

## LINEAR PROGRAMMING INTERPRETATIONS OF MATHER’S VARIATIONAL PRINCIPLE<sup>\*,\*\*</sup>

L.C. EVANS<sup>1</sup> AND D. GOMES<sup>2</sup>

**Abstract.** We discuss some implications of linear programming for Mather theory [13–15] and its finite dimensional approximations. We find that the complementary slackness condition of duality theory formally implies that the Mather set lies in an  $n$ -dimensional graph and as well predicts the relevant nonlinear PDE for the “weak KAM” theory of Fathi [5–8].

**Mathematics Subject Classification.** 90C05, 35F20.

Received October 10, 2001. Revised February 13, 2002.

### 1. INTRODUCTION

This largely heuristic paper discusses some formal implications of linear programming duality theory to Mather’s variational problem in dynamics [13–15]. Our intention in working out this application is mostly to publicize the following ideas for studying nonlinear optimization problems.

#### Two general principles:

- **Relaxation.** It is often useful to “relax” a minimization problem, searching for a probability measure on an appropriate phase space. This procedure in effect “linearizes” the original problem, subject to constraints.
- **Duality.** Once this relaxed problem is found, we can sometimes apply linear programming duality theory. Often the dual formulation contains useful information about the structure of the solution of the original problem.

(Applications to the Monge–Kantorovich mass transfer problem following these guidelines are well-known: see for instance the survey paper [3].)

**Mather’s variational problem.** To formulate our application of these principles in dynamics, we begin with a smooth Lagrangian function  $L : \mathbb{R}^n \times \mathbb{T}^n \rightarrow \mathbb{R}$ ,  $L = L(q, x)$ , where  $q$  denotes velocity in  $\mathbb{R}^n$  and  $x$  denotes position in  $\mathbb{T}^n$ , the flat torus in  $n$  dimensions. Our primary hypotheses are that

the mapping  $x \mapsto L(q, x)$  is  $\mathbb{T}^n$  periodic

---

*Keywords and phrases:* Linear programming, duality, weak KAM theory.

\* Supported in part by NSF Grant DMS-0070480 and by the Miller Institute for Basic Research in Science, UC Berkeley.

\*\* Supported in part by FCT (Portugal) through programs POCTI, POCTI/32931/MAT/2000, BPD 1531/2000.

<sup>1</sup> Department of Mathematics, University of California, Berkeley, CA 94720, USA; e-mail: [evans@math.berkeley.edu](mailto:evans@math.berkeley.edu)

<sup>2</sup> Department of Mathematics, University of Texas, Austin, TX 78712, USA.

for each  $q \in \mathbb{R}^n$ , and

the mapping  $q \mapsto L(q, x)$  is uniformly convex

for each  $x \in \mathbb{T}^n$ .

The original problem is to search for curves  $\mathbf{x}(\cdot) : [0, \infty) \rightarrow \mathbb{R}^n$  that minimize for each time  $T > 0$  the classical action

$$\int_0^T L(\dot{\mathbf{x}}, \mathbf{x}) dt$$

among Lipschitz curves  $\mathbf{y}(\cdot)$  with  $\mathbf{x}(0) = \mathbf{y}(0)$ ,  $\mathbf{x}(T) = \mathbf{y}(T)$ . We call such a curve  $\mathbf{x}(\cdot)$  an *absolute minimizer*, and are interesting in finding absolute minimizers which satisfy for a given vector  $Q \in \mathbb{R}^n$  the asymptotic growth condition

$$\lim_{t \rightarrow \infty} \frac{\mathbf{x}(t)}{t} = Q. \tag{1.1}$$

In general however absolute minimizing curves with these given asymptotics do not exist.

