Optimal Transport Problems For Contact Structures

Fraydoun Rezakhanlou*
UC Berkeley
Department of Mathematics
Berkeley, CA 94720-3840

June 13, 2015

Abstract

In the Monge-Kantorovich transportation problem, we search for a plan that minimizes the cost of transporting mass from a set of locations to another set of locations. According to a result of Moser, two volume forms on a compact Manifold of the same orientation and total volume are isomorphic and a solution to Monge-Kantorovich problem offers a special isomorphism for Moser's result. A celebrated result of Gray asserts that if instead of volume forms we take two contact structures on a compact manifold, then they are isomorphic provided that they can be connected by a smooth arc of contact structures. In this article, we formulate and study optimal transport problems for Gray's result. Our formulations are, in spirit similar to Benamou-Brenier's formulation of Monge-Kantorovich problem.

1 Introduction

In the Monge-Kantorovich transportation problem, we search for a plan that minimizes the cost of transporting mass from a set of locations to another set of locations. More precisely, we fix a measure space M and a measurable cost function $c : M \times M \to [0, \infty)$, and define

$$D(\mu^0, \mu^1) = \inf \left\{ \int c(x, T(x))\mu^0(dx) : \ T : M \to M \text{ measurable and } T^*\mu^1 = \mu^0 \right\}, \tag{1.1}$$

for any pair of measures μ^0 and μ^1 on M. By $T^*\mu^1 = \mu^0$ (or equivalently $\mu^1 = T^\sharp\mu^0$) we mean that for every bounded measurable $f : M \to \mathbb{R}$,

$$\int f \ d\mu^1 = \int f \circ T \ d(T^*\mu^1) = \int f \circ T \ d\mu^0. \tag{1.2}$$

*This work is supported in part by NSF Grant DMS-1407723.

The classical Monge problem corresponds to the case $M=\mathbb{R}^d$ and $c(x,y)=|x-y|$. Various choices of the transport costs lead to various metrics on the space of probability measures. For example, if M is a metric space with metric $d(\cdot,\cdot)$, and (μ^0,μ^1) is a pair of Borel probability measures, then $D_1=D$ defines Kantorovich-Rubinstein distance for the choice $c=d$. On the other hand, the choice of $c=d^2$ leads to the Wasserstein distance $D_2=\sqrt{D}$. Kantorovich formulated a relaxed variant of the original Monge problem to allow multiple choices for $T(x)$. Kantorovich's formulation and duality lead to alternative ways of calculating D. On the other hand Benamou-Brenier formulation is closely related to a classical problem of Moser for volume forms on a compact manifolds. More precisely, let us define

$$\bar{D}(\mu^0,\mu^1)=\inf\left\{\int_0^1\int\frac{1}{2}|Z(x,t)|^2\rho(x,t)\ dxdt:(\rho,Z)\in A(\rho^0,\ \rho^1)\right\}, \tag{1.3}$$

where

$$A(\rho^0,\ \rho^1)=\left\{(\rho,Z):\rho_t+\nabla\cdot(\rho Z)=0,\ \text{and}\ \rho(\cdot,0)=\rho^0,\ \rho(\cdot,1)=\rho^1\right\}. \tag{1.4}$$

As was shown by Benamou and Brenier (see [V1]), we have $\bar{D}(\mu^0,\mu^1)=D(\mu^0,\mu^1)$ provided that $M=\mathbb{R}^d$, $c(x,y)=\frac{1}{2}|x-y|^2$, and $\mu^1=\rho^1dx$ and $\mu^0=\rho^0dx$ are absolutely continuous with respect to Lebesgue measure.

The main objective of this article is to formulate two transportation problems for contact forms that have similar flavor as Monge-Kantorovich problem. Before embarking on this, let us explain how a minimizer to (1.3) offers a special solution to a classical problem in differential geometry. In a classical problem of Moser, we take two volume forms μ^0 and μ^1 of the same orientation on a compact manifold M with $\int_M\mu^0=\int_M\mu^1$, and search for a diffeomorphism T such that $T^*\mu^1=\mu^0$. According to a result of Moser, such a diffeomorphism T always exists. Moser's method of proof is known as the deformation Trick and can be used to give a straightforward arguments for the local equivalence of symplectic structures (Darboux Theorem) and global equivalence of contact structures in the same "homotopy class" (Gray Stability Theorem). In fact Moser's solution for the existence of the diffeomorphism T is in the same vain as Benamou-Brenier's formulation (1.3) and relies on constructing a vector field Z such that its flow at time 1 is exactly the diffeomorphism T we are looking for. In Moser's approach we simply search for a vector field Z such that its flow ϕ_t connects μ^0 to μ^1. The path $\mu(\cdot,t):=\phi^*_{-t}\mu^0=\mu^t$ solves Lie's equation

$$\mu_t+\mathcal{L}_Z\mu=0, \tag{1.5}$$

where the operator $\mathcal{L}_Z$ is the Lie derivative and, in Euclidean setting, the equation (1.5) becomes Liouville's equation that appeared in (1.4). With this new notation, we may rewrite $\bar{D}(\mu^0,\mu^1)$ as

$$\inf\left\{\int_0^1\int_M\frac{1}{2}|Z(\cdot,t)|^2\ \mu(t)\ dt:\ \mu_t+\mathcal{L}_Z\mu=0,\ \text{and}\ \mu(0)=\mu^0,\ \mu(1)=\mu^1\right\}. \tag{1.6}$$

We use this formula as a guidance to formulate other optimization problems involving a pair of contact forms α^0 and α^1. Especially, our optimization problem will offer a special solution to Gray's problem in just the same way that Monge-Kantorovich minimizing T gives a solution to Moser's problem. In Gray's problem we search for a diffeomorphism ϕ such that $\phi^*\alpha^1 = f\alpha^0$ for a scalar-valued function $f > 0$. Such ϕ exists if α^1 can be connected to α^0 by a path of contact forms. See [G] for details.

To prepare for our optimization problems, let us assume that $n \geq 2$ and that M is a $d = 2n - 1$-dimensional Riemannian manifold with a metric g. For every $x \in M$, write $\|\cdot\|_x$ for the length associated with this metric at the point x. Recall that a 1-form α is contact if and only if

$$m_\alpha := \alpha \wedge (d\alpha)^{n-1},$$

is a volume form. The space of all contact forms is denoted by $\mathcal{C} = \mathcal{C}(M)$. We will define two optimization problems for contact forms that can be used to formulate two metrics S and $\hat{S}$ on $\mathcal{C}$.

Given a continuously differentiable paths $\alpha : [0,1] \to \mathcal{C}$, define

$$\mathcal{I}(\alpha(\cdot)) = \inf\left\{\frac{1}{2}\int_0^1\int_M \left(\|Z(x,t)\|_x^2 + V(x,t)^2\right) m_{\alpha(t)}dt : (Z,V) \in A(\alpha(\cdot))\right\}. \tag{1.7}$$

where $A(\alpha(\cdot))$ consists of those (Z,V) such that $\alpha_t + \mathcal{L}_Z\alpha = V\alpha$, in the interval $(0,1)$. As for the analog of $\bar{D}$ for contact forms, we define

$$S(\alpha^0, \alpha^1) = \inf\{\mathcal{I}(\alpha(\cdot)) : \alpha(0) = \alpha^0,\ \alpha(1) = \alpha^1 \quad \text{for } C^1 \text{ paths } \alpha(\cdot)\}, \tag{1.8}$$

for any pair of α^0 and α^1 in $\mathcal{C}$.

From our first set of results (Theorems 2.4 and 2.5 in Section 2) we learn:

- (i) $\mathcal{I}$ can be expressed as an action $\mathcal{I}(\alpha(\cdot)) = \int_0^1 \mathcal{G}(\alpha(t), \dot{\alpha}(t))dt$, where the "Lagrangian" $\mathcal{G}$ is convex in the second argument and can be expressed as an infimum over scalar-valued functions H. The function H plays the role of a Hamiltonian function and minimizing H satisfies an elliptic PDE. This allows us to simplify (1.7) and find a more natural expression for it.

- (ii) Legendre transform $\mathcal{H}(\alpha, Y)$ (defined for vector fields Y) of $\mathcal{G}(\alpha, \nu)$ in the ν-variable can be calculated explicitly. This allows us to express $\mathcal{G}$ as a supremum that is interpreted as the dual variational problem associated with $\mathcal{I}$.

Our optimal transport formulation raises the natural question of what the completion of the space of contact forms with respect to the metric $S(\alpha^0, \alpha^1)$ is. For a comparison, let us recall that if we consider Wasserstein metric $W(\mu^0, \mu^1) = (D(\mu^0, \mu^1))^{1/2}$ and complete the space of volume forms with respect to this metric, then we get the space of probability measures because this metric induces the topology of weak convergence. Completion of the space of contact forms with respect the metric $S(\alpha^0,\alpha^1)$ may yield an important insight into various rigidity questions in Symplectic and Contact Geometry.

There is an alternative choice for the definition of a metric on the space of contact forms that is more natural but requires more preparation to describe. For this alternative metric $\hat{S}$, we vary the Riemannian metric with time so that the selected metric at time t is compatible with the contact form $\alpha(t)$. The compatibility is expressed in terms an important notion in Symplectic/Contact Geometry known as *almost complex structure*. As it turns out, in spite of more complicated transportation problem we are adopting, the expression we get for the Lagrangian $\mathcal{G}$ and its Legendre transform are mathematically more tractable!

The organization of the paper is as follows:

- In Section 2 we give an overview of Monge-Kantorovich transport problem, formulate two optimal transport problems for contact forms and state the main results.
- In Section 3 we prove our main results for the metric S of (1.7).
- In Section 4 we prove similar results for the metric $\hat{S}$.

2 Main Results and Survey

This section is divided into four parts.

