EFFECTIVE HAMILTONIANS AND AVERAGING FOR HAMILTONIAN DYNAMICS II

L. C. Evans [1] and D. Gomes [2]

Abstract. We extend to time-dependent Hamiltonians some of the PDE methods from our previous paper [E-G1], and in particular the theory of "effective Hamiltonians" introduced by Lions, Papanicolaou and Varadhan [L-P-V]. These PDE techniques augment the variational approach of Mather [Mt1-4, M-F] and the weak KAM methods of Fathi [F1-5].

We also provide a weak interpretation of adiabatic invariance of the action and suggest a formula for the Berry–Hannay geometric phase in terms of an effective Hamiltonian.

1. Introduction.

We continue our investigation from [E-G1], employing nonlinear PDE methods to augment the variational approach of Mather [Mt1-4, M-F] and the weak KAM methods of Fathi [F1-5] in the study of Hamiltonian equations with many degrees of freedom. This paper treats Hamiltonians $H = H(p, x, t)$ which depend upon time.

In this first section we outline the basic issues arising as we try to extend our earlier methods to time–dependent Hamiltonians, most notably the differences between asymptotic limits of trajectories with "fast" versus "slow" variations in time. These comments provide an overview of the entire paper.

In §2 we explain how to construct the effective Hamiltonian $\hat{H}$ in the case that H is periodic in both x and t, and along the way we build a viscosity solution w of an appropriate Hamilton–Jacobi cell problem. Section 3 demonstrates how $\hat{H}$ and the related Lagrangian $\hat{L}$ govern asymptotics of Hamiltonian dynamics with fast, periodic variations in both space and time.

Section 4 introduces dynamics with fast, periodic variations in space, but slow variations in time. The rescaled limit is now controlled by $\{\bar{H}(\cdot, t)\}_{t\geq 0}$, the overbar denoting the effective Hamiltonian computed from fast variations in the variable x only, as in [E-G1].

An interesting problem is to understand the relationships between $\hat{H}$ and $\{\bar{H}(\cdot, t)\}_{t\geq 0}$, in other words between averaging in x, t together and averaging in x alone for each fixed

[1] Supported in part by NSF Grants DMS-9424342 and DMS-0070480.

[2] Supported in part by FCT (Portugal) through programs POCTI, POCTI/32931/MAT/2000, BPD 1531/2000 and by NSF grant DMS 97-29992.

time. Our results here are not very definitive, but for the reader's convenience we record some relationships for integrable systems under a strong assumption about the averaging. This leads to the Berry–Hannay phase correction, a formal expression for which in terms of $\hat{H}$ we propose in §6.

Much of what follows appears in somewhat different form in the second author's thesis [G1]. We have in this and our previous paper largely rederived using nonlinear PDE techniques earlier results from dynamics, due to Mather, Mañé, Fathi and others. The appendix clarifies some of the connections between our work (and notation) and theirs.

We are very grateful to A. Fathi for pointing out some errors in an earlier version of this paper.

1.1. Changing variables in x and t.

We begin by taking $H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $H = H(p, x, t)$, to be a smooth time–dependent Hamiltonian, for which we consider the corresponding flow

$$\begin{cases} \dot{\mathbf{x}} = D_p H(\mathbf{p}, \mathbf{x}, t) \\ \dot{\mathbf{p}} = -D_x H(\mathbf{p}, \mathbf{x}, t). \end{cases} \tag{1.1}$$

The dynamics (1.1) transform to read

$$\begin{cases} \dot{\mathbf{X}} = D\hat{H}(\mathbf{P}) \\ \dot{\mathbf{P}} = 0, \end{cases} \tag{1.2}$$

under the canonical change of variables

$$(p, x) \to (P, X), \tag{1.3}$$

provided

$$\begin{cases} p = D_x u(P, x, t) \\ X = D_P u(P, x, t), \end{cases} \tag{1.4}$$

and the generating function $u = u(P, x, t)$ satisfies the Hamilton–Jacobi PDE

$$u_t + H(D_x u, x, t) = \hat{H}(P) \quad \text{in } \mathbb{R}^n \times \mathbb{R}. \tag{1.5}$$

The term $\hat{H}$ on the right hand side is at this point in the discussion just some given function of P only, which does not depend on X. This change of coordinates usually fails in practice, since the PDE (1.5) does not often have a smooth solution and, even if it does, the transformation (1.3), (1.4) is not usually valid globally.

