x_i} \quad ; \quad \lambda^- = \sum_1^N m_i^* \delta_{y_i} \tag{1}$$

subjected to $\sum_1^N m_i = \sum_i^N m_i^* = 1$, then

$$W_p(\lambda) = \left[\min_{\Lambda} \sum_1^N \sum_1^N \lambda^{i,j} |x_i - y_j|^p \right]^{1/p}$$

where $\Lambda = \{\lambda^{i,j}\}$ ie the set of all non-negative $N \times N$ matrices satisfying

$$\sum_{j=1}^n \lambda^{i,j} = m_i \quad ; \quad \sum_{i=1}^n \lambda^{i,j} = m_j^*$$

- Weak continuity: we may approximate $\lambda^{\pm}$ by atomic measures.

- Weak continuity: we may approximate $\lambda^{\pm}$ by atomic measures.
- Then the optimal plan is an atomic measure as well, solvable in the set of bi-stochastic matrices $\{\Lambda_{i,j}\}$.

- Weak continuity: we may approximate $\lambda^{\pm}$ by atomic measures.
- Then the optimal plan is an atomic measure as well, solvable in the set of bi-stochastic matrices $\{\Lambda_{i,j}\}$. .
- This "discrete" plan is an approximation in the weak topology of Λ.

- Weak continuity: we may approximate $\lambda^{\pm}$ by atomic measures.
- Then the optimal plan is an atomic measure as well, solvable in the set of bi-stochastic matrices $\{\Lambda_{i,j}\}$. .
- This "discrete" plan is an approximation in the weak topology of Λ.
- An optimal map is sometimes deterministic:

$$W_p^p(\lambda^+, \lambda^-) = \inf_{T_\# \lambda^+ = \lambda^-} \int |x - T(x)|^p d\lambda^+$$

where $T_\# \lambda^+(B) = \lambda^-\left(T^{-1}(B)\right)$. Then $\Lambda(dxdy) = \lambda^+(dx)\delta_{y-T(x)}dy$ is the optimal plan.

- Weak continuity: we may approximate $\lambda^{\pm}$ by atomic measures.
- Then the optimal plan is an atomic measure as well, solvable in the set of bi-stochastic matrices $\{\Lambda_{i,j}\}$. .
- This "discrete" plan is an approximation in the weak topology of Λ.
- An optimal map is sometimes deterministic:

$$W_p^p(\lambda^+,\lambda^-) = \inf_{T_\#\lambda^+=\lambda^-} \int |x - T(x)|^p d\lambda^+$$

  where $T_\#\lambda^+(B) = \lambda^-\left(T^{-1}(B)\right)$. Then $\Lambda(dxdy) = \lambda^+(dx)\delta_{y-T(x)}dy$ is the optimal plan.
- If $p > 1$ then T is obtained in terms of a "potential function" Φ. In particular, $p = 2$ and λ^+ is continuous w.r to Lebesgue measure than $T(x) = \nabla\Phi(x)$ where Φ is a convex function, and this T is unique. (Brenier, McCann, Gangbo, Caffarelli)

- For the case $p = 1$, the optimal potential ϕ gives only partial information on the optimal mapping

$$T(x) = x + t\nabla\phi(x)$$

where t is unknown (change with x).

- For the case $p = 1$, the optimal potential ϕ gives only partial information on the optimal mapping

$$T(x) = x + t\nabla\phi(x)$$

where t is unknown (change with x).

- The solvability of optimal map in the metric case ($p = 1$) is a difficult problem. First attempt by Sudakov (1979).

- For the case $p = 1$, the optimal potential ϕ gives only partial information on the optimal mapping

$$T(x) = x + t\nabla\phi(x)$$

where t is unknown (change with x).
- The solvability of optimal map in the metric case ($p = 1$) is a difficult problem. First attempt by Sudakov (1979).
- Equivalent formulation (Beckmann (1952))

Definition

$$W_1(\lambda) = \inf \int |d\vec{m}|$$

subject to $\nabla \cdot \vec{m} = \lambda$.

- For the case $p = 1$, the optimal potential ϕ gives only partial information on the optimal mapping

$$T(x) = x + t\nabla\phi(x)$$

where t is unknown (change with x).

- The solvability of optimal map in the metric case ($p = 1$) is a difficult problem. First attempt by Sudakov (1979).
- Equivalent formulation (Beckmann (1952))

Definition

$$W_1(\lambda) = \inf \int |d\vec{m}|$$

subject to $\nabla \cdot \vec{m} = \lambda$.

- The optimal $\vec{m} := \rho\nabla\phi$ yields a complete information on T.

- For the case $p = 1$, the optimal potential ϕ gives only partial information on the optimal mapping

$$T(x) = x + t\nabla\phi(x)$$

where t is unknown (change with x).
- The solvability of optimal map in the metric case ($p = 1$) is a difficult problem. First attempt by Sudakov (1979).
- Equivalent formulation (Beckmann (1952))

Definition

$$W_1(\lambda) = \inf \int |d\vec{m}|$$

subject to $\nabla \cdot \vec{m} = \lambda$.

- The optimal $\vec{m} := \rho\nabla\phi$ yields a complete information on T.
- There is an interest in calculating *the Transport Measure* $\rho := |\vec{m}|$, and verifies

$$\nabla \cdot (\rho\nabla\phi) = \lambda \ .$$

- Next attempts by Gangbo and Evans (1999): Approximating Lip_1 by $|\nabla\phi|_p$ where $p \to \infty$.

- Next attempts by Gangbo and Evans (1999): Approximating Lip_1 by $|\nabla\phi|_p$ where $p \to \infty$.
  minimizing
  $$p^{-1} \int |\nabla\phi|^p - \int \phi d\lambda$$
  leads to
  $$\nabla \cdot \left(|\nabla\phi|^{p-2} \nabla\phi \right) = \lambda$$

- Next attempts by Gangbo and Evans (1999): Approximating Lip_1 by $|\nabla\phi|_p$ where $p \to \infty$.
  minimizing
  $$p^{-1}\int|\nabla\phi|^p - \int \phi d\lambda$$
  leads to
  $$\nabla\cdot\left(|\nabla\phi|^{p-2}\nabla\phi\right) = \lambda$$
  and to the approximation $|\nabla\phi|^{p-2} \to \rho$ while $|\nabla\phi| \to 1$.

- Next attempts by Gangbo and Evans (1999): Approximating Lip_1 by $|\nabla\phi|_p$ where $p \to \infty$.
  minimizing

$$p^{-1}\int |\nabla\phi|^p - \int \phi d\lambda$$

  leads to

$$\nabla \cdot \left(|\nabla\phi|^{p-2}\nabla\phi\right) = \lambda$$

  and to the approximation $|\nabla\phi|^{p-2} \to \rho$ while $|\nabla\phi| \to 1$.
- Other approaches by Trudinger, Wang, Ma, Caffarelli, Feldman, McCann Ambrosio, Pratelli... in the last decade.

Conditional W_1 distance

Conditional W_1 distance

Definition

Define, for $\mu \in \mathcal{B}_1^+(\Omega)$, $\lambda \in \mathcal{B}_0(\Omega)$ and $p > 1$

$$W_1^{(p)}(\lambda\|\mu) := \sup_{0 \not\equiv \nabla\phi \in C^1(\Omega)} \frac{\int_\Omega \phi d\lambda}{\left(\int_\Omega |\nabla\phi|^q d\mu\right)^{1/q}}$$

where $q = p/(p-1)$.

Theorem

$$W_1(\lambda) = \inf_{\mu \in \mathcal{B}_1^+} W_1^{(p)}(\lambda \| \mu)$$

If $p = 2$ then any minimizer μ is a Transport measure supported in an optimal plan of $W_1(\lambda)$.