Following instead the first general principle above, we relax the problem and look instead for a measure  $\mu$  on  $\mathbb{R}^n \times \mathbb{T}^n$  to minimize the generalized action

$$\int_{\mathbb{R}^n} \int_{\mathbb{T}^n} L(q, x) d\mu, \tag{1.2}$$

subject to the constraints that

$$\int_{\mathbb{R}^n} \int_{\mathbb{T}^n} q \cdot D\phi d\mu = 0 \quad \text{for all } \phi \in C^1(\mathbb{T}^n), \tag{1.3}$$

$$\mu(\mathbb{R}^n \times \mathbb{T}^n) = 1, \quad \mu \geq 0, \tag{1.4}$$

and

$$\int_{\mathbb{R}^n} \int_{\mathbb{T}^n} q d\mu = Q. \tag{1.5}$$

Condition (1.4) says of course that  $\mu$  is a probability measure and (1.5) generalizes (1.1). The requirement (1.3), that the measure  $\mu$  be flow invariant, generalizes the classical requirement that the action be computed along a curve. Mather has shown that in general there exists a minimizing measure in this sense.

We will discuss in this paper what information can be extracted from the second general principle mentioned above, namely application of linear programming duality theory. We do so firstly, and rigorously, for a natural finite-dimensional approximation to problem (1.2–1.5). Secondly we explain a formal application of duality for the true, infinite dimensional problem. Our main observation is that the duality theory immediately implies that the minimizing measure  $\mu$  lies on a graph in the  $q$ -direction over  $\mathbb{T}^n$ . Furthermore duality predicts the structure of the Hamiltonian PDE of the “weak KAM” theory of Fathi [6–8] and [5].

Now this and much more has already been discovered in the references cited above. Our point is therefore only to emphasize that the general principles above contain clues as to some interesting mathematics, which can indeed be made rigorous. It is surprising to us that general abstraction can point in such fruitful directions, and we hope that this viewpoint may usefully apply to other problems.

Our thanks to the referee for carefully reading this paper.

We are pleased to contribute to this memorial volume in honor of Professor J.-L. Lions, the author of many influential books in PDE theory. Our favorites are *Quelques méthodes de résolution des problèmes aux limites non linéaires* and *Asymptotic Analysis of Periodic Structures* (written with Bensoussan and Papanicolaou).

2. LINEAR PROGRAMMING: DISCRETE CASE

2.1. General linear programming

Let us review the basic structure of duality theory for finite-dimensional linear programming; cf. Bertsimas and Tsitsiklis [2], Franklin [9], Lax ([11], Chap. 13), etc.

First, we record the notation that if  $x \in \mathbb{R}^N$  is a vector,  $x = (x_1, x_2, \dots, x_N)$ , we write  $x \geq 0$  to mean  $x_i \geq 0$  for  $i = 1, \dots, N$ .

We are given vectors  $c \in \mathbb{R}^N$ ,  $b \in \mathbb{R}^M$  and an  $M \times N$  matrix  $A$ . The *primal problem* is to find  $\hat{x} \in \mathbb{R}^N$  to

$$(P) \quad \begin{cases} \text{minimize } c \cdot x, \\ \text{subject to the constraints that} \\ Ax = b, \ x \geq 0; \end{cases}$$

and the corresponding *dual problem* is to find  $\hat{y} \in \mathbb{R}^M$  to

$$(D) \quad \begin{cases} \text{maximize } y \cdot b, \\ \text{subject to the constraints that} \\ A^T y \leq c. \end{cases}$$

A vector  $x \in \mathbb{R}^N$  satisfying  $Ax = b$ ,  $x \geq 0$  is called *feasible for (P)*;  $y \in \mathbb{R}^M$  is *feasible for (D)* if  $A^T y \leq c$ .

The central assertion of finite dimensional linear programming is this:

**Theorem 2.1 (Strong duality).** *Assume that the collections of feasible vectors for (P) and for (D) are both nonempty. Then*

- (i) *problem (P) has a solution  $\hat{x}$  and (D) has a solution  $\hat{y}$ ;*
- (ii) *We have*

$$c \cdot \hat{x} = \hat{y} \cdot b \tag{2.1}$$

or, equivalently,

$$\hat{x} \cdot (A^T \hat{y} - c) = 0. \tag{2.2}$$

This last equality expresses the *complementary slackness condition*. This identity implies for each  $i = 1, \dots, N$ , that either  $\hat{x}_i = 0$  or  $(A^T \hat{y} - c)_i = 0$ , or both.