1. In the first part, Subsection 2.1, we give an overview of Monge-Kantorovich transportation problem. This Subsection will be used to motivate some of the definitions and results we discuss in Subsections 2.2 and 2.4.
2. In the second part, we give our first formulation of an optimal transport problem for contact forms and state two theorems for it.
3. Inspired by our optimization formulation in Subsection 2.2, we define a Riemannian metric on the space of contact forms in Subsection 2.3.
4. The fourth part is devoted to our second metric $\hat{S}$.

2.1 Monge-Kantorovich Problem

The expressions D and $\bar{D}$ of (1.1) and (1.3) offer two ways of measuring some kind of distance between probability measures. Kantorovich formulated a relaxed variant of the original Monge problem to allow multiple choices for $T(x)$. More precisely, Kantorovich's transport problem is defined by

$$\hat{D}(\mu^0,\mu^1) = \inf\left\{\int c(x,y)\mu(dx,dy) : \pi_x^\sharp\mu = \mu^0 \text{ and } \pi_y^\sharp\mu = \mu^1\right\}, \tag{2.1}$$

where $\pi_x(x,y) = x,\ \pi_y(x,y) = y$ so that $\pi_x^\sharp\mu = \mu^0$ and $\pi_y^\sharp\mu = \mu^1$ are the marginals of μ. In other words we require that the measure μ to be a *coupling* of the measures μ^0 and μ^1. For many cases of interest, $\hat{D} = D$. We can further relax Kantorovich's optimization problem by combining it with Benamou-Brenier's problem. To this end, let us write $C^1 = C^1([0,1];\mathbb{R}^d)$ for the space of continuously differential functions $x : [0,1] \to \mathbb{R}^d$ with its standard metric. Let $\mathcal{M}$ denote the space of nonnegative Borel measures on C^1. We also write $\pi_t(x(\cdot)) = x(t)$ for the evaluation at time t. We now define

$$\tilde{D}(\mu^0,\mu^1) = \inf_{P\in\mathcal{M}}\left\{\int\left[\int_0^1 \frac{1}{2}\left|\dot{x}(t)\right|^2\ dt\right]\ P(dx(\cdot)) : \pi_0^\sharp P = \mu^0 \text{ and } \pi_1^\sharp P = \mu^1\right\}, \tag{2.2}$$

Recall that D, $\bar{D}$, $\hat{D}$ and $\tilde{D}$ are defined by (1.1), (1.3), (2.1), and (2.2) respectively.

Theorem 2.1 **(i)** *Assume that $M = \mathbb{R}^d$ is equipped with its Borel σ-algebra and $c(x,y) = \frac{1}{2}|x-y|^2$. If $\mu^1 = \rho^1 dx$ and $\mu^0 = \rho^0 dx$ are absolutely continuous with respect to the Lebesgue measure and*

$$\int |x|^2\mu^0(dx) + \int |x|^2\mu^1(dx) < \infty, \tag{2.3}$$

then $D(\mu^0,\mu^1) = \hat{D}(\mu^0,\mu^1)$.
(ii) *Assume also that ρ^0 and ρ^1 in part* **(i)** *are of compact support. Then*

$$D(\mu^0,\mu^1) = \bar{D}(\mu^0,\mu^1) = \tilde{D}(\mu^0,\mu^1).$$

Remark 2.1

(i) We refer to [V1] for the proof of $D = \bar{D}$ and $D = \hat{D}$. The equality $D = \bar{D}$ is due to Benamou-Brenier [BB].
(ii) To explain $D = \bar{D}$ heuristically, write $\{\phi_t : t \in \mathbb{R}\}$ for the flow of the ODE

$$\frac{dx}{dt} = Z(x,t); \quad x(0) = a. \tag{2.4}$$

In other words, $x(t) = \phi_t(a)$. Now if $\mu^t = \phi_{-t}^*\mu^0$, then $\mu^t(dx) = \rho(x,t)dx$ satisfies Liouville's equation (1.5). In Benamou-Brenier's problem we try to minimize the averaged kinetic energy

$$\int_0^1\int \frac{1}{2}\left|\frac{d\phi_t(a)}{dt}\right|^2 \rho(a,0)\ dadt = \int_0^1\int \frac{1}{2}|Z(x,t)|^2\rho(x,t)\ dxdt.$$

What Theorem 2.1**(i)** asserts is that the minimizing velocity field $\bar{Z}$ has a t-linear flow of the form

$$\phi_t(x) = tT(x) + (1-t)x, \tag{2.5}$$

for a suitable homeomorphism T. Indeed, $\bar{Z} = (T - id) \circ \phi_{-t}$ for such a flow, and the averaged kinetic energy simplifies to

$$\int_0^1 \int \frac{1}{2}|\bar{Z}|^2 \ \rho \ dxdt = \int_0^1 \int \frac{1}{2}|\bar{Z}|^2 \ \phi_{-t}^* \mu^0(dx)dt = \int_0^1 \int \frac{1}{2}|\bar{Z} \circ \phi_t|^2 \ \rho^0 \ dxdt$$
$$= \int_0^1 \int \frac{1}{2}|x - T(x)|^2 \ \rho^0 \ dxdt.$$

The last expression on the right-hand side is exactly what we minimize in (1.1).

(iii) Basically what the equality $D = \hat{D}$ asserts is that the minimizing measure $\bar{\mu}$ of the Kantorovich's optimization problem lies on a graph of a function T and takes the form

$$\bar{\mu}(dx, dy) = \delta_{T(x)}(dy)\mu_0(dx).$$

(iv) In Moser's approach we simply search for a vector field Z such that its flow ϕ_t connects μ^0 to μ^1 by a line segment; $\phi_{-t}^* \mu^0 = \mu(t) = t\mu^1 + (1-t)\mu^0$. We note that Moser's solution is different from Benamou-Brenier's solution because in the latter, we have $\phi_t(x) = tT(x) + (1-t)x$ instead.

(v) What the equality $D = \tilde{D}$ really says is that the minimizing measure $\bar{P}$ in the variational problem (2.2) comes from solutions of an ODE of the form (2.4). More precisely, for any bounded continuous function $F : C^1 \to \mathbb{R}$, we have

$$\int F \ d\bar{P} = \int F(\phi_\cdot(a))\mu^0(da).$$

□

One of the main advantage of the relaxed variational problem (2.1) is that it has a nice dual formulation. This method of dualization is a cornerstone of Kantorovich's formulation and is a basic trick in many optimization problems. To this end, let us assume that M is a Polish (complete separable metric) space.The dual problem for (2.1) takes the form

$$\hat{D}^*(\mu^0, \mu^1) = \sup_{J^1, J^0 \in C_b(M)} \left\{ \int J^1 \ d\mu^1 - \int J^0 \ d\mu^0 : J^1(y) - J^0(x) \le c(x, y) \right\}, \tag{2.6}$$

where $C_b(M)$ denotes the space of bounded continuous functions $J : M \to \mathbb{R}$. Similarly the dual problem for (1.3) is given by

$$\bar{D}^*(\mu^0, \mu^1) = \sup \left\{ \int J(\cdot, 1) \ d\mu^1 - \int J(\cdot, 0) \ d\mu^0 : \ J_t + \frac{1}{2}|J_x|^2 = 0 \right\}. \tag{2.7}$$

Here the supremum is over continuous functions $J : \mathbb{R}^d \times [0,\infty) \to \mathbb{R}$ that satisfy the Hamilton-Jacobi equation

$$J_t + \frac{1}{2}|J_x|^2 = 0, \tag{2.8}$$

in the viscosity sense. The following result is well-known and can be found in [V1]:

Theorem 2.2 **(i)** *If the cost function c is lower semi-continuous, then $\hat{D} = \hat{D}^*$.*

(ii) *Under the assumptions of Theorem 2.1* **(ii)**, *we have $D = \hat{D} = \hat{D}^* = \bar{D}^*$.*

Remark 2.2

(i) To explain the equality $\hat{D}^* = \bar{D}^*$, recall that by Hopf-Lax-Oleinik Formula, the solution J of the equation (2.8) is given by

$$J(x,t) = \inf_y \left\{ J(y,0) + \frac{|x-y|^2}{2t} \right\}. \tag{2.9}$$

Hence,

$$J(x,1) - J(y,0) \le \frac{1}{2}|x-y|^2.$$

In other words, we use the the Hamilton-Jacobi equation (2.8) to interpolate between J^1 and J^0 of variational problem (2.6).

(ii) To explain the equality $\bar{D} = \bar{D}^*$, observe that if J satisfies (2.8) and ρ satisfies (1.4) for the choice of $Z = J_x$, then

$$\begin{aligned}\int J(\cdot,1)\, d\mu^1 - \int J(\cdot,0)\, d\mu^0 &= \int_0^1 \int (J\rho)_t \, dxdt = \int_0^1 \int (J_t\rho + J\rho_t)\, dxdt \\ &= -\int_0^1 \int \left[\frac{1}{2}|J_x|^2\rho + J\nabla\cdot(\rho J_x)\right] dxdt \\ &= \frac{1}{2}\int_0^1 \int |J_x|^2 \rho \, dxdt = \frac{1}{2}\int_0^1 \int |Z|^2\rho\, dxdt,\end{aligned}$$

where we integrated by parts to arrive at the fourth equality.

(iii) As an immediate consequence of Part **(ii)** of this remark and Remark 2.1 **(ii)** we learn

$$J_x(x,0) = Z(x,0) = T(x) - x, \quad \text{or} \quad T(x) = \nabla\left(J(x,0) + \frac{1}{2}|x|^2\right),$$

provided that $J(\cdot,0)$ is differentiable.