Theorem

$$W_1(\lambda) = \inf_{\mu \in \mathcal{B}_1^+} W_1^{(p)}(\lambda \| \mu)$$

If $p = 2$ then any minimizer μ is a Transport measure supported in an optimal plan of $W_1(\lambda)$.

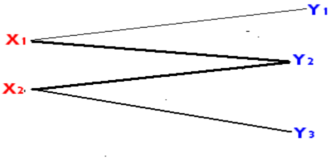
Example: $\lambda = m_1\delta_{x_1} + m_2\delta_{x_2} - m_1^*\delta_{y_1} - m_2^*\delta_{y_2} - m_3^*\delta_{y_3}$

Theorem

$$W_1(\lambda) = \inf_{\mu \in \mathcal{B}_1^+} W_1^{(p)}(\lambda \| \mu)$$

If $p = 2$ then any minimizer μ is a Transport measure supported in an optimal plan of $W_1(\lambda)$.

Example: $\lambda = m_1\delta_{x_1} + m_2\delta_{x_2} - m_1^*\delta_{y_1} - m_2^*\delta_{y_2} - m_3^*\delta_{y_3}$



Disadvantage of using $W_1^{(p)}(\lambda\|\mu)$ for calculating transport measures:

Disadvantage of using $W_1^{(p)}(\lambda \| \mu)$ for calculating transport measures:
$W_1^{(p)}(\lambda \| \mu)$ is not continuous in μ.

Disadvantage of using $W_1^{(p)}(\lambda\|\mu)$ for calculating transport measures:
$W_1^{(p)}(\lambda\|\mu)$ is not continuous in μ. In particular

$$W_1^{(p)}(\lambda\|\mu_n) = \infty$$

for any atomic measure μ_n.

Disadvantage of using $W_1^{(p)}(\lambda\|\mu)$ for calculating transport measures: $W_1^{(p)}(\lambda\|\mu)$ is not continuous in μ. In particular

$$W_1^{(p)}(\lambda\|\mu_n) = \infty$$

for any atomic measure μ_n.

Thus, we cannot approximate μ as a limit of atomic measures.

"Proof":

$$\inf_{\mu\in\mathcal{B}_1^+}\sup_{0\not\equiv\phi\in C^1(\Omega)}\frac{\int_\Omega \phi d\lambda}{\left(\int_\Omega |\nabla\phi|^q d\mu\right)^{1/q}} = \sup_{0\not\equiv\phi\in C^1(\Omega)}\inf_{\mu\in\mathcal{B}_1^+}\frac{\int_\Omega \phi d\lambda}{\left(\int_\Omega |\nabla\phi|^q d\mu\right)^{1/q}}$$

while

$$\sup_{\mu\in\mathcal{B}_1^+}\int_\Omega |\nabla\phi|^q d\mu = \sup_{x\in\Omega}|\nabla\phi(x)|^q = Lip^q(\phi)$$

"Proof":

$$\inf_{\mu\in\mathcal{B}_1^+}\sup_{0\not\equiv\phi\in C^1(\Omega)}\frac{\int_\Omega\phi d\lambda}{\left(\int_\Omega|\nabla\phi|^q d\mu\right)^{1/q}}=\sup_{0\not\equiv\phi\in C^1(\Omega)}\inf_{\mu\in\mathcal{B}_1^+}\frac{\int_\Omega\phi d\lambda}{\left(\int_\Omega|\nabla\phi|^q d\mu\right)^{1/q}}$$

while

$$\sup_{\mu\in\mathcal{B}_1^+}\int_\Omega|\nabla\phi|^q d\mu=\sup_{x\in\Omega}|\nabla\phi(x)|^q=Lip^q(\phi)$$

In case $\lambda^+=\sum_1^N m_i\delta_{x_i}$; $\lambda^-=\sum_1^N m_i^*\delta_{y_i}$ the optimal μ is given by

$$\mu=\sum_i^N\sum_i^N\frac{\lambda^{i,j}}{|x_i-y_j|}\delta_{[x_i,y_j]}$$

with $\sum_i\lambda^{i,j}=m_j^*$; $\sum_j\lambda^{i,j}=m_i$ are the optimal transports.

Theorem

For $p > 1$

$$W_1^{(p)}(\lambda||\mu) = \Gamma - \lim_{\varepsilon\to 0} \varepsilon^{-1} W_p\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right)$$

Theorem

For $p > 1$

$$W_1^{(p)}(\lambda||\mu) = \Gamma - \lim_{\varepsilon\to 0} \varepsilon^{-1} W_p\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right)$$

$$W_1(\lambda) = \lim_{\varepsilon\to 0} \varepsilon^{-1} \inf_{\mu\in\mathcal{B}_1^+} W_p\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right)$$

Theorem

For $p > 1$

$$W_1^{(p)}(\lambda||\mu) = \Gamma - \lim_{\varepsilon\to 0} \varepsilon^{-1} W_p\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right)$$

$$W_1(\lambda) = \lim_{\varepsilon\to 0} \varepsilon^{-1} \inf_{\mu\in\mathcal{B}_1^+} W_p\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right)$$

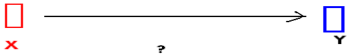
Remark

$$W_p^\varepsilon(\lambda\|\mu) := \varepsilon^{-1} W_p\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right)$$

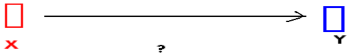
is weakly continuous in μ.

Let $\lambda = \delta_x - \delta_y$.

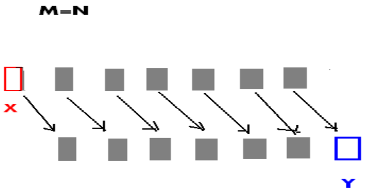
Let $\lambda = \delta_x - \delta_y$.



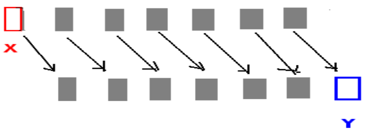
Let $\lambda = \delta_x - \delta_y$.



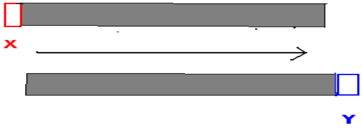
If $\varepsilon = 1/n$ then μ is displayed in the $n-$ gray shadows

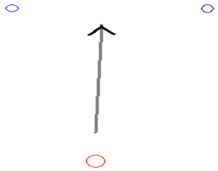

M=N
X
Y

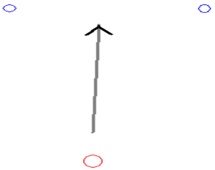
M=N

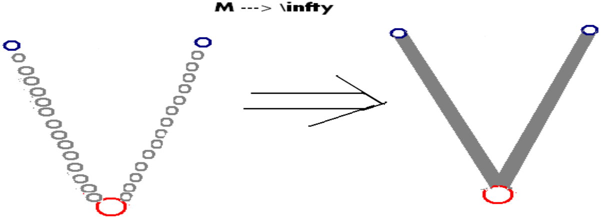


M---> \infty








M ---> \infty

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)
- A different approach, using $W_1^{(p)}(\lambda\|\mu)$:Can we restrict the conditioning measure μ to obtain optimal networks?

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)
- A different approach, using $W_1^{(p)}(\lambda\|\mu)$:Can we restrict the conditioning measure μ to obtain optimal networks?
- Suppose

$$\widehat{W}^{\Gamma}(\lambda) := \inf_{\mu} W_1^{(p)}(\lambda\|\mu)$$

where we minimize on probability measures μ *supported on a given graph* Γ*.*

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)
- A different approach, using $W_1^{(p)}(\lambda\|\mu)$:Can we restrict the conditioning measure μ to obtain optimal networks?
- Suppose

$$\widehat{W}^{\Gamma}(\lambda) := \inf_{\mu} W_1^{(p)}(\lambda\|\mu)$$

where we minimize on probability measures μ *supported on a given graph Γ.*

-

$$\widehat{W}^{\Gamma}(\lambda) = \min_{\Lambda} \int\int D_{\Gamma}(x,y)d\Lambda(x,y)$$

where D_{Γ} is the distance reduced to Γ.