See Chapter 4 in Bertsimas and Tsitsiklis [2] for more, and note in particular the chart on page 151 of this reference.

2.2. Discrete approximation of Mather's problem

We as follows discretize velocity-position space  $\mathbb{R}^n \times \mathbb{T}^n$ . Fix large positive integers  $K, L$ , and introduce the multiindices

$$\begin{cases} i = (i_1, \dots, i_n), \ j = (j_1, \dots, j_n), \ \text{where} \\ i_k \in \{0, \dots, K - 1\}, \ j_k \in \{-L, \dots, L\} \ \text{for } k = 1, \dots, n. \end{cases} \tag{2.3}$$

We write  $i \in I$  and  $j \in J$  if (2.3) holds, and note that  $|I| = K^n$ ,  $|J| = (2L + 1)^n$ . We define as well the *discrete Lagrangian*  $\lambda = ((\lambda_{ij}))$ , where

$$\lambda_{ij} := L \left( \frac{j}{K}, \frac{i}{K} \right). \tag{2.4}$$

(These is a slight notational problem here, since the letter  $L$  is also used in (2.3). When  $L$  denotes the Lagrangian in this section, we will always make this clear by displaying its arguments.)

The discrete *primal problem* is to find  $\mu = ((\mu_{ij}))$  to

$$\text{minimize } \frac{1}{|I||J|} \sum_{i,j} \lambda_{ij} \mu_{ij} \tag{2.5}$$

subject to the following constraints. We firstly require that

$$\frac{1}{|I||J|} \sum_{k=1}^n \sum_j j_k (\mu_{i-e_{k,j}} - \mu_{ij}) = 0 \quad \text{for all } i \in I. \tag{2.6}$$

Here we assume that all quantities are periodic in the variables  $i$ , and write  $e_k := (0, \dots, 1, \dots, 0)$ , where the 1 lies in the  $k^{\text{th}}$ -slot. The identities (2.6) form the discrete analogue of the flow-invariance requirement (1.3). We also want  $\mu$  to be a discrete probability measure, and so ask that

$$\frac{1}{|I||J|} \sum_{i,j} \mu_{ij} = 1, \quad \mu_{ij} \geq 0 \quad \text{for all } i, j. \tag{2.7}$$

Lastly we enforce the analogue of (1.5):

$$\frac{1}{|I||J|} \sum_{i,j} \frac{j_k}{K} \mu_{ij} = Q_k \quad \text{for } k \in \{1, \dots, n\}, \tag{2.8}$$

where  $Q = (Q_1, \dots, Q_n)$  is a given vector in  $\mathbb{R}^n$ .

In summary, we want to find  $\mu$  to minimize (2.5) subject to (2.6–2.8). There are in all  $M = |I| + 1 + n = K^n + 1 + n$  equality constraints, for the  $N = |I||J| = (2L + 1)^n K^n$  unknown components of  $\mu = ((\mu_{ij}))$ .

### 2.3. The dual problem

The foregoing problem is of the form (P) for

$$c = \frac{1}{|I||J|} ((\lambda_{ij})), \quad b = |I||J| \left( \underbrace{0, \dots, 0}_{|I| \text{ zeros}}, 1, Q_1, \dots, Q_n \right) \tag{2.9}$$

and an appropriate  $M \times N$  matrix  $A$ .