(iv) What we are optimizing in (2.2) is the average of the action with respect to a path measure P. More generally, we may start with a Lagrangian function $L(x, v)$ that is convex in v and define

$$\tilde{D}(\mu^0, \mu^1) = \inf_{P \in \mathcal{M}} \left\{ \int \left[\int_0^1 L(x(t), \dot{x}(t))\ dt \right]\ P(dx(\cdot)) : \pi_0^\sharp P = \mu^0 \text{ and } \pi_1^\sharp P = \mu^1 \right\},$$

This has a dual formulation as

$$\bar{D}^*(\mu^0, \mu^1) = \sup \left\{ \int J(1, \cdot)\ d\mu^1 - \int J(0, \cdot)\ d\mu^0 :\ J_t + H(x, J_x) = 0 \right\}, \tag{2.10}$$

where H is the Legendre Transform of L in v variable:

$$H(x, p) = \sup(v \cdot p - L(x, v)).$$

(See [V2] for more details.) When L depends on x, there is no analog of D or $\hat{D}$ formulation. This is because the minimizing P comes from a vector field $Z(x, t) = H_p(x, J_x(x, t))$, with J a maximizer of (2.10); Z-flows are simply the characteristic curves of the corresponding Hamilton-Jacobi equation and these curves are no longer straight lines. The point is that if the flow is linear as in (2.5), then it is uniquely determined by its end points. This is no longer the case if L is not independent of x. Nonetheless, we may define

$$c(x, y) = \inf \left\{ \int \int_0^1 L(x(t), \dot{x}(t))\ dt\ :\ x(0) = x,\ x(1) = y \right\},$$

and express $\tilde{D}$ as

$$\tilde{D}(\mu^0, \mu^1) = \inf_{P \in \mathcal{M}} \left\{ \int c(x, y)\ \mu(dx, dy)\ :\ \pi_x^\sharp \mu = \mu^0 \text{ and } \pi_y^\sharp \mu = \mu^1 \right\}.$$

$\square$

We end this subsection with a variant of Monge-Kantorovich-Moser's problem that is not conservative and is closely related to the metric S that will be defined in Subsection 2.2 below. By "nonconservative" we mean that the measures μ^0 and μ^1 do not have to be of the same total mass in order to have a finite "distance". This time we define

$$\bar{E}(\mu^0, \mu^1) = \inf \left\{ \int_0^1 \int \frac{1}{2} \left(|Z|^2 + V^2 \right) \rho\ dxdt : (\rho, Z, V) \in B(\rho^0,\ \rho^1) \right\}, \tag{2.11}$$

where

$$B(\rho^0,\ \rho^1) = \left\{ (\rho, Z, V) : \rho_t + \nabla \cdot (\rho Z) = nV\rho, \text{ and } \rho(\cdot, 0) = \rho^0,\ \rho(\cdot, 1) = \rho^1 \right\}. \tag{2.12}$$

In view of (1.5), the equation satisfied by the volume form $\mu(t) = \rho(x,t)dx$ can be written as

$$\mu_t + \mathcal{L}_Z \mu = nV\mu, \tag{2.13}$$

As for the dual problem,

$$\bar{E}^*(\mu^0, \mu^1) = \sup \left\{ \int J(1,\cdot)\ d\mu^1 - \int J(0,\cdot)\ d\mu^0 :\ J_t + \frac{1}{2}|J_x|^2 + \frac{n^2}{2}J^2 = 0 \right\}. \tag{2.14}$$

The analog of Theorem 2.1**(ii)** for this variant of $\bar{D}$ is Theorem 2.3.

Theorem 2.3 *Under the assumptions of Theorem 2.1*(**ii**)*, we have $\bar{E} = \bar{E}^*$.*

Remark 2.3 As we will see in Section 3, the relationship between the minimizing (Z, V) in (2.13) and the maximizing J in (2.14) is

$$Z = J_x, \quad V = nJ. \tag{2.15}$$

□

2.2 The Metric S

The primary purpose of this paper is to formulate a transportation problem for contact forms that has a similar flavor as Monge-Kantorovich problem. As we mentioned before, we may use the representation (1.6) as a guideline to define some kind of "distance" between k-forms. Let us recall that if we consider Wasserstein metric $W(\mu^0, \mu^1) = (D(\mu^0, \mu^1))^{1/2}$ and complete the space of volume forms of a given orientation with respect to this metric, then we get the space of measures because this metric induces the topology of weak convergence. Note that when we talk about volume forms of the same orientation (and total volumes), we are tacitly assuming a non-degeneracy. In the Euclidean setting, this means that for example, our volume forms are of the form $\mu^1 = \rho^1 dx$ and $\mu^0 = \rho^0 dx$ with both ρ^0 and ρ^1 positive. In fact the argument of Moser relies on the fact that if we take two (non-degenerate) volume forms μ^1 and μ^0 of the same total volume and orientation, then there is a path of (non-degenerate) forms of the same orientation that would connect μ^1 to μ^0, namely $\mu(t) = t\mu^1 + (1-t)\mu^0$. Now if we write $\mathcal{V}$ for the space of all volume forms of a given orientation and total volume 1, then this space is convex and its completion with respect to the Wasserstein distance is simply the space of probability measures. There is a subtlety in defining such distances for forms because a priori we do not have a candidate for the completion of our space with respect to the metric we wish to define. Also, we can define a distance between two forms only if there is a path of non-degenerate forms connecting them. What non-degeneracy we have in mind for forms? We choose a non-degeneracy so that reasonably nice paths exist to connect them! In view of the formula (1.5), we would like to have a path that comes from a vector field Z. Let us explain this by recalling three classical results in Symplectic/Contact Geometry.

1. In the case of 2-forms, we restrict ourselves to symplectic (non-degenerate closed) forms. By the non-degeneracy of a 2-form ω we mean that if $\omega_x(v_1, v_2) = 0$ for every v_2, then $v_1 = 0$. Given a point x, and symplectic forms ω^0 and ω^1, we can always find a path of sympletic forms connecting them provided that we stay close to x. In fact $\omega(t) = t\omega^1 + (1-t)\omega^0$ does the job. Sufficiently close to x, we can even guarantee that $\omega(t)$ is exact and that $\omega_t = \mathcal{L}_Z\omega$, for some vector field Z. The flow ϕ of Z produces an isomorphism between ω_0 and ω_1; in fact $\phi_1^*\omega^0 = \omega^1$. In summary, we have shown Darboux' theorem with the aid of the *Moser's deformation trick.*

2. In the case of 1-forms, recall that we write $\mathcal{C}$ for the space of contact forms. If $(\alpha(t) : t \in [0, 1])$ is a smooth path in $\mathcal{C}$, then there exists a vector field Z and a (potential) function V such that $\alpha_t + \mathcal{L}_Z\alpha = V\alpha$. This fact allows us to apply Moser's deformation trick to deduce the celebrated Stability Theorem of Gray: If M is compact and α^1 and α^0 are in the same (pathwise) connected component of $\mathcal{C}$, then α^1 and α^0 induce isomorphic contact structures. That is, there exists a diffeomorphism T such that $T^*\eta^1 = \eta^0$ where $\eta_x^i = \{v : \alpha_x^i(v) = 0\}$.

3. By a deep result of Eliashberg, if the dimension is 3, then any two *overtwisted* contact forms in the same homotopy class can be connected by a smooth path of contact forms. (We refer to [G] for the definition of overtwisted forms and the proof of Eliashberg's theorem.) In other words, if a smooth path of non-degenerate forms connects two overtwisted contact forms, then this path can be smoothly deformed to a path of contact forms.

We already defined contact forms in Section 1. Here we give an equivalent definition (see [R1] for example) that is more practical and natural.

Definition 2.1. A contact form α on an odd dimensional manifold M is a 1-form that satisfies the following two conditions:

(i) The set $l_x = \{v : d\alpha_x(v, w) = 0 \text{ for all } w \in T_xM\}$ is a line.

(ii) $\alpha_x(v) \neq 0$ for every nonzero $v \in l_x$.

Equivalently, a 1-form α is contact, if and only if $m_\alpha = \alpha \wedge (d\alpha)^{n-1}$, is a volume form. □

Contact forms are non-degenerate in a rather strong sense. To explain this, let us also define the kernel of α

$$\eta_x^\alpha = \eta_x = \{v : \alpha_x(v) = 0\}, \tag{2.16}$$

and observe that the contact condition really means that l_x and η_x give a decomposition of T_xM that depends solely on α:

$$T_xM = \eta_x \oplus l_x. \tag{2.17}$$

We also define the Reeb vector field $R(x) = R^\alpha(x)$ to be the unique vector such that

$$R(x) \in l_x, \quad \alpha_x(R(x)) = 1. \tag{2.18}$$

Hence a contact form always induces an orientation on M. To this end, let us fix a volume form $\bar{m}$ and define

$$\begin{aligned} \mathcal{V} &= \mathcal{V}(\bar{m}) = \{\mu : \mu \text{ has the same orientation as } \bar{m}\}, \\ \mathcal{C} &= \mathcal{C}(\bar{m}) = \{\alpha : \alpha \text{ is contact and } m_\alpha \in \mathcal{V}(\bar{m})\}. \end{aligned} \tag{2.19}$$

For our purposes, we also need to assume that M is a Riemannian manifold equipped with a metric g. We also write $\|v\|_x = \sqrt{g_x(v,v)}$.