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)
- A different approach, using $W_1^{(p)}(\lambda\|\mu)$:Can we restrict the conditioning measure μ to obtain optimal networks?
- Suppose

$$\widehat{W}^{\Gamma}(\lambda) := \inf_{\mu} W_1^{(p)}(\lambda\|\mu)$$

 where we minimize on probability measures μ *supported on a given graph* Γ.

-

$$\widehat{W}^{\Gamma}(\lambda) = \min_{\Lambda} \int \int D_{\Gamma}(x,y) d\Lambda(x,y)$$

 where D_{Γ} is the distance reduced to Γ.

- Can we obtain a formulation of "optimal network" by restricting to a set of probability measures supported of graphs of prescribed length?

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)
- A different approach, using $W_1^{(p)}(\lambda\|\mu)$:Can we restrict the conditioning measure μ to obtain optimal networks?
- Suppose

$$\widehat{W}^{\Gamma}(\lambda) := \inf_{\mu} W_1^{(p)}(\lambda\|\mu)$$

  where we minimize on probability measures μ *supported on a given graph* Γ*.*
-

$$\widehat{W}^{\Gamma}(\lambda) = \min_{\Lambda} \int\int D_{\Gamma}(x,y) d\Lambda(x,y)$$

  where D_Γ is the distance reduced to Γ.
- Can we obtain a formulation of "optimal network" by restricting to a set of probability measures supported of graphs of prescribed length?
- Disadvantage: This is a formidable set, not natural, not compact.

Optimal Networks

- Underlining idea: Cost of transforation depends on the flux as well. It is a increasing, concave function of the flux. (Butazzio, Xia, Stepanov ....)
- A different approach, using $W_1^{(p)}(\lambda\|\mu)$:Can we restrict the conditioning measure μ to obtain optimal networks?
- Suppose

$$\widehat{W}^{\Gamma}(\lambda) := \inf_{\mu} W_1^{(p)}(\lambda\|\mu)$$

  where we minimize on probability measures μ *supported on a given graph* Γ*.*
-

$$\widehat{W}^{\Gamma}(\lambda) = \min_{\Lambda} \int\int D_{\Gamma}(x,y)d\Lambda(x,y)$$

  where D_{Γ} is the distance reduced to Γ.
- Can we obtain a formulation of "optimal network" by restricting to a set of probability measures supported of graphs of prescribed length?
- Disadvantage: This is a formidable set, not natural, not compact. Certainly cannot be approximated by atomic measures!

Equivalent formulation:

Theorem

For $p > 1$

$$\lim_{M\to\infty} M^{1-1/p} \min_{\mu\in\mathcal{B}_M^+} W_p\left(\mu+\lambda^+,\mu+\lambda^-\right) = W_1(\lambda^+,\lambda^-)$$

where $\mathcal{B}_M^+$ *stands for the set of all positive Borel measures* μ *normalized by* $\int d\mu = M$.

Equivalent formulation:

Theorem

For $p > 1$

$$\lim_{M\to\infty} M^{1-1/p} \min_{\mu\in\mathcal{B}_M^+} W_p\left(\mu+\lambda^+,\mu+\lambda^-\right) = W_1(\lambda^+,\lambda^-)$$

where $\mathcal{B}_M^+$ *stands for the set of all positive Borel measures* μ *normalized by* $\int d\mu = M$.

Suppose we replace the condition $M \to \infty$ by the condition $n \to \infty$ where μ is restricted to the set of atomic measures $\mathcal{B}^{n,+}$ of (at most) n atoms?

Equivalent formulation:

Theorem

For $p > 1$

$$\lim_{M\to\infty} M^{1-1/p} \min_{\mu\in\mathcal{B}_M^+} W_p\left(\mu+\lambda^+, \mu+\lambda^-\right) = W_1(\lambda^+, \lambda^-)$$

where $\mathcal{B}_M^+$ *stands for the set of all positive Borel measures* μ *normalized by* $\int d\mu = M$.

Suppose we replace the condition $M \to \infty$ by the condition $n \to \infty$ where μ is restricted to the set of atomic measures $\mathcal{B}^{n,+}$ of (at most) n atoms?

Theorem

For any $q > 1$ *and* $\lambda = \sum_1^N m_i\delta_{x_i} - m_i^*\delta_{y_i}$.

Equivalent formulation:

Theorem

For $p > 1$

$$\lim_{M\to\infty} M^{1-1/p} \min_{\mu\in\mathcal{B}^+_M} W_p\left(\mu+\lambda^+, \mu+\lambda^-\right) = W_1(\lambda^+, \lambda^-)$$

where $\mathcal{B}^+_M$ *stands for the set of all positive Borel measures* μ *normalized by* $\int d\mu = M$.

Suppose we replace the condition $M \to \infty$ by the condition $n \to \infty$ where μ is restricted to the set of atomic measures $\mathcal{B}^{n,+}$ of (at most) n atoms?

Theorem

For any $q > 1$ *and* $\lambda = \sum_1^N m_i\delta_{x_i} - m_i^*\delta_{y_i}$.

$$\lim_{n\to\infty} n^{1-1/p} \inf_{\mu\in\mathcal{B}^{n,+}} W_p\left(\mu+\lambda^+, \mu+\lambda^-\right) = \widehat{W}^{(p)}(\lambda)$$

Recall

Definition

$$W_1(\lambda) = \inf \int |d\vec{m}|$$

subject to $\nabla \cdot \vec{m} = \lambda$.

Recall

Definition

$$W_1(\lambda) = \inf \int |d\vec{m}|$$

subject to $\nabla \cdot \vec{m} = \lambda$.

Definition

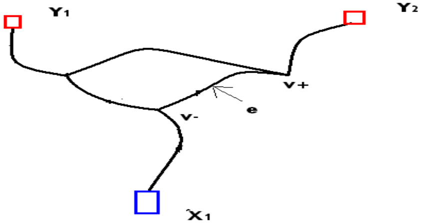
(Xia) For $p > 1$ and λ an atomic metric

$$\widehat{W}^{(p)}(\lambda) = \inf \int |\frac{d\vec{m}}{dH_1}|^{1/p} dH_1$$

subject to $\nabla \cdot \vec{m} = \lambda$.

An oriented, weighted graph (γ, m) associated with λ is a graph γ composed of vertices $V(\gamma)$ and edges $E(\gamma)$ and a function $m : E \to \mathbb{R}^+$:

An oriented, weighted graph (γ, m) associated with λ is a graph γ composed of vertices $V(\gamma)$ and edges $E(\gamma)$ and a function $m : E \to \mathbb{R}^+$:



Definition

- The set of all weighted graphs associated with λ is denoted by $\Gamma(\lambda)$.

Definition

- The set of all weighted graphs associated with λ is denoted by $\Gamma(\lambda)$.
- $$\widehat{W}^{(p)}(\lambda) := \inf_{(\gamma,m)\in\Gamma(\lambda)} \sum_{e\in E(\gamma)} |e| m_e^{1/p}$$

Examples:

- $p = 1$ (Reduced to the metric Monge problem),

Definition

- The set of all weighted graphs associated with λ is denoted by $\Gamma(\lambda)$.
- $$\widehat{W}^{(p)}(\lambda) := \inf_{(\gamma,m)\in\Gamma(\lambda)} \sum_{e\in E(\gamma)} |e| m_e^{1/p}$$

Examples:

- $p = 1$ (Reduced to the metric Monge problem),
- $p = 0$ (Reduced to Steiner problem of minimal graphs)

We may associate a weighted graph (γ, m) with an optimal plan:

$$e \iff (i,j); \gamma_{i,j} > 0 \qquad , m_e = \gamma_{i,j}$$

We may associate a weighted graph (γ, m) with an optimal plan:

$$e \Longleftrightarrow (i,j); \gamma_{i,j} > 0 \qquad , m_e = \gamma_{i,j}$$

Postulate

- $\{x_1, \ldots, y_1, \ldots\} \subset V(\gamma)$.