We do not attempt to write out  $A$  explicitly, but as follows use duality to compute  $A^T$ . Let us take  $x = \mu \in \mathbb{R}^N$ ,  $\mu = ((\mu_{ij}))$  as above, and also write  $y \in \mathbb{R}^M$  as

$$y = (v_i \ (i \in I), w_0, w_1, \dots, w_n). \tag{2.10}$$

Then

$$\begin{aligned} A^T y \cdot x = y \cdot Ax &= \frac{1}{|J|} \sum_{k=1}^n \sum_{i,j} v_i j_k (\mu_{i-e_k, j} - \mu_{ij}) \\ &+ \frac{1}{|I||J|} \sum_{i,j} w_0 \mu_{ij} + \frac{1}{|I||J|} \sum_{k=1}^n \sum_{i,j} w_k \frac{j_k}{K} \mu_{ij} \\ &= \frac{1}{|I||J|} \sum_{i,j} \mu_{ij} \left( \sum_{k=1}^n (v_{i+e_k} - v_i) j_k + w_0 + \sum_{k=1}^n \frac{w_k j_k}{K} \right). \end{aligned}$$

Hence the  $(i, j)^{\text{th}}$  component of  $A^T y$  is

$$\frac{1}{|I||J|} \left( \sum_{k=1}^n (v_{i+e_k} - v_i) j_k + w_0 + \sum_{k=1}^n \frac{w_k j_k}{K} \right), \tag{2.11}$$

where  $y$  has the form (2.10).

Suppose now that  $\hat{x} = \mu, \hat{y} = (v_i (i \in I), w_0, w_1, \dots, w_n)$  are optimal. According to the constraints and the complementary slackness condition (2.2), we have

$$\begin{cases} w_0 + \sum_{k=1}^n \left( w_k + \frac{v_{i+e_k} - v_i}{1/K} \right) \frac{j_k}{K} \leq \lambda_{ij} \\ \text{for all indices } i \in I, j \in J, \text{ with equality if } \mu_{ij} > 0. \end{cases} \tag{2.12}$$

**2.4. Interpretations of complementary slackness**

**a. Geometry of the support of  $\mu$**

For each fixed multiindex  $i \in I$ , define the function

$$\psi(q) := L \left( q, \frac{i}{K} \right) - \sum_{k=1}^n \left( w_k + \frac{v_{i+e_k} - v_i}{1/K} \right) q_k - w_0 \tag{2.13}$$

for  $q \in \mathbb{R}^n, q = (q_1, \dots, q_n)$ . Then condition (2.12) states

$$\psi \left( \frac{j}{K} \right) \geq 0 \text{ for all } j \in J, \text{ with } \psi \left( \frac{j}{K} \right) = 0 \text{ if } \mu_{ij} > 0. \tag{2.14}$$

But the Lagrangian  $L$ , and so also  $\psi$ , are uniformly convex in  $q$ . Consequently  $\psi \left( \frac{j}{K} \right) = 0$  for at most two adjacent lattice points in any coordinate direction, as illustrated. Hence

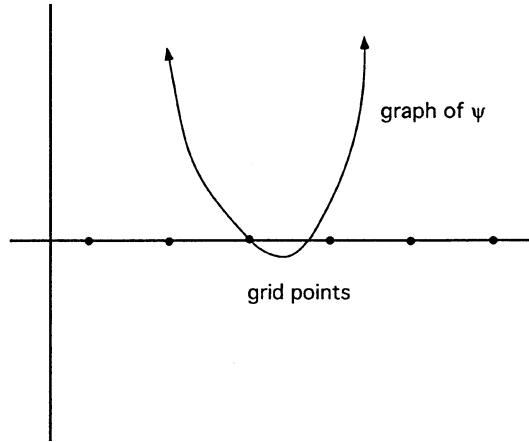
$$\begin{cases} \text{for each index } i \in I, \mu_{ij} > 0 \text{ for at most } 2^n \text{ values of } j \in J. \text{ These} \\ \text{points lie in a subset of the vertices of a cube of side length } 1/K. \end{cases} \tag{2.15}$$

So duality theory predicts that the support of  $\mu$  is approximately a graph.