Now given $p \in [1,\infty)$ and two C^1 paths $\alpha : [0,1] \to \mathcal{C}$ and $\mu : [0,1] \to \mathcal{V}$, define

$$\begin{aligned} \mathcal{I}_p(\alpha(\cdot),\mu(\cdot)) &: = \inf\left\{p^{-1}\int_0^1\int_M \left(\|Z(x,t)\|_x^p + V(x,t)^p\right)\mu(t)dt : (Z,V) \in A(\alpha(\cdot)) \cap B(\mu(\cdot))\right\}, \\ \mathcal{I}_p(\alpha(\cdot)) &: = \mathcal{I}_p(\alpha(\cdot), m_{\alpha(\cdot)}), \end{aligned}$$

where $A(\alpha(\cdot))$ consists of those (Z,V) such that

$$\alpha_t + \mathcal{L}_Z\alpha = V\alpha, \tag{2.20}$$

for t in the interval $(0,1)$, and $B(\mu(\cdot))$ consists of those (Z,V) such that

$$\mu_t + \mathcal{L}_Z\mu = nV\mu, \tag{2.21}$$

for t in the interval $(0,1)$. Let us remark that if $\alpha(\cdot)$ satisfies (2.20), then $\mu(\cdot) = m_{\alpha(\cdot)}$ satisfies (2.21). As a result

$$\mathcal{I}_p(\alpha(\cdot)) = \inf\left\{p^{-1}\int_0^1\int_M \left(\|Z(x,t)\|_x^p + |V(x,t)|^p\right) m_{\alpha(t)}dt : (Z,V) \in A(\alpha(\cdot))\right\}. \tag{2.22}$$

We are now ready to define an analog of $\bar{D}$ for 1-forms. Given (α^0,μ^0) and (α^1,μ^1) in $\mathcal{C}\times\mathcal{V}$, we define

$$\begin{aligned} S_p(\alpha^0,\mu^0;\alpha^1,\mu^1) = \inf\big\{\mathcal{I}_p(\alpha(\cdot),\mu(\cdot)) : (\alpha(0),\mu(0)) = (\alpha^0,\mu^0),\ (\alpha(1),\mu(1)) = (\alpha^1,\mu^1), \\ \text{for } C^1 \text{ paths } (\alpha(\cdot),\mu(\cdot))\big\}. \end{aligned}$$

Similarly, given α^0 and α^1 in $\mathcal{C}$, we set

$$S_p(\alpha^0,\alpha^1) = \inf\big\{\mathcal{I}_p(\alpha(\cdot)) : \alpha(0)=\alpha^0,\ \alpha^1=\alpha(1) \quad \text{for } C^1 \text{ paths } \alpha(\cdot)\big\}. \tag{2.23}$$

Remark 2.3

(i) If there is no C^1 path in $\mathcal{C}$ connecting α^0 to α^1, then by definition $S(\alpha^0,\alpha^1)$ is infinite.

(ii) If there is a continuously differentiable path in $\mathcal{C}$ connecting α^0 to α^1 and M is a compact manifold, then $S(\alpha^0,\alpha^1)<\infty$ because we can always find (Z,V) such that (2.20) is valid.
(iii) Evidently

$$S_2(\alpha^0,\mu^0;\alpha^1,\mu^1)\ge \bar{E}(\mu^0,\mu^1), \quad S_2(\alpha^0,\alpha^1)\ge \bar{E}(m_{\alpha^0},m_{\alpha^1}),$$

where, as in (2.11),

$$\bar{E}(\mu^0,\mu^1)=\inf\{\mathcal{J}(\mu(\cdot)) : \mu(0)=\mu^0,\ \mu^1=\mu(1)\},$$

with

$$\mathcal{J}(\mu(\cdot)) := \inf\left\{\frac{1}{2}\int_0^1\int_M \left(\|Z(x,t)\|_x^2+V(x,t)^2\right)\mu(t)dt : (Z,V)\in B(\mu(\cdot))\right\}.$$

□

For our next result, we express $\mathcal{I}$ as the action of a certain Lagrangian function. To prepare for this, we make some definitions.

Definition 2.2

(i) A vector field X is called an *α-contact vector field* if $\mathcal{L}_X\alpha = f\alpha$ for some scalar-valued continuous function f. As it is well-known, given a "Hamiltonian" $H : M\to\mathbb{R}$, there exists a unique contact α-vector field $X_H = X_{H,\alpha}$ such that $i_{X_H}\alpha = \alpha(X_H) = H$. The function f can be expressed in terms of H with the aid of the Reeb's vector field $R = R^\alpha$; indeed, $f = dH(R^\alpha)$, and as a result,

$$\mathcal{L}_{X_H}\alpha = dH(R^\alpha)\alpha.$$

(ii) Given any 1-form ν, we write $Z_{\nu,\alpha} = Z$ and $V_{\nu,\alpha} = V$ for a pair of vector field and continuous function that are uniquely determined by the requirements

$$i_Z\alpha = 0, \quad \nu+\mathcal{L}_Z\alpha = \nu + i_Z d\alpha = V\alpha. \tag{2.24}$$

□

Remark 2.4 In terms of the Reeb's vector field, we simply have $V_{\nu,\alpha} = \nu(R^\alpha)$. On the other hand, since $d\alpha_{|\eta}$ is symplectic for $\eta=\eta^\alpha$, the linear transformation $\mathcal{F}_\alpha(Z) = i_Z d\alpha$ is invertible from the space $\mathcal{X}_\alpha$ of η^α-valued vector fields, onto the space Λ_α of 1-forms γ such that $\gamma(R^\alpha) = 0$. Using the inverse of $\mathcal{F}_\alpha$, we can write

$$(Z_{\nu,\alpha}, V_{\nu,\alpha}) = \left(\hat{\mathcal{F}}_\alpha(\nu), \nu(R^\alpha)\right),$$

where

$$\hat{\mathcal{F}}_\alpha(\nu) = \mathcal{F}_\alpha^{-1}\left(\nu(R^\alpha)\alpha - \nu\right)$$

Similarly, $X_H = \hat{\mathcal{F}}_\alpha(dH) + HR^\alpha$. □

To explain further the significance of contact forms and the meaning of various concepts we have introduced so far, let us discuss the Euclidean case in Example 2.1.

Example 2.1 In the case of $M = \mathbb{R}^{2n-1}$ or $\mathbb{T}^{2n-1}$, with $n \geq 2$, we may express a 1-form α as $\alpha = u \cdot dx$ for a vector field u. Moreover

$$\beta_x(v_1, v_2) = d\alpha_x(v_1, v_2) = C(u)v_1 \cdot v_2,$$

where $C(u) = (Du)^* - Du$. (we are writing A^* for the transpose of A.) Since $C^* = -C$, we have that $\det C = (-1)^d \det C$. This implies that C cannot be invertible if the dimension is odd. Hence the null space l_x of $C(u)(x)$ is never trivial and in part (i) of Definition 2.1, we are assuming the this null space has the smallest possible dimension; this is the best non-degeneracy we can afford. The second condition of Definition 2.1 in this case simply means that $u(x) \cdot R(x) \neq 0$. Of course R is chosen so that $u(x) \cdot R(x) \equiv 1$. Writing $u^\perp$ and $R^\perp$ for the space of vectors perpendicular to u and R respectively, then $\eta = u^\perp$, and we may define a matrix $C'(u)$ which is not exactly the inverse of $C(u)$ (because $C(u)$ is not invertible), but it is specified uniquely by two requirements:

- (i) $C'(u)$ restricted to $R^\perp$ is the inverse of $C(u) : u^\perp \to R^\perp$.
- (ii) $C'(u)u = 0$.

The contact vector field associated with H is given by

$$X_H = -C'(u)\nabla H + HR.$$

Also, if $\nu = w \cdot dx$, then $Z_{w,u} = Z_{\nu,\alpha}$ is given by $Z_{w,u} = -C'(u)w$. Moreover, by a slight abuse of notation the equation (2.20) takes the form $u_t + \mathcal{L}_Z u = Vu$, where

$$\mathcal{L}_Z u = \nabla(u \cdot Z) + C(u)Z. \tag{2.25}$$

Let us write $m_\alpha = \rho_u(x)\ dx^1 \wedge \cdots \wedge dx^{2n-1}$. We fix the orientation by simply defining

$$\mathcal{C}^+ = \{u : u \text{ is a } C^1 \text{ vector field with } \rho_u(x) > 0 \text{ for every } x\}, \tag{2.26}$$

In particular, when $n = 2$, the form $\alpha = u \cdot dx$ is contact if and only if $\rho = u \cdot \xi$ is never 0, where $\xi = \nabla \times u$ is the curl (vorticity) of u. In this case the Reeb vector field is given by $R = \xi/(u \cdot \xi)$, and

$$\mathcal{L}_Z u = \nabla(u \cdot Z) + \xi \times Z, \tag{2.27}$$

$$\mathcal{C}^+ = \{u : u \text{ is a } C^1 \text{ vector field with } u(x) \cdot \xi(x) > 0 \text{ for every } x\}. \tag{2.28}$$

We also write $\bar{u} = u/\rho$. The vector field $Z_w = Z_{\nu,\alpha}$ for $\nu = w \cdot dx$ and the contact vector field associated with H are given by

$$Z_w = \bar{u} \times w, \quad X_H = \bar{u} \times \nabla H + HR. \tag{2.29}$$

□

In Theorem 2.4, we give a Lagrangian formulation for the expression $\mathcal{I}_p$.

Theorem 2.4 *We have $\mathcal{I}_p(\alpha(\cdot)) = \int_0^1 \mathcal{G}_p(\alpha(t), \dot{\alpha}(t))dt$, where*

$$\mathcal{G}_p(\alpha, \nu) = \inf_H \; p^{-1} \int_M \left[\|Z_{\nu,\alpha} + X_{H,\alpha}\|^p + |(\nu + dH)(R^\alpha)|^p \right] \, m_\alpha. \tag{2.30}$$

Moreover the function $\mathcal{G}_p$ is convex in ν-variable.

As in the Monge-Kantorovich problem, we would like to work out a dual formulation of our optimal problem associated with $\mathcal{I}_p$ and find expressions analogous to (2.6) and (2.7). In our setting it boils down to calculating the Legendre transform $\mathcal{H}_p(\alpha, Y)$ of the Lagrange-type functional $\mathcal{G}_p(\alpha, \nu)$. To prepare for the calculation of $\mathcal{H}_p$, we make some definitions.