We may associate a weighted graph (γ, m) with an optimal plan:

$$e \Longleftrightarrow (i,j); \gamma_{i,j} > 0 \qquad , m_e = \gamma_{i,j}$$

Postulate

- $\{x_1, \dots, y_1, \dots\} \subset V(\gamma)$.
- *For each* $i \in \{1, N\}$, $\sum_{\{e, x_i \in \partial^+ e\}} m_e = m_i$ *and* $\sum_{\{e, y_i \in \partial^- e\}} m_e = m_i^*$, *where* $\partial^\pm e := v_e^\pm$.

We may associate a weighted graph (γ, m) with an optimal plan:

$$e \Longleftrightarrow (i,j); \gamma_{i,j} > 0 \qquad , m_e = \gamma_{i,j}$$

Postulate

- $\{x_1, \ldots, y_1, \ldots\} \subset V(\gamma)$.
- *For each* $i \in \{1, N\}$, $\sum_{\{e, x_i \in \partial^+ e\}} m_e = m_i$ *and* $\sum_{\{e, y_i \in \partial^- e\}} m_e = m_i^*$, *where* $\partial^\pm e := v_e^\pm$.
- *For each* $v \in V(\gamma) - \{x_1, \ldots y_N\}$, $\sum_{\{e; v \in \partial^+ e\}} m_e = \sum_{\{e; v \in \partial^- e\}} m_e$.

We may associate a weighted graph (γ, m) with an optimal plan:

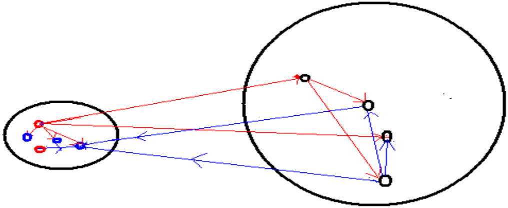
$$e \Longleftrightarrow (i,j); \gamma_{i,j} > 0 \qquad , m_e = \gamma_{i,j}$$

Postulate

- $\{x_1, \dots, y_1, \dots\} \subset V(\gamma)$.
- *For each* $i \in \{1, N\}$, $\sum_{\{e, x_i \in \partial^+ e\}} m_e = m_i$ *and* $\sum_{\{e, y_i \in \partial^- e\}} m_e = m_i^*$, *where* $\partial^\pm e := v_e^\pm$.
- *For each* $v \in V(\gamma) - \{x_1, \dots y_N\}$, $\sum_{\{e; v \in \partial^+ e\}} m_e = \sum_{\{e; v \in \partial^- e\}} m_e$.

Lemma

There exists an optimal plan $\{\gamma\}$ *whose graph contains at most* $2N^3$ *nodes of order* ≥ 3.



$\Sigma_o = m_o$ $\Sigma_o = m_o^*$ $\Sigma_o = \Sigma_o$

The set $\mathcal{B}^{n,+}$ is, evidently, not a compact one. Still we claim

Lemma

For each $n \in \mathbb{N}$, a minimizer $\mu_n \in \mathcal{B}^{n,+}$

$$\overline{W}_q(\lambda) := \inf_{\mu \in \mathcal{B}^{n,+}} W_q\left(\mu + \lambda^+, \mu + \lambda^-\right) = W_q\left(\mu_n + \lambda^+, \mu_n + \lambda^-\right)$$

exists.

The set $\mathcal{B}^{n,+}$ is, evidently, not a compact one. Still we claim

Lemma

For each $n \in \mathbb{N}$, a minimizer $\mu_n \in \mathcal{B}^{n,+}$

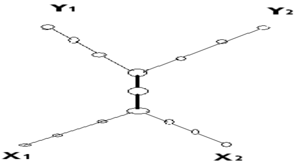
$$\overline{W}_q(\lambda) := \inf_{\mu \in \mathcal{B}^{n,+}} W_q\left(\mu + \lambda^+, \mu + \lambda^-\right) = W_q\left(\mu_n + \lambda^+, \mu_n + \lambda^-\right)$$

exists.

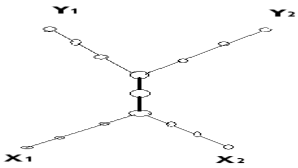
Theorem

Let μ_n be a regular minimizer of $W_q\left(\mu + \lambda^+, \mu + \lambda^-\right)$ in $\mathcal{B}^{n,+}$. Then the associated optimal plan spans a reduced weighted tree $(\hat{\gamma}_n, m_n)$ which converges (in Hausdorff metric) to an optimal graph $(\hat{\gamma}, m) \in \Gamma(\lambda)$ as $n \to \infty$,

One direction



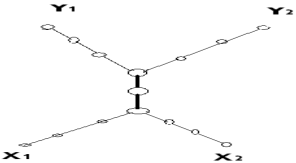
One direction



Definition

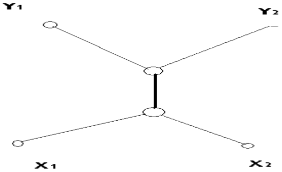
Reduced graph: Remove all nodes of degree =2.

One direction



Definition

Reduced graph: Remove all nodes of degree =2.



$$\sum_{e\in E(\hat{\gamma})} m_e^{1/p}|e| \leq \left(\sum_{e\in E(\hat{\gamma})} m_e|e|^p\right)^{1/p} |E(\hat{\gamma})|^{(p-1)/p} .$$

$$\sum_{e\in E(\hat{\gamma})} m_e^{1/p}|e| \leq \left(\sum_{e\in E(\hat{\gamma})} m_e|e|^p\right)^{1/p} |E(\hat{\gamma})|^{(p-1)/p} .$$

$$\sum_{e\in E(\hat{\gamma})} m_e|e|^p = W_p^p(\lambda^+ + \mu, \lambda^- + \mu)$$

$$\sum_{e\in E(\hat{\gamma})} m_e^{1/p}|e| \leq \left(\sum_{e\in E(\hat{\gamma})} m_e|e|^p\right)^{1/p} |E(\hat{\gamma})|^{(p-1)/p} .$$

$$\sum_{e\in E(\hat{\gamma})} m_e|e|^p = W_p^p(\lambda^+ + \mu, \lambda^- + \mu)$$

From Lemma: $|E(\hat{\gamma})|^{(p-1)/p} = n^{(p-1)/p} + o(n)$.

$$\sum_{e\in E(\hat{\gamma})} m_e^{1/p}|e| \le \left(\sum_{e\in E(\hat{\gamma})} m_e|e|^p\right)^{1/p} |E(\hat{\gamma})|^{(p-1)/p} .$$

$$\sum_{e\in E(\hat{\gamma})} m_e|e|^p = W_p^p(\lambda^+ + \mu, \lambda^- + \mu)$$

From Lemma: $|E(\hat{\gamma})|^{(p-1)/p} = n^{(p-1)/p} + o(n)$. Hence:

$$n^{(1-p)/p} W_p(\lambda^+ + \mu, \lambda^- + \mu) \ge \sum_{e\in E(\hat{\gamma})} m_e^{1/p}|e|$$