**b. Approximate solutions of the corrector PDE**

According to (2.12),

$$w_0 + (w + D_K v) \cdot \frac{j}{K} \leq \lambda_{ij} = L \left( \frac{j}{K}, \frac{i}{K} \right) \tag{2.16}$$



where  $w = (w_1, \dots, w_n)$ ,  $D_K v := \left( \dots, \frac{v_i + \epsilon_k - v_i}{1/K} \dots \right)$ , with equality if  $\mu_{ij} > 0$ . Therefore for each  $i \in I$  we have

$$H_K \left( w + D_K v, \frac{i}{K} \right) := \sup_{j \in J} \left\{ \frac{j}{K} \cdot (w + D_K v) - L \left( \frac{j}{K}, \frac{i}{K} \right) \right\} \leq -w_0,$$

with equality if there exists  $j \in J$  with  $\mu_{ij} > 0$ . We write

$$P_K := w, \quad \bar{H}_K(P_K) := -w_0.$$

Then

$$H_K \left( P_K + D_K v, \frac{i}{K} \right) \leq \bar{H}_K(P_K) \quad \text{for } i \in I, j \in J$$

and

$$H_K \left( P_K + D_K v, \frac{i}{K} \right) = \bar{H}_K(P_K) \quad \text{whenever } \mu > 0.$$

This is a discrete analogue of the generalized eikonal equation  $H(P + Dv, x) = \bar{H}(P)$  from weak KAM theory, as discussed in the next section.

**Remark.** The foregoing suggests a numerical approximation scheme for the generalized eikonal equation. It is perhaps interesting that this procedure does not involve any sort of upwinding method, to compute the correct, viscosity solution across shocks. Presumably this is because, as discussed for instance in [5], there are no discontinuities of  $Dv$  on the support of projection  $\sigma$  of the Mather measure onto  $\mathbb{T}^n$ .

### 3. LINEAR PROGRAMMING: CONTINUUM CASE

In this section we discuss analogues for Mather’s problem (1.2–1.5) of the foregoing discrete calculations. These involve infinite dimensional linear programming issues, for which the book of Anderson and Nash [1] provides a good introduction. Our viewpoint is not that infinite dimensional linear programming immediately provides us with useful theorems. For instance there are subtle problems in finding the proper abstract spaces. Rather we take on the more modest task of pointing out that certain known facts from weak KAM theory are predicted by linear programming, at least heuristically.

### 3.1. Infinite dimensional linear programming

One infinite dimensional analogue of the linear programming problem discussed in Section 2.1 is this. Let  $X, Y$  be real topological vector spaces, and assume  $X$  is endowed with a partial ordering, which is compatible with the linear structure and which induces a partial ordering on its dual space  $X^*$ . We denote the pairing between a space and its dual by  $\langle \cdot, \cdot \rangle$ .

Fix  $c^* \in X^*, b \in Y$  and suppose  $A : X \rightarrow Y$  is a continuous linear operator, with adjoint  $A^* : Y^* \rightarrow X^*$ . The *primal problem* is to find  $\hat{x} \in X$  to

$$(P) \quad \begin{cases} \text{minimize } \langle c^*, x \rangle, \\ \text{subject to the constraints that} \\ Ax = b, \quad x \geq 0; \end{cases}$$

and the *dual problem* is to find  $\hat{y}^* \in Y^*$  to

$$(D) \quad \begin{cases} \text{maximize } \langle y^*, b \rangle, \\ \text{subject to the constraints that} \\ A^*y^* \leq c^*. \end{cases}$$

In this setting we do not generally have a strong duality assertion like Theorem 2.1. We can however assert *weak duality*:

**Theorem 3.1 (Weak duality).**

(i) Assume that  $x$  is feasible for (P) and  $y^*$  is feasible for (D). Then

$$\langle y^*, b \rangle \leq \langle c^*, x \rangle.$$

(ii) Furthermore, if  $\hat{x}$  is feasible for (P) and  $\hat{y}^*$  is feasible for (D), and if

$$\langle \hat{y}^*, b \rangle = \langle c^*, \hat{x} \rangle, \tag{3.1}$$

then  $\hat{x}$  is optimal for (P) and  $\hat{y}^*$  is optimal for (D).