Definition 2.3

(i) Observe that the Riemannian metric g allows us to define an operator $\sharp = \sharp^g$ that maps 1-forms to vector fields by requiring

$$g_x\left((\sharp\alpha)_x, v\right) = \alpha_x(v),$$

for every vector $v \in T_xM$. This duality also induces a metric on 1-forms; we define a norm $\|\cdot\|_x^*$ on T_x^*M by

$$\|\alpha_x\|_x^{*2} := \|(\sharp\alpha)_x\|_x^2 = 2 \sup_{v \in T_xM} \left(\alpha_x(v) - \frac{1}{2}\|v\|_x^2 \right) = \sup_{0 \neq v \in T_xM} \frac{\alpha_x(v)^2}{\|v\|_x^2}. \tag{2.31}$$

(Recall that $\|v\|_x^2 = g_x(v, v)$.)

(ii) We write vol_g for the volume form associated with the metric g. The density of m_α with respect to this volume form is denoted by ρ_α:

$$m_\alpha = \rho_\alpha \; vol_g.$$

(iii) Let us assume that M is a manifold without boundary. The divergence of a vector filed Y with respect to the metric g is denoted by $div_g Y = div Y$. More precisely, for every C^1 function $H : M \to \mathbb{R}$ of compact support,

$$\int_M (div_g Y) \; H \; vol_g = -\int_M dH(Y) \; vol_g. \tag{2.32}$$

(iv) We write $\mathcal{X}$ for the space of all continuous vector fields on M, and $\hat{\mathcal{X}}$ for the set of pairs $\hat{Z} = (Z, V)$ such that $Z \in \mathcal{X}$ and $V : M \to \mathbb{R}$ is a continuous function. We also write Λ for the space of 1-forms on M, and $\hat{\Lambda}$ for the space of pairs (γ, f) with $\gamma \in \Lambda$ and $f : M \to \mathbb{R}$ a continuous function. With slight abuse of the notation, we write $\|(\gamma_x, f(x))\|_x^{*2} = \|\gamma_x\|_x^{*2} + f(x)^2$.

(v) Given a 1-form α, we define two linear operators $\hat{\mathcal{L}}_\alpha : \hat{\mathcal{X}} \to \Lambda$ and $\hat{\mathcal{L}}_\alpha^* : \mathcal{X} \to \hat{\Lambda}$ by

$$\hat{\mathcal{L}}_\alpha \hat{Z} = \hat{\mathcal{L}}_\alpha(Z, V) = \mathcal{L}_Z \alpha - V\alpha, \qquad \hat{\mathcal{L}}_\alpha^* Y = -(\mathcal{B}_\alpha Y, \alpha(Y)), \tag{2.33}$$

where the operator $\mathcal{B}_\alpha : \mathcal{V} \to \Lambda$ is defined by the formula

$$\mathcal{B}_\alpha Y := i_Y d\alpha + (div_g Y)\, \alpha. \tag{2.34}$$

Indeed if we define two pairings

$$\langle \cdot, \cdot \rangle : \Lambda \times \mathcal{X} \to \mathbb{R}, \quad \langle \cdot, \cdot \rangle' : \hat{\Lambda} \times \hat{\mathcal{X}} \to \mathbb{R},$$

by

$$\langle \gamma, Y \rangle = \int_M \gamma(Y) \; vol_g, \quad \langle (\gamma, f), (Z, V) \rangle' = \int_M (\gamma(Z) + fV) \; vol_g,$$

then, using (2.34), what we really have is that the operator $\hat{\mathcal{L}}_\alpha^*$ is the dual of the operator $\hat{\mathcal{L}}_\alpha$;

$$\langle \hat{\mathcal{L}}_\alpha \hat{Z}, Y \rangle = \langle \hat{\mathcal{L}}_\alpha^* Y, \hat{Z} \rangle'. \tag{2.35}$$

□

Theorem 2.5 *Assume that M is a closed C^1 manifold, $\alpha \in \mathcal{C}$, and ν is a C^1 1-form. We also assume that $\rho_\alpha > 0$ everywhere on M. For $p > 1$ we have*

$$\mathcal{G}_p(\alpha, \nu) = \sup_Y \left(\int_M \nu(Y)\ vol_g - \mathcal{H}_q(\alpha, Y) \right),$$

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

$$\mathcal{H}_q(\alpha, Y) = \int_M \frac{1}{q\rho_\alpha^{p-1}} \left[\|\mathcal{B}_\alpha Y\|^{*q} + \alpha(Y)^q \right] vol_g = \int_M \frac{1}{q\rho_\alpha^{p-1}}\ \|\hat{\mathcal{L}}_\alpha^* Y\|^{*q}\ vol_g. \tag{2.36}$$

2.3 Symplectization and Riemannian Structures on $\mathcal{C}$

In this Subsection, we first describe a Riemanian metric on volume forms (or more generally probability measures) that was initiated by Otto (see [V1]). We then discuss a standard trick in Symplectic Geometry that is known as *symplectization* and would allow us to give a more natural description for the operators $\hat{\mathcal{L}}_\alpha$ and $\mathcal{B}_\alpha$ of (2.33) and (2.34). Symplectization also helps us to find a natural Riemannian Structure on $\mathcal{C}$ that has the same flavor as Otto's Riemannian metric on space of probability measures .

Let us write $\mathcal{P}$ for the space of probability measures μ on $\mathbb{R}^d$ such that $\int |x|^2 \mu(dx) < \infty$. After the work of F. Otto, we may equip $\mathcal{P}$ with a Riemannian metric which is closely related to the Wasserstein distance. To explain this, first we need to make sense of the tangent bundle of $\mathcal{P}$. The idea is that if $\mu^t(dx) = \rho(x,t)dx$ is a smooth path of measures that passes through $\mu^0(dx) = m(x)dx$, then the equation (1.5) is valid for some vector field Z. On the space of such vector fields, we use $L^2(\mu^0) = L^2(m)$ inner product, and this in turn induces a Riemmanian metric on $\mathcal{P}$. For this to work, we need a unique way of assigning a vector field Z to a smooth path that passes through μ^0. Note that the relationship between $r(x) = \rho_t(x, 0)$, and $Z^0 = Z(\cdot, 0)$ is given by

$$r + \nabla \cdot (mZ^0) = 0. \tag{2.37}$$

Given r, we wish to select a unique $\bar{Z}$ that satisfies (2.37). On account of our minimization problem (1.3), we select $\bar{Z}$ to minimize

$$\int \frac{1}{2} |Z^0|^2 m\ dx,$$

among all possible solutions of (2.37). Given a solution Z^0 of (1.3), observe that we also have

$$r + \nabla \cdot \big(m(Z^0 + Y/m) \big) = 0,$$

for any divergence free vector field Y. Hence, the minimizing $\bar{Z}$ is orthogonal to such divergence free vector fields;

$$\int \bar{Z}(Y/m) m\ dx = \int \bar{Z} Y\ dx = 0.$$

As a result, $\bar{Z} = \nabla\phi$ must be a gradient vector field. In summary, given $m \in \mathcal{P}$, our potential tangent "vector" r to $\mathcal{P}$ at m takes the form

$$r = -\nabla \cdot (m\nabla\phi),$$

an elliptic PDE for the scalar-valued function ϕ. Motivated by this, we define the tangent fiber $\hat{\mathcal{T}}_m\mathcal{P}$ at the measure $\mu^0(dx) = m(x)dx$ to be the $L^2(m)$-closure of

$$\hat{\mathcal{T}}^0_m\mathcal{P} = \{\nabla\phi \ : \ \phi \in C^\infty\}, \tag{2.38}$$

and equip $\hat{\mathcal{T}}^0_m\mathcal{P}$ with the inner product

$$\langle \nabla\phi, \nabla\phi' \rangle_m = \int \nabla\phi \cdot \nabla\phi' \ m \ dx.$$

We also define

$$\mathcal{T}^0_m\mathcal{P} = \{r = -\nabla(m\nabla\phi) \ : \ \phi \in C^\infty\}. \tag{2.39}$$

Furthermore, if $r = -\nabla(m\nabla\phi)$, $r' = -\nabla(m\nabla\phi')$, then

$$\langle r, r' \rangle_m = \int \nabla\phi \cdot \nabla\phi' \ m \ dx. \tag{2.40}$$

defines an inner product on $\mathcal{T}^0_m\mathcal{P}$.

As in the Otto's work, we would like to define a Riemannian metric on the tangent bundle $\mathcal{TC}$. For this we would like to assign a unique vector field to every tangent element. More precisely, given any contact form $\nu \in \mathcal{T}^0_\alpha\mathcal{C}$, we would like to assign a unique vector field Z and scalar-valued function V such that

$$\nu + \mathcal{L}_Z\alpha = V\alpha. \tag{2.41}$$

To describe our choice of (Z, V), we first discuss the notion of symplectization.