Generalization to Lagrangian on manifolds and relation with the Weak KAM Theory

Lagrangian-Hamiltonian duality $(x, v) \in \mathbb{T}\Omega$:

$$l(x, v) = \sup_{p \in T^*\Omega} \langle p, v \rangle - h(x, p)$$

Generalization to Lagrangian on manifolds and relation with the Weak KAM Theory

Lagrangian-Hamiltonian duality $(x, v) \in \mathbb{T}\Omega$:

$$l(x, v) = \sup_{p \in T^*\Omega} \langle p, v \rangle - h(x, p)$$

$$\sup_{\mu \in \mathcal{B}_1^+} \inf_{\phi \in C^1(\Omega)} \int_\Omega h(x, d\phi) d\mu = \underline{E} =: \inf_{\phi \in C^1(\Omega)} \sup_{x \in \Omega} h(x, d\phi)$$

Generalization to Lagrangian on manifolds and relation with the Weak KAM Theory

Lagrangian-Hamiltonian duality $(x, v) \in \mathbb{T}\Omega$:

$$l(x, v) = \sup_{p \in T^*\Omega} \langle p, v \rangle - h(x, p)$$

$$\sup_{\mu \in \mathcal{B}_1^+} \inf_{\phi \in C^1(\Omega)} \int_\Omega h(x, d\phi) d\mu = \underline{E} =: \inf_{\phi \in C^1(\Omega)} \sup_{x \in \Omega} h(x, d\phi)$$

Example $l(x, v) = |v|^2/2 - V(x)$, $h(x, p) = |p|^2/2 + V(x)$

$$\sup_{\mu \in \mathcal{B}_1^+} \inf_{\phi \in C^1(\Omega)} \int_\Omega \left(|\nabla \phi|^2/2 + V(x) \right) d\mu = \sup_{x \in \Omega} V(x) \ .$$

$$C_T(x,y) := \inf_{\vec{z}(0)=x, \vec{z}(T)=y} \int_0^T l\left(\vec{z}(s), \dot{\vec{z}}(s)\right) ds \quad , \quad T > 0 \; .$$

$$C_T(x,y) := \inf_{\vec{z}(0)=x, \vec{z}(T)=y} \int_0^T l\left(\vec{z}(s), \dot{\vec{z}}(s)\right) ds \quad , \quad T > 0 \ .$$

Then

$$C_T(\mu) := C_T(\mu,\mu) = \min_{\Lambda \in \mathcal{P}(\mu,\mu)} \int_{M\times M} C_T(x,y) d\Lambda(x,y)$$

$$C_T(x,y) := \inf_{\vec{z}(0)=x, \vec{z}(T)=y} \int_0^T l\left(\vec{z}(s), \dot{\vec{z}}(s)\right) ds \quad , \quad T > 0 \; .$$

Then

$$C_T(\mu) := C_T(\mu,\mu) = \min_{\Lambda \in \mathcal{P}(\mu,\mu)} \int_{M \times M} C_T(x,y) d\Lambda(x,y)$$

Theorem

(Buffoni and Bernard)

$$\min_{\mu \in \mathcal{B}_1^+} C_T(\mu) = -T\underline{E}$$

where the minimizers coincide, for any $T > 0$, with the projected Mather measure.

The action C_T induces a "metric" on Ω:

Definition

$$(x,t) \in \Omega \times \Omega \mapsto D_E(x,y) = \inf_{T>0} C_T(x,y) + TE \ .$$

The action C_T induces a "metric" on Ω:

Definition

$$(x,t) \in \Omega \times \Omega \mapsto D_E(x,y) = \inf_{T>0} C_T(x,y) + TE \ .$$

Lemma

$D_E(x,y) = -\infty$ *for any* $x, y \in \Omega$ *if* $E < \underline{E}$.

The action C_T induces a "metric" on Ω:

Definition

$$(x,t) \in \Omega \times \Omega \mapsto D_E(x,y) = \inf_{T>0} C_T(x,y) + TE \ .$$

Lemma

$D_E(x,y) = -\infty$ for any $x, y \in \Omega$ if $E < \underline{E}$. If $E \geq \underline{E}$ then $D_E(x,x) = 0$ for any $x \in \Omega$.

The action C_T induces a "metric" on Ω:

Definition

$$(x,t) \in \Omega \times \Omega \mapsto D_E(x,y) = \inf_{T>0} C_T(x,y) + TE \ .$$

Lemma

$D_E(x,y) = -\infty$ for any $x, y \in \Omega$ if $E < \underline{E}$. If $E \geq \underline{E}$ then $D_E(x,x) = 0$ for any $x \in \Omega$.

Example

For $l(x,v) = |v|^2/2$ we get $C_T(x,y) = |x-y|^2/2T$ while $D_E(x,y) = \sqrt{2E}|x-y|$ if $E \geq 0$, $D_E(x,y) = -\infty$ if $E < 0$. Here $\underline{E} = 0$.

The action C_T induces a "metric" on Ω:

Definition

$$(x,t)\in\Omega\times\Omega\mapsto D_E(x,y)=\inf_{T>0} C_T(x,y)+TE\ .$$

Lemma

$D_E(x,y)=-\infty$ for any $x,y\in\Omega$ if $E<\underline{E}$. If $E\geq\underline{E}$ then $D_E(x,x)=0$ for any $x\in\Omega$.

Example

For $l(x,v)=|v|^2/2$ we get $C_T(x,y)=|x-y|^2/2T$ while $D_E(x,y)=\sqrt{2E}|x-y|$ if $E\geq 0$, $D_E(x,y)=-\infty$ if $E<0$. Here $\underline{E}=0$.

Lemma

$$D_E(\lambda):=D_E(\lambda^+,\lambda^-)=\sup\left\{\int\phi d\lambda\ ;\ \ \phi(x)-\phi(y)\leq D_E(x,y)\right\}$$

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi\in C^1(\Omega)} \int_\Omega -h(x,d\phi)d\mu + \phi d\lambda$$

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi\in C^1(\Omega)} \int_\Omega -h(x,d\phi)d\mu + \phi d\lambda$$

$$\widehat{\mathcal{C}}(\lambda) := \inf_{\mu\in\mathcal{B}_1^+} \widehat{\mathcal{C}}(\lambda;\mu)$$

.

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi\in C^1(\Omega)} \int_\Omega -h(x,d\phi)d\mu + \phi d\lambda$$

$$\widehat{\mathcal{C}}(\lambda) := \inf_{\mu\in\mathcal{B}_1^+} \widehat{\mathcal{C}}(\lambda;\mu)$$

.

Theorem

If $\lambda \in \mathcal{B}_0$, the following definitions are equivalent:

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi \in C^1(\Omega)} \int_\Omega -h(x, d\phi) d\mu + \phi d\lambda$$

$$\widehat{\mathcal{C}}(\lambda) := \inf_{\mu \in \mathcal{B}_1^+} \widehat{\mathcal{C}}(\lambda;\mu)$$

.

Theorem

If $\lambda \in \mathcal{B}_0$*, the following definitions are equivalent:*

1. $\widehat{\mathcal{C}}_T(\lambda) := T\widehat{\mathcal{C}}\left(\frac{\lambda}{T}\right)$

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi\in C^1(\Omega)} \int_\Omega -h(x,d\phi)d\mu + \phi d\lambda$$

$$\widehat{\mathcal{C}}(\lambda) := \inf_{\mu\in\mathcal{B}_1^+} \widehat{\mathcal{C}}(\lambda;\mu)$$

.

Theorem

If $\lambda \in \mathcal{B}_0$, the following definitions are equivalent:

1. $\widehat{\mathcal{C}}_T(\lambda) := T\widehat{\mathcal{C}}\left(\frac{\lambda}{T}\right)$
2. $\widehat{\mathcal{C}}_T(\lambda) := \sup_{E\geq\underline{E}} D_E(\lambda) - ET$.