Assertion (3.1) is equivalent to

$$\langle A^*\hat{y}^* - c^*, \hat{x} \rangle = 0. \tag{3.2}$$

The identity (3.2) is the *complementary slackness condition*.

It is also useful to define the *duality gap*

$$\Delta := \inf\{\langle c^*, x \rangle \mid Ax = b, \quad x \geq 0\} - \sup\{\langle y^*, b \rangle \mid A^*y^* \leq c^*\}. \tag{3.3}$$

### 3.2. Primal problem

In this section we cast Mather's variational problem into the foregoing framework. Let us firstly take

$$X = \mathcal{M}(\mathbb{R}^n \times \mathbb{T}^n), \tag{3.4}$$

the space of Radon measures on  $\mathbb{R}^n \times \mathbb{T}^n$ , and

$$Y = C^1(\mathbb{T}^n)^* \times \mathbb{R} \times \mathbb{R}^n. \tag{3.5}$$

Let  $c^* = L \in X^*, b = (0^*, 1, Q) \in Y$ , where  $0^*$  denotes the zero functional.

If  $\mu \in X$ , we write

$$A\mu := \left( L_\mu, \mu(\mathbb{R}^n \times \mathbb{T}^n), \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} q \, d\mu \right). \quad (3.6)$$

In this expression  $L_\mu$  denotes the linear functional defined for each  $\phi \in C^1(\mathbb{T}^n)$  by the formula

$$L_\mu \phi := \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} q \cdot D\phi(x) \, d\mu. \quad (3.7)$$

The primal problem (P) is therefore to find a measure  $\mu$  to

$$\text{minimize } \langle c^*, \mu \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} L \, d\mu,$$

subject to the requirements that

$$\mu \geq 0, \quad A\mu = b. \quad (3.8)$$

### 3.3. Dual problem

Let  $y^* = (v, w_0, w) \in Y^* = C^1(\mathbb{T}^n)^{**} \times \mathbb{R} \times \mathbb{R}^n$ . We will vastly simplify, by *simply assuming hereafter that in fact*  $v \in C^1(\mathbb{T}^n)$ .

We compute  $A^*y^*$  by calculating for  $x = \mu$  that

$$\langle A^*y^*, x \rangle = \langle y^*, Ax \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} q \cdot Dv(x) + w_0 + w \cdot q \, d\mu.$$

We deduce

$$A^*y^* = w_0 + q \cdot (w + Dv).$$

So the dual problem is to find  $y^* = (v, w_0, w)$  to

$$\text{maximize } \langle y^*, b \rangle = w_0 + Q \cdot w, \quad (3.9)$$

subject to the pointwise constraints that

$$w_0 + q \cdot (w + Dv) \leq L(q, x) \quad \text{in } \mathbb{R}^n \times \mathbb{T}^n. \quad (3.10)$$

### 3.4. Interpretations of complementary slackness

**a. Geometry of the support of  $\mu$ .** Assume now that  $\hat{x} = \mu$  and  $\hat{y}^* = (v, w_0, w)$  are optimal, and as before suppose as well that  $v \in C^1(\mathbb{T}^n)$ . Then the complementary slackness condition implies

$$w_0 + q \cdot (w + Dv) = L(q, x) \quad \text{on } \text{spt}(\mu). \quad (3.11)$$

For each fixed  $x \in \mathbb{T}^n$ , define the function

$$\psi(q) := L(q, x) - q \cdot (w + Dv) - w_0.$$

Then

$$\psi \geq 0 \quad \text{on } \mathbb{R}^n, \quad \text{and } \psi = 0 \quad \text{on } \text{spt}(\mu). \quad (3.12)$$



Since  $L$  and therefore  $\psi$  are uniformly convex in the variable  $q$ , it follows for each  $x \in \mathbb{T}^n$  that the support of  $\mu$  in  $\mathbb{R}^n \times \{x\}$  consists of at most one point  $q \in \mathbb{R}^n$ . Since the gradient of  $\psi(q)$  vanishes at its minimum, we in fact have  $D_q L(q, x) = w + Dv(x)$ . Consequently, the Mather set  $\text{spt}(\mu)$  lies on the  $n$ -dimensional graph  $q = (D_q L(\cdot, x))^{-1}(w + Dv(x))$  over  $\mathbb{T}^n$ .