For a symplectization of M, we set

$$\begin{aligned}
&\hat{M} = M \times (0, \infty) = \{(x, \tau) : \ x \in M, \ \tau > 0\}, \\
&\hat{g}_{(x,\tau)}((v, a), (v', a')) = g_x(v, v') + aa', \quad \|(v, a)\|^2_{(x,\tau)} = \|v\|^2_x + a^2 \\
&\hat{\alpha}_{(x,\tau)}(v, a) = e^{-\tau}\alpha_x(v), \quad \hat{H}(x, \tau) = e^{-\tau}H(x).
\end{aligned}$$

Observe that $T_{(x,\tau)}\hat{M} = T_xM \times \mathbb{R}$. We can readily show

$$d\hat{\alpha} = e^{-\tau}(d\alpha + \alpha \wedge d\tau), \quad (d\hat{\alpha})^n = ne^{-n\tau} \ m_\alpha \wedge d\tau,$$

and that $(\hat{M}, d\hat{\alpha})$ is a symplectic manifold. Moreover, if $\hat{Z} = (Z, V)$, then

$$\mathcal{L}_{\hat{Z}}\hat{\alpha} = e^{-\tau}\left(\mathcal{L}_Z\alpha - V\alpha\right) = e^{-\tau}\hat{\mathcal{L}}_\alpha\hat{Z},$$

where $\hat{\mathcal{L}}_\alpha\hat{Z}$ is the same operator we defined in (2.33). There are two straightforward consequences of this formula:

- The contact vector field X_H associated with the Hamiltonian function H yields a Hamiltonian vector field for the symplectic form $d\hat{\alpha}$; if $\hat{X}_H = (X_H, dH(R^\alpha))$, then $\mathcal{L}_{\hat{X}_H}\hat{\alpha} = 0$ and $i_{\hat{X}_H} d\hat{\alpha} = -d\hat{H}$.
- The equation (2.41) simply means that for $\hat{\nu} = e^{-\tau}\nu$, we have

$$\hat{\nu} + \mathcal{L}_{\hat{Z}}\hat{\alpha} = 0. \tag{2.42}$$

On account of the minimization problem (1.7), we select a solution $\tilde{Z} = (\bar{Z}, \bar{V})$ of (2.42) that minimizes

$$\int_{\hat{M}} \frac{1}{2}\|\hat{Z}\|^2\ (d\hat{\alpha})^n.$$

We note that if $\hat{Z}$ satisfies (2.42), then $\hat{Z} + \hat{X}_H$ also satisfies (2.42), for every smooth H. Hence the minimizing $\tilde{Z}$ must satisfy

$$\int_{\hat{M}} \hat{g}(\tilde{Z}, \hat{X}_H)\ (d\hat{\alpha})^n = 0, \tag{2.43}$$

for every smooth function $H : M \to \mathbb{R}$. The meaning of the requirement (2.43) is that $\tilde{Z}$ is orthogonal to the kernel of $\hat{\mathcal{L}}_\alpha$ with respect to the $L^2((d\hat{\alpha})^n)$ inner product. This is equivalent to saying that $\tilde{Z}$ is in the range of the adjoint operator $\hat{\mathcal{L}}_\alpha^*$. In other words, the requirement (2.43) really means that for some vector field W,

$$\tilde{Z} = \hat{\mathcal{A}}_\alpha W := \rho_\alpha^{-1}(\mathcal{A}_\alpha W, \alpha(W)) := \rho_\alpha^{-1}\left(\sharp\mathcal{B}_\alpha W, \alpha(W)\right), \tag{2.44}$$

where the operator $\sharp$ was defined right before (2.31) (Recall that $m_\alpha = \rho_\alpha vol_g$.) Indeed as in (2.35),

$$\begin{aligned}
\int_{\hat{M}} \hat{g}(\hat{\mathcal{A}}_\alpha W, \hat{X}_H)\ (d\hat{\alpha})^n &= \int_{\hat{M}} \rho_\alpha^{-1}\left(g(\mathcal{A}_\alpha W, X_H) + \alpha(W) dH(R^\alpha)\right)\ (d\hat{\alpha})^n \\
&= n\int_M \rho_\alpha^{-1}\left(g(\mathcal{A}_\alpha W, X_H) + \alpha(W) dH(R^\alpha)\right)\ m_\alpha \\
&= n\int_M \left(\mathcal{B}_\alpha W(X_H) + \alpha(W) dH(R^\alpha)\right)\ vol_g \\
&= n\int_M \left(-i_{X_H} d\alpha(W) + (div_g W)\alpha(X_H) + \alpha(W) dH(R^\alpha)\right)\ vol_g \\
&= n\int_M \left((div_g W)\alpha(X_H) - d(\alpha(X_H))(W)\right)\ vol_g = 0.
\end{aligned}$$

Motivated by (2.44) and (2.42), we set

$$\begin{aligned}
&\hat{\mathcal{T}}_\alpha^0\mathcal{C} = \left\{\hat{\mathcal{A}}_\alpha W\ :\ W \text{ is a smooth vector field}\right\}, \\
&\mathcal{T}_\alpha^0\mathcal{C} = \left\{-\mathcal{L}_{\rho_\alpha^{-1}\sharp\mathcal{B}_\alpha W}\alpha + \rho_\alpha^{-1}\alpha(W)\alpha\ :\ W \text{ is a smooth vector field}\right\},
\end{aligned} \tag{2.45}$$

as in (2.38) and (2.39). Again, if

$$\nu = -\mathcal{L}_{\rho_\alpha^{-1}\sharp\mathcal{B}_\alpha W}\alpha + \rho_\alpha^{-1}\alpha(W)\alpha, \quad \nu' = -\mathcal{L}_{\rho_\alpha^{-1}\sharp\mathcal{B}_\alpha W'}\alpha + \rho_\alpha^{-1}\alpha(W')\alpha,$$

we define the inner product

$$\langle \nu, \nu' \rangle_\alpha = \int_M [\langle \mathcal{B}_\alpha W, \mathcal{B}_\alpha W' \rangle^* + \alpha(W)\alpha(W')]\, \rho\, dvol_g.$$

2.4 The Metric $\hat{S}$

So far we have formulated optimal transport problems for contact structures on a fixed Riemannian manifold. In fact there is a natural way of defining an optimal transport problem for which the Riemannian metric is compatible with the contact form and as a result changes with time. For such compatible metrics many of our formulas simplify. To explain this, let us first discuss the notion of *almost complex structures*.

Definition 2.4

(i) Given a contact form α, we set $\mathcal{J}^\alpha$ to denote the set of almost complex structures J that are compatible with α. More precisely, $J \in \mathcal{J}^\alpha$ is a x-continuous family $(J_x : \eta_x^\alpha \to \eta_x^\alpha)_{x \in M}$ such that $J^2 = -id$, and $g_x^J(v,w) = d\alpha_x(v, J_x w)$ defines a Riemannian metric on η. The latter condition means that $g_x^J(v,w) = g_x^J(w,v)$ and that $g_x^J(v,v) > 0$ for every non-zero $v \in \eta_x$. To extend g^J to a metric on M, we write $\pi_x : T_x M \to \eta_x$ for the R^α-projection

$$\pi_x(v) = v - \alpha_x(v) R^\alpha(x),$$

and set

$$\bar{g}_x^J(v,w) = g_x^J(\pi_x(v), \pi_x(w)) + \alpha_x(v)\alpha_x(w), \quad \|v\|_x^J = \bar{g}_x^J(v,v)^{1/2}.$$

We note that by definition

$$\alpha(v) = \bar{g}(R^\alpha, v), \qquad 1 = \alpha(R^\alpha) = \bar{g}(R^\alpha, R^\alpha). \tag{2.46}$$

(ii) Motivated by (2.46), we say that a pair $(\bar{g}, R)$ is *contact* if $\bar{g}$ is a Riemaniann metric, R is a unit vector field, and the form $R^{\bar{g}}$, defind by

$$\alpha^{\bar{g},R}(v) := \bar{g}(R, v),$$

is a contact form. Given a contact form α, we write $\hat{\mathcal{J}}^\alpha$ for the set of Riemannian metrics that are compatible with α. In other words, $\bar{g} \in \hat{\mathcal{J}}^\alpha$ if and only if $\alpha = \alpha^{\bar{g},R}$ for some vector field R. □

Finally we write div_J for the divergence associated with metric $\bar{g}^J$.

Given a path $((\alpha(t),\bar g(t)):\ t\in[0,1])$ such that $\bar g(t)\in\hat{\mathcal J}^{\alpha(t)}$ for every $t\in[0,1]$, we define

$$\hat{\mathcal I}_2(\alpha(\cdot),\bar g(\cdot))=\inf\left\{\frac{1}{2n!}\int_0^1\int_M\left(\left(\|Z(x,t)\|_x^{J(t)}\right)^2+V(x,t)^2\right)m_{\alpha(t)}\ dt:\ (Z,V)\in A(\alpha(t))\right\}.$$

Note that our definition for $\hat{\mathcal I}_2$ now has the additional factor $(n!)^{-1}$ and this very much related to the appearance of this factor in (2.51).

Theorem 2.6 *We have*

$$\hat{\mathcal I}_2(\alpha(\cdot),\bar g(\cdot))=\int_0^1\hat{\mathcal G}_2(\bar g(t),\dot\alpha(t))dt.$$

where

$$\hat{\mathcal G}_2(\bar g,\nu)=\inf_{f\in C^1}\int_M\frac12\left[(\|\nu-df\|^{\bar g*})^2+f^2\right]\ vol_{\bar g},\tag{2.47}$$

where $\|\cdot\|^{\bar g*}$ *was defined by* (2.31). *Moreover*

$$\hat{\mathcal G}_2(\bar g,\nu)=\sup_Y\left(\int\nu(Y)\ vol_{\bar g}-\hat{\mathcal H}(\bar g,Y)\right),\tag{2.48}$$

with

$$\hat{\mathcal H}(\bar g,Y)=\frac12\int_M\left[(\|Y(x)\|_x^{\bar g})^2+(div_{\bar g}Y(x))^2\right]\ vol_{\bar g}.$$

Remark 2.4 The minimizing (Z,V) in the variational problem associated with $\hat{\mathcal I}_2$ is related to the maximizing Y in the variational problem associated with $\mathcal G_2$ by

$$\bar g(Z,v)=(\mathcal B_\alpha Y)(v),\quad \alpha(Y)=V.$$

From this we learn

$$\pi_xZ=J_x\pi_xY,\quad \alpha(Z)=div_JY,\quad \alpha(Y)=V.$$

We may rewrite this as

$$Y=-J_x\pi_xZ+VR^\alpha,\quad \alpha(Z)=div_JY.$$

The pair (Z,V) must satisfy

$$div_J\left(-J_x\pi_xZ+VR^\alpha\right)=\alpha(Z).$$

This is simply the Euler-Lagrange equation for the minimizer in the definition of $\hat{\mathcal{I}}$. However, (Z,V) can be expressed as

$$(\hat{\mathcal{F}}(\nu+dH)+HR^\alpha, (\nu+dH)(R^\alpha)),$$

according to Theorem 2.6. Substituting this in the above equation yields an elliptic PDE for H:

$$div_J\left(-J_x\hat{\mathcal{F}}(\nu+dH)+(\nu+dH)(R^\alpha)R^\alpha\right)=H. \tag{2.49}$$