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi\in C^1(\Omega)} \int_\Omega -h(x,d\phi)d\mu + \phi d\lambda$$

$$\widehat{\mathcal{C}}(\lambda) := \inf_{\mu\in\mathcal{B}_1^+} \widehat{\mathcal{C}}(\lambda;\mu)$$

.

Theorem

If $\lambda \in \mathcal{B}_0$, the following definitions are equivalent:

1. $\widehat{\mathcal{C}}_T(\lambda) := T\widehat{\mathcal{C}}\left(\frac{\lambda}{T}\right)$
2. $\widehat{\mathcal{C}}_T(\lambda) := \sup_{E\geq\underline{E}} D_E(\lambda) - ET$.
3. $\widehat{\mathcal{C}}_T(\lambda) := \inf_{\mu\in\mathcal{B}_1^+} \sup_{\phi\in C^1(M)} \int_M -Th(x,d\phi)d\mu + \phi d\lambda$.

Definition

$$\widehat{\mathcal{C}}(\lambda;\mu) := \sup_{\phi\in C^1(\Omega)} \int_\Omega -h(x,d\phi)d\mu + \phi d\lambda$$

$$\widehat{\mathcal{C}}(\lambda) := \inf_{\mu\in\mathcal{B}_1^+} \widehat{\mathcal{C}}(\lambda;\mu)$$

.

Theorem

If $\lambda\in\mathcal{B}_0$, the following definitions are equivalent:

1. $\widehat{\mathcal{C}}_T(\lambda) := T\widehat{\mathcal{C}}\left(\frac{\lambda}{T}\right)$
2. $\widehat{\mathcal{C}}_T(\lambda) := \sup_{E\geq\underline{E}} D_E(\lambda) - ET$.
3. $\widehat{\mathcal{C}}_T(\lambda) := \inf_{\mu\in\mathcal{B}_1^+}\sup_{\phi\in C^1(M)}\int_M -Th(x,d\phi)d\mu + \phi d\lambda$.

In particular, for $\lambda = \delta_x - \delta_y$,

$$\widehat{\mathcal{C}}_T(x,y) := \sup_{E\geq\underline{E}} D_E(x,y) - ET$$

From definition

$$C_T(x,y) \geq \widehat{\mathcal{C}}_T(x,y)$$

From definition

$$C_T(x,y) \geq \widehat{\mathcal{C}}_T(x,y)$$

In general, strict inequality. However, if $T << 1$ we get equality under mild conditions.

From definition

$$C_T(x,y) \geq \widehat{\mathcal{C}}_T(x,y)$$

In general, strict inequality. However, if $T << 1$ we get equality under mild conditions.

Theorem

For any $\lambda \in \mathcal{B}_0$,

$$\widehat{\mathcal{C}}_T(\lambda;\mu) = \Gamma - \lim_{\varepsilon\to 0} \varepsilon^{-1} C_{\varepsilon T}(\mu + \varepsilon\lambda^-, \mu + \varepsilon\lambda^+) \ .$$

From definition

$$C_T(x,y) \geq \widehat{\mathcal{C}}_T(x,y)$$

In general, strict inequality. However, if $T << 1$ we get equality under mild conditions.

Theorem

For any $\lambda \in \mathcal{B}_0$,

$$\widehat{\mathcal{C}}_T(\lambda;\mu) = \Gamma - \lim_{\varepsilon\to 0} \varepsilon^{-1} C_{\varepsilon T}(\mu+\varepsilon\lambda^-, \mu+\varepsilon\lambda^+) \; .$$

$$\widehat{\mathcal{C}}_T(\lambda) = \lim_{\varepsilon\to 0} \inf_{\mu\in\mathcal{B}_1^+} \varepsilon^{-1} C_{\varepsilon T}(\mu+\varepsilon\lambda^-, \mu+\varepsilon\lambda^+) \; .$$

Sketch of proof

Sketch of proof

Easy Lemma

Sketch of proof

Easy Lemma

Lemma

For any $\mu \in \mathcal{B}_1^+$, $\lambda = \lambda^+ - \lambda^- \in \mathcal{B}_0$

$$\liminf_{\varepsilon \to 0} \varepsilon^{-1} C_{\varepsilon T}(\mu + \varepsilon\lambda^-, \mu + \varepsilon\lambda^+) \geq \widehat{\mathcal{C}}_T(\lambda \| \mu) \ .$$

Sketch of proof

Easy Lemma

Lemma

For any $\mu \in \mathcal{B}_1^+$, $\lambda = \lambda^+ - \lambda^- \in \mathcal{B}_0$

$$\liminf_{\varepsilon \to 0} \varepsilon^{-1} C_{\varepsilon T}(\mu + \varepsilon\lambda^-, \mu + \varepsilon\lambda^+) \geq \widehat{\mathcal{C}}_T(\lambda \| \mu) \ .$$

Harder lemma

Sketch of proof

Easy Lemma

Lemma

For any $\mu \in \mathcal{B}_1^+$, $\lambda = \lambda^+ - \lambda^- \in \mathcal{B}_0$

$$\liminf_{\varepsilon \to 0} \varepsilon^{-1} C_{\varepsilon T}(\mu + \varepsilon\lambda^-, \mu + \varepsilon\lambda^+) \geq \widehat{\mathcal{C}}_T(\lambda \| \mu) \ .$$

Harder lemma

Lemma

For $T > 0$,

$$\widehat{\mathcal{C}}_T(\lambda) \geq \limsup_{\varepsilon \to 0} \varepsilon^{-1} \inf_{\mu \in \mathcal{B}_1^+} C_{\varepsilon T}(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-) \ .$$

Proof of "hard" Lemma

Proof of "hard" Lemma

Given $\varepsilon > 0$ let

$$D_E^\varepsilon(x,y) := \inf_{n\in\mathbb{N}} \left[C_{\varepsilon n T}(x,y) + \varepsilon n E T\right] \ .$$

Evidently, $D_E^\varepsilon(x,y)$ is continuous on $M \times M$ locally uniformly in $E \geq \underline{E}$. Moreover,

$$\lim_{\varepsilon \searrow 0} D_E^\varepsilon = D_E$$

uniformly on $M \times M$ and locally uniformly in $E \geq \underline{E}$ as well.

Proof of "hard" Lemma

Given $\varepsilon > 0$ let

$$D_E^\varepsilon(x,y) := \inf_{n\in\mathbb{N}} \left[C_{\varepsilon n T}(x,y) + \varepsilon n E T\right] \ .$$

Evidently, $D_E^\varepsilon(x,y)$ is continuous on $M \times M$ locally uniformly in $E \geq \underline{E}$. Moreover,

$$\lim_{\varepsilon \searrow 0} D_E^\varepsilon = D_E$$

uniformly on $M \times M$ and locally uniformly in $E \geq \underline{E}$ as well.
We now decompose $M \times M$ into mutually disjoint Borel sets Q_n:

$$M \times M = \cup_n Q_n^\varepsilon \ , \quad Q_n^\varepsilon \cap Q_{E,n'}^\varepsilon = \emptyset \ \text{ if } n \neq n'$$

such that

$$Q_n^\varepsilon \subset \left\{(x,y) \in M \times M \ ; \quad D_E^\varepsilon(x,y) = C_{\varepsilon n T}(x,y) + \varepsilon n E T\right\} \ .$$

Let $\Lambda_\varepsilon^E \in \mathcal{P}(\lambda^+, \lambda^-)$ be an optimal plan for

$$\mathcal{D}_E^\varepsilon(\lambda) = \int_{M\times M} D_E^\varepsilon(x,y) d\Lambda_\varepsilon^E = \min_{\Lambda\in\mathcal{P}(\lambda^+,\lambda^-)} \int_{M\times M} D_E^\varepsilon(x,y) d\Lambda \ ,$$

and $\Lambda_\varepsilon^n = \Lambda_\varepsilon^E \lfloor_{Q_n^\varepsilon}$, the restriction of Λ_ε^E to Q_n^ε.