**b. Solutions of the generalized eikonal PDE.** Let

$$H(p, x) := \max_{q \in \mathbb{R}^n} \{q \cdot p - L(q, x)\}$$

denote the Hamiltonian associated with the Lagrangian  $L$ . In view of (3.10), we have

$$H(w + Dv, x) := \max_{q \in \mathbb{R}^n} \{q \cdot (w + Dv) - L(q, x)\} \leq -w_0 \quad \text{on } \mathbb{T}^n, \tag{3.13}$$

with equality on the support of  $\sigma := \text{proj}_x \mu$ , the projection of  $\mu$  onto the  $x$ -variables. We change notation, now to write

$$P := w, \quad \bar{H}(P) := -w_0. \tag{3.14}$$

Then

$$H(P + Dv, x) \leq \bar{H}(P) \quad \text{on } \mathbb{T}^n,$$

with

$$H(P + Dv, x) = \bar{H}(P) \quad \text{on } \text{spt}(\sigma). \tag{3.15}$$

This is the generalized eikonal equation, the basic PDE for “weak KAM theory”: see Fathi [6–8] and [5]. We observe also that the Mather set must lie on the graph  $q = D_p H(P + Dv(x), x)$ .

In view of the change of notation (3.14), the maximization principle (3.9) is now to find

$$\max_P (Q \cdot P - \bar{H}(P)) =: \bar{L}(Q),$$

the effective Lagrangian evaluated at the given vector  $Q$ .

**Remark.** Observe that linear programming does not predict that the generalized eikonal PDE hold everywhere, but only on  $\text{spt}(\sigma)$ , which is the projection of the Mather set onto  $\mathbb{T}^n$ . This is consistent with the alternative variational method developed in [4].

**c. Duality gap.** The foregoing deductions we made under the assumption that we had (i) a measure  $\hat{x} = \mu$  solving the primal problem and (ii) a triple  $\hat{y}^* = (v, w_0, w) \in C^1(\mathbb{T}^n) \times \mathbb{R} \times \mathbb{R}^n$  solving the dual problem (D).

However, we are not aware of any general theorem asserting the existence of a solution of (D) as above, and consequently the conclusions before are unjustified. However, we can in fact show rigorously that there is no duality gap: recall the definition (3.3). Indeed, as explained in the work of Mather and of Fathi, we have

$$\inf \{ \langle c^*, x \rangle \mid Ax = b, x \geq 0 \} = \bar{L}(Q), \tag{3.16}$$

where  $\bar{L}$  is the effective Lagrangian, the convex dual of  $\bar{H}$ . On the other hand, according to Lions *et al.* [12], Fathi [5–8], etc., for each  $P \in \mathbb{R}^n$  there exists a Lipschitz continuous function  $v$  solving

$$H(P + Dv, x) = \bar{H}(P) \quad \text{in } \mathbb{T}^n$$

in the sense of viscosity solutions. As in [5], we introduce the mollified function  $v^\varepsilon := \eta_\varepsilon * v$ , where  $\eta_\varepsilon$  is a radial convolution kernel in the variables  $x$ . Since the PDE  $H(P + Dv, x) = \bar{H}(P)$  holds pointwise a.e., we have

$$H(P + Dv^\varepsilon, x) \leq \bar{H}(P) + C\varepsilon \tag{3.17}$$

for each  $x \in \mathbb{T}^n$ .