□

Remark 2.5 In view of symplectization we defined in Subsection 2.3, we may extend both $\bar{g}^J$ and J to $\hat{M}=M\times(0,\infty)$ in the following way: We define $\hat{J}_x: T_xM\times\mathbb{R}\to T_xM\times\mathbb{R}$, by

$$\hat{J}_{(x,\tau)}(v,\theta)=(J_x(\pi_x v)+\theta R^\alpha(x), -\alpha_x(v)), \tag{2.50}$$

and a metric $\hat{g}$ on the manifold $\hat{M}$ by

$$\begin{aligned}\hat{g}^J_x((v,\theta),(w,\theta'))&=g^J_x(\pi_x(v),\pi_x(w))+\alpha_x(v)\alpha_x(w)+\theta\theta',\\ \|(v,\theta)\|^J_x&=\hat{g}^J_x((v,\theta),(v,\theta))^{1/2}.\end{aligned} \tag{2.51}$$

Note that $\hat{J}$ is an almost complex structure for the symplectic manifold $(\hat{M},d\hat{\alpha})$ and the associated metric is given by $e^{-\tau}\hat{g}$;

$$e^{-\tau}\hat{g}^J_{(x,\tau)}((v,\theta),(w,\theta'))=d\hat{\alpha}_{(x,\tau)}((v,\theta),J_{(x,\tau)}(w,\theta')).$$

The corresponding volume form is given by

$$vol_{\hat{g}}:=e^{n\tau}(n!)^{-1}(d\hat{\alpha})^n=(n!)^{-1}m_\alpha\wedge d\tau. \tag{2.52}$$

3 Proofs of Theorems 2.3, 2.4 and 2.5

Proof of Theorems 2.4 The proof of (2.30) is straightforward and is an immediate consequence of the fact that if $\alpha:(0,1)\to\mathcal{C}$ is a C^1-path, then

- the pair $(\bar{Z},\bar{V})=(Z_{\dot{\alpha},\alpha},V_{\dot{\alpha},\alpha})$ belongs to $A(\alpha(\cdot))$, and
- for any $(Z,V)\in A(\alpha(\cdot))$, we have $\mathcal{L}_{Z-\bar{Z}}\alpha=(V-\bar{V})\alpha$, or equivalently $Z-\bar{Z}=X_{H,\alpha}$ and $V-\bar{V}=dH(R^\alpha)$ for $H=\alpha(Z-\bar{Z})$.

For the convexity of $\mathcal{G}_p$ in ν variable, observe that if we write $\hat{Z} = (Z, V)$ and $\|\hat{Z}\|^p$ for $\|Z\|^p + |V|^p$, then

$$\mathcal{G}_p(\alpha, \nu) = \inf \left\{ p^{-1} \int \|\hat{Z}\|^p \; m_\alpha : \; \nu + \hat{\mathcal{L}}_\alpha \hat{Z} = 0 \right\}, \tag{3.1}$$

where $\hat{\mathcal{L}}_\alpha$ was defined by (2.33). From this, the convexity is immediate; the expression

$$\theta \mathcal{G}_p(\alpha, \nu_1) + (1 - \theta) \mathcal{G}_p(\alpha, \nu_2),$$

is equal to

$$\begin{aligned}
&\inf \left\{ p^{-1} \int \left[\theta \|\hat{Z}_1\|^p + (1-\theta)\|\hat{Z}_2\|^p \right] \; m_\alpha : \; \nu_i + \hat{\mathcal{L}}_\alpha \hat{Z}_i = 0 \text{ for } i = 1, 2 \right\} \\
&\geq \inf \left\{ p^{-1} \int \|\theta \hat{Z}_1 + (1-\theta)\hat{Z}_2\|^p \; m_\alpha : \; \nu_i + \hat{\mathcal{L}}_\alpha \hat{Z}_i = 0 \text{ for } i = 1, 2 \right\} \\
&\geq \inf \left\{ p^{-1} \int \|\hat{Z}\|^p \; m_\alpha : \; (\theta \nu_1 + (1-\theta)\nu_2) + \hat{\mathcal{L}}_\alpha \hat{Z} = 0 \right\} \\
&= \mathcal{G}_p(\alpha, \theta\nu_1 + (1-\theta)\nu_2),
\end{aligned}$$

because $\hat{\mathcal{L}}_\alpha$ is linear. This completes the proof. $\square$

Proof of Theorems 2.5 The equation (2.37) is a consequence of *Minimax Principle* and (3.1): When $p > 1$,

$$\begin{aligned}
\mathcal{G}_p(\alpha, \nu) &= \inf_{\hat{Z} \in \hat{\mathcal{X}}} \sup_{Y \in \mathcal{X}} \left\{ p^{-1} \int \|\hat{Z}\|^p \; m_\alpha + \int \left(\hat{\mathcal{L}}_\alpha \hat{Z} + \nu \right)(Y) \; vol_g \right\} \\
&= \inf_{\hat{Z} \in \hat{\mathcal{X}}} \sup_{Y \in \mathcal{X}} \left\{ \int \nu(Y) \; vol_g - \left[\langle -\hat{\mathcal{L}}_\alpha^* Y, \hat{Z} \rangle' - p^{-1} \int \|\hat{Z}\|^p \; m_\alpha \right] \right\} \\
&= \sup_{Y \in \mathcal{X}} \left\{ \int \nu(Y) \; vol_g - \sup_{\hat{Z} \in \hat{\mathcal{X}}} \left[\langle -\hat{\mathcal{L}}_\alpha^* Y, \hat{Z} \rangle' - p^{-1} \int \|\hat{Z}\|^p \; \rho_\alpha \; vol_g \right] \right\} \\
&= \sup_{Y \in \mathcal{X}} \left\{ \int \nu(Y) \; vol_g - \int_M \frac{1}{q \rho_\alpha^{p-1}} \; \|\hat{\mathcal{L}}_\alpha^* Y\|^{*q} \; vol_g \right\},
\end{aligned} \tag{3.2}$$

where we used (2.35) for the second equality and interchanged supremum with infimum for the third equality. This interchange requires a justification. For this we need to specify an appropriate topologies on the spaces $\mathcal{X}$ and $\hat{\mathcal{X}}$. To make sense of $\hat{\mathcal{L}}_\alpha^* Y$, we need $Y \in C^1$. In fact we equip $\mathcal{X}$ with the standard C^1 topology. More precisely, $\mathcal{X}$ is a Banach space with norm $\|Y\|_{C^1} = \sup_x(\|Y(x)\|\|_x + \|\nabla Y(x)\|_x)$. However the space $\hat{\mathcal{X}}$ is equipped with the weak L^p topology. In other words, $\hat{X}_n = (X_n, V_n) \to \hat{X} = (X, V)$ means that

$$\lim_{n\to\infty} \int [g(X_n, A) + V_n W] \ vol_g = \int [g(X, A) + VW],$$

for every $\hat{A} = (A, W) \in L^q$. To interchange supremum with infimum (3.2), we apply classical Sion's theorem (see [K] or [R2]). For this, we need to verify the following conditions for the functional

$$F(\hat{Z}, Y) = \int \nu(Y) \ vol_g + \langle \hat{\mathcal{L}}^*_\alpha Y, \hat{Z}\rangle' + p^{-1} \int \|\hat{Z}\|^p \ m_\alpha :$$

- (i) F is convex in $\hat{Z}$, and concave and upper semi-continuous in Y.
- (ii) The set $E_Y(r) = \{\hat{Z} \ : \ F(\hat{Z}, Y) \le r\}$ is compact for every $r \in \mathbb{R}$.

To complete the proof of (2.37), it remains to verify (ii) because (i) is obvious. Let us write $c_0 = \int_M vol_g$ for the total volume of M, and

$$\|d\alpha\|^*_x := \sup_{|v_1|=|v_2|=1} |d\alpha_x(v_1, v_2)|, \qquad \|d\alpha\|^*_C = \sup_x \|d\alpha\|^*_x,$$
$$\|\alpha\|^*_C := \sup_x \|\alpha_x\|^*_x, \qquad \|\alpha\|^*_{C^1} = \|\alpha\|^*_C + \|d\alpha\|^*_C.$$

Now, using the elementary inequality

$$ab = \left(2^{-1/p}\rho^{1/p}a\right)\left(2^{1/p}\rho^{1/p}b\right) \le \frac{a^p}{2p} + \frac{2^{q/p}b^q}{q\rho^{q/p}},$$

we have

$$\begin{aligned} F(\hat{Z}, Y) &= \int \left[\nu(Y) - d\alpha(Y, Z) - (div_g Y)\alpha(Z) - \alpha(Y)V + p^{-1}\|\hat{Z}\|^p\rho_\alpha\right] \ vol_g \\ &\ge -c_0\|\nu\|^*_C\|Y\|_{C^1} - \int \|d\alpha\|^*_x\|Y\|_x\|Z\|_x \ vol_g - \|\alpha\|^*_C\|Y\|_{C^1} \int \|Z\|_x \ vol_g \\ &\qquad - \int \|\alpha\|^*_x\|Y\|_x|V(x)| \ vol_g + p^{-1}\int \|\hat{Z}\|^p \ \rho_\alpha \ vol_g \\ &\ge -c_0\|\nu\|^*_C\|Y\|_{C^1} - 2\|\alpha\|^*_{C^1}\|Y\|_{C^1}\int \|\hat{Z}\|_x \ vol_g + p^{-1}\int \|\hat{Z}\|^p \ \rho_\alpha \ vol_g \\ &\ge -c_0\|\nu\|^*_C\|Y\|_{C^1} - \frac{1}{2p}\int \|\hat{Z}\|^p \ \rho_\alpha \ vol_g + p^{-1}\int \|\hat{Z}\|^p \ \rho_\alpha \ vol_g \\ &\qquad\qquad - c_0 2^{q+q/p}q^{-1}\rho^{-q/p}\|\alpha\|^{*q}_{C^1}\|Y\|^q_{C^1} \\ &= -\|\nu\|^*_C\|Y\|_{C_1} - c_0 2^{q+q/p}q^{-1}\rho^{-q/p}\|\alpha\|^{*q}_{C^1}\|Y\|^q_{C^1} + \frac{1}{2p}\int \|\hat{Z}\|^p \ \rho_\alpha \ vol_g. \end{aligned}$$