Let $\Lambda_\varepsilon^E \in \mathcal{P}(\lambda^+, \lambda^-)$ be an optimal plan for

$$\mathcal{D}_E^\varepsilon(\lambda) = \int_{M\times M} D_E^\varepsilon(x,y)d\Lambda_\varepsilon^E = \min_{\Lambda\in\mathcal{P}(\lambda^+,\lambda^-)} \int_{M\times M} D_E^\varepsilon(x,y)d\Lambda \ ,$$

and $\Lambda_\varepsilon^n = \Lambda_\varepsilon^E\lfloor_{Q_n^\varepsilon}$, the restriction of Λ_ε^E to Q_n^ε. Set $\lambda_n^\pm$ to be the marginals of Λ_ε^n on the first and second factors of $M \times M$.

Let $\Lambda_\varepsilon^E \in \mathcal{P}(\lambda^+, \lambda^-)$ be an optimal plan for

$$\mathcal{D}_E^\varepsilon(\lambda) = \int_{M\times M} D_E^\varepsilon(x,y) d\Lambda_\varepsilon^E = \min_{\Lambda\in\mathcal{P}(\lambda^+,\lambda^-)} \int_{M\times M} D_E^\varepsilon(x,y) d\Lambda \ ,$$

and $\Lambda_\varepsilon^n = \Lambda_\varepsilon^E \lfloor_{Q_n^\varepsilon}$, the restriction of Λ_ε^E to Q_n^ε. Set $\lambda_n^\pm$ to be the marginals of Λ_ε^n on the first and second factors of $M \times M$. Then $\sum_{n=1}^\infty \Lambda_\varepsilon^n = \Lambda_\varepsilon^E$ and

$$\sum_{n=1}^\infty \lambda_n^\pm = \lambda^\pm$$

Remark

Note that $Q_n^\varepsilon = \emptyset$ for all but a finite number of $n \in \mathbb{N}$. In particular, the sum contains only a finite number of non-zero terms.

Let $|\lambda_n| := \int_M d\lambda_n^{\pm} \equiv \int_{M\times M} d\Lambda_n^{\varepsilon}$. The *averaged flight time* is

$$\langle T\rangle^{\varepsilon} := \varepsilon T \sum_{n=1}^{\infty} n|\lambda_n|$$

Let $|\lambda_n| := \int_M d\lambda_n^{\pm} \equiv \int_{M\times M} d\Lambda_n^{\varepsilon}$. The *averaged flight time* is

$$\langle T\rangle^{\varepsilon} := \varepsilon T \sum_{n=1}^{\infty} n|\lambda_n|$$

We observe that $\langle T\rangle^{\varepsilon} \in \partial_E \mathcal{D}_E^{\varepsilon}(\lambda)$, where ∂_E is the super gradient as a function of E. At this stage we choose E depending on ε, T such that

$$\langle T\rangle^{\varepsilon} = T + 2\varepsilon T|\lambda^{\pm}|$$

Let $|\lambda_n| := \int_M d\lambda_n^{\pm} \equiv \int_{M\times M} d\Lambda_n^{\varepsilon}$. The *averaged flight time* is

$$\langle T\rangle^{\varepsilon} := \varepsilon T \sum_{n=1}^{\infty} n|\lambda_n|$$

We observe that $\langle T\rangle^{\varepsilon} \in \partial_E \mathcal{D}_E^{\varepsilon}(\lambda)$, where ∂_E is the super gradient as a function of E. At this stage we choose E depending on ε, T such that

$$\langle T\rangle^{\varepsilon} = T + 2\varepsilon T|\lambda^{\pm}|$$

Let $\widehat{\Lambda}_{\varepsilon}^{n} \in \mathcal{B}^{+}(TM)$ satisfying

$$\left(I \oplus Exp_{(I)}^{(t=\varepsilon nT)}\right)_{\#} \widehat{\Lambda}_{\varepsilon}^{n} = \Lambda_{\varepsilon}^{n}$$

Let $|\lambda_n| := \int_M d\lambda_n^\pm \equiv \int_{M\times M} d\Lambda_n^\varepsilon$. The *averaged flight time* is

$$\langle T\rangle^\varepsilon := \varepsilon T \sum_{n=1}^{\infty} n|\lambda_n|$$

We observe that $\langle T\rangle^\varepsilon \in \partial_E \mathcal{D}_E^\varepsilon(\lambda)$, where ∂_E is the super gradient as a function of E. At this stage we choose E depending on ε, T such that

$$\langle T\rangle^\varepsilon = T + 2\varepsilon T|\lambda^\pm|$$

Let $\widehat{\Lambda}_\varepsilon^n \in \mathcal{B}^+(TM)$ satisfying

$$\left(I \oplus Exp_{(I)}^{(t=\varepsilon n T)}\right)_\# \widehat{\Lambda}_\varepsilon^n = \Lambda_\varepsilon^n$$

Use $\widehat{\Lambda}_\varepsilon^n$ to define

$$\lambda_n^j := \left(Exp_{(I)}^{(t=\varepsilon n T)}\right)_\# \widehat{\Lambda}_\varepsilon^n \in \mathcal{B}^+(M)$$

for $j = 0, 1 \ldots n$.

Let $|\lambda_n| := \int_M d\lambda_n^\pm \equiv \int_{M\times M} d\Lambda_n^\varepsilon$. The *averaged flight time* is

$$\langle T\rangle^\varepsilon := \varepsilon T \sum_{n=1}^\infty n|\lambda_n|$$

We observe that $\langle T\rangle^\varepsilon \in \partial_E \mathcal{D}_E^\varepsilon(\lambda)$, where ∂_E is the super gradient as a function of E. At this stage we choose E depending on ε, T such that

$$\langle T\rangle^\varepsilon = T + 2\varepsilon T|\lambda^\pm|$$

Let $\widehat{\Lambda}_\varepsilon^n \in \mathcal{B}^+(TM)$ satisfying

$$\left(I \oplus Exp_{(I)}^{(t=\varepsilon n T)}\right)_\# \widehat{\Lambda}_\varepsilon^n = \Lambda_\varepsilon^n$$

Use $\widehat{\Lambda}_\varepsilon^n$ to define

$$\lambda_n^j := \left(Exp_{(I)}^{(t=\varepsilon n T)}\right)_\# \widehat{\Lambda}_\varepsilon^n \in \mathcal{B}^+(M)$$

for $j = 0, 1 \ldots n$. Note that $\lambda_n^0 = \lambda_n^+ \quad , \lambda_n^n = \lambda_n^-$

Let $|\lambda_n| := \int_M d\lambda_n^{\pm} \equiv \int_{M\times M} d\Lambda_n^{\varepsilon}$. The *averaged flight time* is

$$\langle T\rangle^{\varepsilon} := \varepsilon T \sum_{n=1}^{\infty} n|\lambda_n|$$

We observe that $\langle T\rangle^{\varepsilon} \in \partial_E \mathcal{D}_E^{\varepsilon}(\lambda)$, where ∂_E is the super gradient as a function of E. At this stage we choose E depending on ε, T such that

$$\langle T\rangle^{\varepsilon} = T + 2\varepsilon T|\lambda^{\pm}|$$

Let $\widehat{\Lambda}_{\varepsilon}^{n} \in \mathcal{B}^{+}(TM)$ satisfying

$$\left(I \oplus Exp_{(I)}^{(t=\varepsilon n T)}\right)_{\#} \widehat{\Lambda}_{\varepsilon}^{n} = \Lambda_{\varepsilon}^{n}$$