Now take  $y_\varepsilon^* = (v^\varepsilon, -\bar{H}(P) - C\varepsilon, P)$ . Owing to (3.17), we have

$$\begin{aligned} w_0 + q \cdot (w + Dv^\varepsilon) &= -\bar{H}(P) - C\varepsilon + q \cdot (P + Dv^\varepsilon) \\ &\leq \max_r \{r \cdot (P + Dv^\varepsilon) - L(r, x)\} - \bar{H}(P) - C\varepsilon + L(q, x) \\ &= H(P + Dv^\varepsilon, x) - \bar{H}(P) - C\varepsilon + L(q, x) \\ &\leq L(q, x) \quad \text{in } \mathbb{R}^n \times \mathbb{T}^n. \end{aligned}$$

Hence  $y_\varepsilon^*$  is feasible for the dual problem. Consequently,

$$\begin{aligned} \sup\{\langle y^*, b \rangle \mid A^* y^* \leq c^*\} &\geq -\bar{H}(P) - C\varepsilon + Q \cdot P \\ &= \bar{L}(Q) - C\varepsilon \\ &= \inf\{\langle c^*, x \rangle \mid Ax = b, x \geq 0\} - C\varepsilon. \end{aligned}$$

Since this holds for each  $\varepsilon > 0$ , the duality gap  $\Delta$  is zero.

**Remark.** In light of the foregoing, we propose these interpretations:

- the vector  $P$  is the Lagrange multiplier for the constraint (1.5), that the measure  $\mu$  have the rotation vector  $Q$ ;
- the number  $\bar{H}(P)$  is the Lagrange multiplier for the constraint (1.4) that  $\mu$  be a probability measure;
- The function  $v$ , solving the generalized eikonal equation, is the Lagrange multiplier for the constraint (1.3) that  $\mu$  be flow invariant.

## REFERENCES

- [1] E.J. Anderson and P. Nash, *Linear Programming in Infinite Dimensional Spaces*. Wiley (1987).
- [2] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*. Athena Scientific (1997).
- [3] L.C. Evans, *Partial differential equations and Monge–Kantorovich mass transfer (survey paper)*. Available at the website of LCE, at [math.berkeley.edu](http://math.berkeley.edu)
- [4] L.C. Evans, Some new PDE methods for weak KAM theory. *Calc. Var. Partial Differential Equations* (to appear).
- [5] L.C. Evans and D. Gomes, Effective Hamiltonians and averaging for Hamiltonian dynamics I. *Arch. Rational Mech. Anal.* **157** (2001) 1-33.
- [6] A. Fathi, Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens. *C. R. Acad. Sci. Paris Sér. I Math.* **324** (1997) 1043-1046.
- [7] A. Fathi, Solutions KAM faibles conjuguées et barrières de Peierls. *C. R. Acad. Sci. Paris Sér. I Math.* **325** (1997) 649-652.
- [8] A. Fathi, Weak KAM theory in Lagrangian Dynamics, Preliminary Version. *Lecture Notes* (2001).
- [9] J. Franklin, *Methods of Mathematical Economics*. SIAM, *Classics in Appl. Math.* **37** (2002).
- [10] D. Gomes, *Numerical methods and Hamilton–Jacobi equations* (to appear).
- [11] P. Lax, *Linear Algebra*. John Wiley (1997).
- [12] P.-L. Lions, G. Papanicolaou and S.R.S. Varadhan, *Homogenization of Hamilton–Jacobi equations*. CIRCA (1988) (unpublished).
- [13] J. Mather, Minimal measures. *Comment. Math Helvetici* **64** (1989) 375-394.
- [14] J. Mather, Action minimizing invariant measures for positive definite Lagrangian systems. *Math. Z.* **207** (1991) 169-207.
- [15] J. Mather and G. Forni, *Action minimizing orbits in Hamiltonian systems*. Transition to Chaos in Classical and Quantum Mechanics, edited by S. Graffi. Springer, *Lecture Notes in Math.* **1589** (1994).