From this we learn

$$E_Y(r) \subseteq \left\{ \hat{Z} : \int \|\hat{Z}\|^p \ \rho_\alpha \ vol_g \le r'(r, Y) \right\},$$

for a constant r' that depends on r and Y. Here we are using our assumptions $p > 1$ and $\rho_\alpha > 0$. Since bounded subsets of L^p are weakly precompacts, we learn that the set $E_Y(r)$ is precompact. On the other hand, we can readily show that the function F is lower semi-continuous in Y-variable with respect to the L^p weak topology. This completes the proof of property (ii) of Sion's theorem. This in turn completes the proof of Theorem 2.5. $\square$

4 Proof of Theorem 2.6

Proof of Theorem 2.6 For the proof of (2.48), take a metric $\bar{g} \in \hat{\mathcal{J}}^\alpha$ and write g for the restriction of $\bar{g}$ to $\eta = \eta^\alpha$. Recall that $d\alpha(v, w) = g(Jv, w)$ for $v, w \in \eta$ and that $v = \pi(v) + \alpha(v)R$, where $\pi(v) \in \eta$ denotes the η-component of v. Observe that if $f = \alpha(Z)$, then we can write

$$\begin{aligned}
(\mathcal{L}_Z\alpha - V\alpha)(Y) &= d\alpha(Z, Y) - V\alpha(Y) + df(Y) \\
&= g(J\pi(Z), \pi(Y)) - V\alpha(Y) + df(Y) = \bar{g}(J\pi(Z) - VR^\alpha, Y) + df(Y) \\
&=: -\bar{g}(A, Y) + df(Y).
\end{aligned}$$

As a result,

$$\nu + \mathcal{L}_Z\alpha - V\alpha = \nu - A^{\bar{g}} + df = 0, \quad \text{or} \quad \nu + df = A^{\bar{g}}.$$

On the other hand,

$$\left(\|A\|^{\bar{g}}\right)^2 + f^2 = \left(\|J\pi(Z)\|^g\right)^2 + V^2 + \alpha(Z)^2 = \left(\|Z\|^{\bar{g}}\right)^2 + V^2.$$

From this we learn that if the right-hand side is denoted by $\mathcal{G}'(\bar{g}, \nu)$, then

$$\mathcal{G}'(\bar{g}, \nu) \le \hat{\mathcal{G}}(\bar{g}, \nu). \tag{4.1}$$

It remains to prove the reverse inequality. In the definition of $\mathcal{G}'$, we may take infimum over $f \in H^1(M)$, that is, those f that are weakly differentiable with $\sharp df \in L^2(M)$. We now assert that the infimum $\bar{f}$ is attained and we have the relationships

$$\sharp(\nu + d\bar{f}) =: \bar{Y}, \qquad \bar{f} = div_{\bar{g}}\bar{Y}. \tag{4.2}$$

To see this, observe that we may take a sequence f_n such that

$$\mathcal{G}'(\bar{g}, \nu) = \lim_{n\to\infty} \int \frac{1}{2} \left(\|\nu + df_n\|^{*2} + f_n^2 \right) \ vol_{\bar{g}}.$$

Since $\sup_n \int f_n^2 \; vol_{\bar{g}} < \infty$, we may assume that $f_n \to \bar{f}$ for a subsequence. On the other hand, if we set $\bar{Y} = \sharp(\nu + d\bar{f})$, then $\bar{Y}$ is the weak limit of the sequence $\sharp(\nu + df_n)$ and by the lower semi-continuity,

$$\liminf_{n\to\infty} \int \frac{1}{2}\left(\|\nu + df_n\|^{*2} + f_n^2\right) \; vol_{\bar{g}} \geq \int \frac{1}{2}\left(\|\bar{Y}\|^2 + \bar{f}^2\right) \; vol_{\bar{g}}.$$

As a result,

$$\mathcal{G}'(\bar{g}, \nu) = \int \frac{1}{2}\left(\|\bar{Y}\|^2 + \bar{f}^2\right) \; vol_{\bar{g}},$$

which simply means that the infimum is achieved at $\bar{f}$. Now taking a smooth h and using

$$\int \frac{1}{2}\left(\|\nu + d(\bar{f} + th)\|^{*2} + (f + th)^2\right) \; vol_{\bar{g}} \geq \int \frac{1}{2}\left(\|\bar{Y}\|^2 + \bar{f}^2\right) \; vol_{\bar{g}},$$

for every $t \in \mathbb{R}$, we deduce

$$\int \left[g\left(\sharp(\nu + d\bar{f}), \sharp dh\right) + \bar{f}h\right] \; vol_{\bar{g}} = \int \left[-\left(div_{\bar{g}}\bar{Y}\right)h + \bar{f}h\right] \; vol_{\bar{g}} = 0.$$

Since h is an arbitrary smooth function, we deduce (4.2). Moreover,

$$\int \|\bar{Y}\|^2 \; vol_{\bar{g}} = \int \left(\nu(\bar{Y}) + d\bar{f}(\bar{Y})\right) \; vol_{\bar{g}} = \int \left(\nu(\bar{Y}) - \bar{f}(div_{\bar{g}}Y)\right) \; vol_{\bar{g}},$$

by (4.2). Hence

$$\mathcal{G}'(\bar{g}, \nu) = \int \frac{1}{2}\left(\|\bar{Y}\|^2 + \bar{f}^2\right) \; vol_{\bar{g}} = \int \left(\nu(\bar{Y}) - \frac{1}{2}\bar{f}^2 - (div_{\bar{g}}Y))^2\right) \; vol_{\bar{g}}.$$

Finally

$$\begin{aligned}
\hat{\mathcal{G}}_2(\alpha, J, \nu) &= \inf_{\hat{Z}\in\hat{\mathcal{X}}} \sup_{Y\in\mathcal{X}} \left\{\frac{1}{2}\int \left(\|\hat{Z}\|^J\right)^2 \; vol_J + \int \left(\hat{\mathcal{L}}_\alpha \hat{Z} + \nu\right)(Y) \; vol_J\right\} \\
(4.3) \qquad &= \inf_{\hat{Z}\in\hat{\mathcal{X}}} \sup_{Y\in\mathcal{X}} \left\{\int \nu(Y) \; vol_J - \left[\langle -\hat{\mathcal{L}}_\alpha^* Y, \hat{Z}\rangle' - \frac{1}{2}\int \left(\|\hat{Z}\|^J\right)^2 \; vol_J\right]\right\} \\
&\geq \sup_{Y\in\mathcal{X}} \left\{\int \nu(Y) \; vol_J - \sup_{\hat{Z}\in\hat{\mathcal{X}}} \left[\langle -\hat{\mathcal{L}}_\alpha^* Y, \hat{Z}\rangle' - \frac{1}{2}\int \left(\|\hat{Z}\|^J\right)^2 \; vol_J\right]\right\} \\
&=: \sup_Y \left(\int \nu(Y) \; vol_J - \hat{\mathcal{H}}'(\alpha, J, Y)\right),
\end{aligned}$$

Set $\bar{Y} = (Y, -div_J Y)$. We have

$$
\begin{aligned}
< \hat{\mathcal{L}}_\alpha^* Y, \hat{Z} >' &= -\int_M [(div_J Y)\alpha(Z) + d\alpha(Y,Z) + V\alpha(Y)] \ vol_J \\
&= -\int_{\hat{M}} e^{-\tau} [(div_J Y)\alpha(Z) + d\alpha(Y,Z) + V\alpha(Y)] \ vol_J \wedge d\tau \\
&= \int_{\hat{M}} d\hat{\alpha}(\hat{Z}, \bar{Y}) \ vol_J \wedge d\tau = -\int_{\hat{M}} e^{-\tau} \hat{g}^J(\hat{Z}, \hat{J}\bar{Y}) \ vol_J \wedge d\tau.
\end{aligned}
$$

From this and (4.3) we deduce,

$$
\begin{aligned}
\hat{\mathcal{H}}'(\alpha, J, \nu) &= \sup_{\hat{Z} \in \hat{\mathcal{X}}} \left[\int_{\hat{M}} e^{-\tau} \left(\hat{g}^J(\hat{Z}, \hat{J}\bar{Y}) - \frac{1}{2} \left(\|\hat{Z}\|^J \right)^2 \right) \ vol_J \wedge d\tau \right] \\
&= \int_{\hat{M}} e^{-\tau} \frac{1}{2} \left(\|\hat{J}\bar{Y}\|^J \right)^2 \ vol_J \wedge d\tau = \int_M \frac{1}{2} \left(\|\bar{Y}\|^J \right)^2 \ vol_J.
\end{aligned}
$$

□

References

[BB] J-D. Benamou , Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, **84**, 375–393 (2000).

[G] H. Greiges, An Introduction to Contact Topology. Cambridge University Press, Cambridge, 2008.

[K] H. Komiya, Elementary proof for Sion's minimax theorem. Kodai Math. J. **11** (1988), no. 1, 5–7.

[R1] F. Rezakhanlou, Lectures on Symplectic Geometry. http://math.berkeley.edu/ rezakhan

[R2] F. Rezakhanlou, Lectures on Large Deviation Principles. http://math.berkeley.edu/ rezakhan

[V1] C. Villani, Topics in Optimal Transportaion. Graduate Studies in Mathematics, AMS, 2003.

[V2] C. Villani, Optimal Transport: Old and New. Springer Verlag (Grundlehren der mathematischen Wissenschaften), 2008.