Use $\widehat{\Lambda}_{\varepsilon}^{n}$ to define

$$\lambda_n^j := \left(Exp_{(I)}^{(t=\varepsilon n T)}\right)_{\#} \widehat{\Lambda}_{\varepsilon}^{n} \in \mathcal{B}^{+}(M)$$

for $j = 0, 1 \ldots n$. Note that $\lambda_n^0 = \lambda_n^+ \quad , \lambda_n^n = \lambda_n^-$

$$C_{\varepsilon n T}(\lambda_n^+, \lambda_n^-) + \varepsilon n E T|\lambda_n| = \sum_{j=0}^{n-1} \left[C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1}) + \varepsilon E T|\lambda_n|\right]$$

$$\mathcal{D}_E^{\varepsilon}(\lambda) = \sum_{n=1}^{\infty} D_E^{\varepsilon}(\lambda_n) = \sum_{n=1}^{\infty} \left[C_{\varepsilon n T}(\lambda_n^+, \lambda_n^-) + \varepsilon n E T |\lambda_n| \right]$$

$$= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \left(C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1}) + \varepsilon E T |\lambda_n| \right) \ . \quad (2)$$

$$\mathcal{D}_E^\varepsilon(\lambda) = \sum_{n=1}^{\infty} D_E^\varepsilon(\lambda_n) = \sum_{n=1}^{\infty} \left[C_{\varepsilon n T}(\lambda_n^+, \lambda_n^-) + \varepsilon n E T |\lambda_n| \right]$$

$$= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \left(C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1}) + \varepsilon E T |\lambda_n| \right) \ . \quad (2)$$

Let now

$$\mu^{\varepsilon, E} = \varepsilon \sum_{n=1}^{\infty} \sum_{j=1}^{n-1} \lambda_n^j \ .$$

$$\mathcal{D}_E^{\varepsilon}(\lambda) = \sum_{n=1}^{\infty} D_E^{\varepsilon}(\lambda_n) = \sum_{n=1}^{\infty} \left[C_{\varepsilon n T}(\lambda_n^+, \lambda_n^-) + \varepsilon n E T |\lambda_n| \right]$$

$$= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \left(C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1}) + \varepsilon E T |\lambda_n| \right) \quad . \quad (2)$$

Let now

$$\mu^{\varepsilon, E} = \varepsilon \sum_{n=1}^{\infty} \sum_{j=1}^{n-1} \lambda_n^j \quad .$$

Note that

$$\mu^{\varepsilon, E} = \varepsilon \sum_{n=1}^{\infty} \sum_{j=0}^{n} \lambda_n^j - \varepsilon \sum_{n=1}^{\infty} \lambda_n^0 - \varepsilon \sum_{n=1}^{\infty} \lambda_n^n \; .$$

We obtain

$$\left|\mu^{\varepsilon,E}\right| = \varepsilon\sum_{n=1}^{\infty}(n+1)|\lambda_n^{\pm}| - 2\varepsilon|\lambda^{\pm}| = 1 \implies \mu^{\varepsilon,E} \in \mathcal{B}_1^{+} \ .$$

We obtain

$$\left|\mu^{\varepsilon,E}\right| = \varepsilon\sum_{n=1}^{\infty}(n+1)|\lambda_n^{\pm}| - 2\varepsilon|\lambda^{\pm}| = 1 \implies \mu^{\varepsilon,E} \in \mathcal{B}_1^{+} \ .$$

$$\sum_{n=1}^{\infty}\sum_{j=0}^{n-1} C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1})$$

We obtain

$$\left|\mu^{\varepsilon,E}\right| = \varepsilon\sum_{n=1}^{\infty}(n+1)|\lambda_n^{\pm}| - 2\varepsilon|\lambda^{\pm}| = 1 \implies \mu^{\varepsilon,E} \in \mathcal{B}_1^{+} \; .$$

$$\sum_{n=1}^{\infty}\sum_{j=0}^{n-1} C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1})$$

$$\geq C_{\varepsilon T}\left(\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\lambda_n^j, \sum_{n=1}^{\infty}\sum_{j=1}^{n}\lambda_n^{j+1}\right) =$$

We obtain

$$\left|\mu^{\varepsilon,E}\right| = \varepsilon\sum_{n=1}^{\infty}(n+1)|\lambda_n^{\pm}| - 2\varepsilon|\lambda^{\pm}| = 1 \implies \mu^{\varepsilon,E} \in \mathcal{B}_1^+ \ .$$

$$\sum_{n=1}^{\infty}\sum_{j=0}^{n-1} C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1})$$

$$\geq C_{\varepsilon T}\left(\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\lambda_n^j, \sum_{n=1}^{\infty}\sum_{j=1}^{n}\lambda_n^{j+1}\right) =$$

$$\varepsilon^{-1}C_{\varepsilon T}\left(\varepsilon\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\lambda_n^j, \varepsilon\sum_{n=1}^{\infty}\sum_{j=1}^{n}\lambda_n^{j+1}\right)$$

We obtain

$$\left|\mu^{\varepsilon,E}\right| = \varepsilon\sum_{n=1}^{\infty}(n+1)|\lambda_n^{\pm}| - 2\varepsilon|\lambda^{\pm}| = 1 \implies \mu^{\varepsilon,E} \in \mathcal{B}_1^+ \ .$$

$$\sum_{n=1}^{\infty}\sum_{j=0}^{n-1} C_{\varepsilon T}(\lambda_n^j, \lambda_n^{j+1})$$

$$\geq C_{\varepsilon T}\left(\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\lambda_n^j, \sum_{n=1}^{\infty}\sum_{j=1}^{n}\lambda_n^{j+1}\right) =$$

$$\varepsilon^{-1}C_{\varepsilon T}\left(\varepsilon\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\lambda_n^j, \varepsilon\sum_{n=1}^{\infty}\sum_{j=1}^{n}\lambda_n^{j+1}\right)$$

$$= \varepsilon^{-1}C_{\varepsilon T}\left(\mu^{\varepsilon,E} + \varepsilon\lambda^+, \mu^{\varepsilon,E} + \varepsilon\lambda^-\right) \ .$$

$$
\begin{aligned}
\mathcal{D}^{\varepsilon}_{E}(\lambda) - \langle T \rangle^{\varepsilon} E &\geq \varepsilon^{-1} C_{\varepsilon T}\left(\mu^{\varepsilon,E} + \varepsilon\lambda^{+}, \mu^{\varepsilon,E} + \varepsilon\lambda^{-}\right) \\
&\geq \varepsilon^{-1} \inf_{\mu \in \mathcal{B}^{+}_{1}} C_{\varepsilon T}\left(\mu + \varepsilon\lambda^{+}, \mu + \varepsilon\lambda^{-}\right) \ . \qquad (3)
\end{aligned}
$$

$$\mathcal{D}_E^{\varepsilon}(\lambda) - \langle T\rangle^{\varepsilon} E \geq \varepsilon^{-1} C_{\varepsilon T}\left(\mu^{\varepsilon,E} + \varepsilon\lambda^{+}, \mu^{\varepsilon,E} + \varepsilon\lambda^{-}\right)$$
$$\geq \varepsilon^{-1} \inf_{\mu\in\mathcal{B}_1^{+}} C_{\varepsilon T}\left(\mu + \varepsilon\lambda^{+}, \mu + \varepsilon\lambda^{-}\right) . \quad (3)$$

Finally,

$$\widehat{\mathcal{C}}_T(\lambda) \geq \mathcal{D}_E(\lambda) - TE =$$
$$\lim_{\varepsilon\to 0} \mathcal{D}_E^{\varepsilon}(\lambda) - \langle T\rangle^{\varepsilon} E \geq \limsup_{\varepsilon\to 0} \varepsilon^{-1} \inf_{\mu\in\mathcal{B}_1^{+}} C_{\varepsilon T}\left(\mu + \varepsilon\lambda^{+}, \mu + \varepsilon\lambda^{-}\right) . \quad (4)$$