On the other hand, we can under reasonable assumptions on H build a weak solution of (1.5) for a suitably defined function $\hat{H}$. This we show in §2 below, following Lions–Papanicolaou–Varadhan [L-P-V]. The key hypothesis will be that H is $\mathbb{T}^n$-periodic in the space variable x, where $\mathbb{T}^n$ denotes the unit cube in $\mathbb{R}^n$ with opposite faces identified, and furthermore that H is $[0,1]$-periodic in the time variable t.

It turns out that given $P \in \mathbb{R}^n$ there exists a *unique* real number λ for which the cell problem

$$\begin{cases} w_t + H(P + D_x w, x, t) = \lambda & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w \;\text{ is } \mathbb{T}^{n+1}\text{–periodic} \end{cases}$$

has a solution, where $\mathbb{T}^{n+1} := \mathbb{T}^n \times [0,1]$. We define $\hat{H}(P) := \lambda$, to rewrite the foregoing as

$$\begin{cases} w_t + H(P + D_x w, x, t) = \hat{H}(P) & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w \;\text{ is } \mathbb{T}^{n+1}\text{–periodic.} \end{cases} \tag{1.6}$$

And so if we also set

$$u(P,x,t) := P \cdot x + w(P,x,t), \tag{1.7}$$

we recover (1.5).

The overall goal now is to study the *effective* or *averaged Hamiltonian* $\hat{H}$, so defined, and to try to understand how $\hat{H}$ and u encode information about the dynamics (1.1). This we at least partially accomplish by turning our attention exclusively to Hamiltonians which are uniformly convex in the momenta.

To understand the basic issues, let us for the moment change from the Hamiltonian viewpoint and so let $L = L(q,x,t)$ denote the corresponding Lagrangian. We consider curves $\mathbf{x}(\cdot)$ which are *absolutely minimizing* for the action. This means that

$$\int_0^T L(\dot{\mathbf{x}}, \mathbf{x}, t)\, dt \le \int_0^T L(\dot{\mathbf{y}}, \mathbf{y}, t)\, dt \tag{1.8}$$

for each time $T > 0$ and each Lipschitz curve $\mathbf{y}(\cdot)$ with $\mathbf{x}(0) = \mathbf{y}(0)$, $\mathbf{x}(T) = \mathbf{y}(T)$. The corresponding momentum is

$$\mathbf{p} := D_q L(\dot{\mathbf{x}}, \mathbf{x}, t),$$

and $(\mathbf{x}(\cdot), \mathbf{p}(\cdot))$ solve the system of ODE (1.1).

We propose now to rescale in time, defining

$$\mathbf{x}_\varepsilon(t) := \varepsilon \mathbf{x}\left(\frac{t}{\varepsilon}\right), \;\; \mathbf{p}_\varepsilon := \mathbf{p}\left(\frac{t}{\varepsilon}\right); \tag{1.9}$$

then

$$
\begin{cases} \dot{\mathbf{x}}_\varepsilon = D_pH\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right) \\ \dot{\mathbf{p}}_\varepsilon = -\frac{1}{\varepsilon}D_xH\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right). \end{cases} \tag{1.10}
$$

We study the oscillations in $\mathbf{p}_\varepsilon(\cdot)$ and $\mathbf{x}_\varepsilon(\cdot)$ (mod $\mathbb{T}^n$) as $\varepsilon \to 0$ and demonstrate in §2,3 how these are governed by certain Young measures, whose structure in turn can be described, at least in part, in terms of $\hat{H}$ and u. Note that we have rescaled so that (1.10) entails fast, periodic changes in the time variable.

1.2. Slow variations in time.

It turns out to be interesting as well to consider slow variations in time. For this, take $\varepsilon > 0$ and assume the curve $\mathbf{x}^\varepsilon(\cdot)$ is an *ε-absolute minimizer* of the action, meaning

$$
\int_0^T L(\dot{\mathbf{x}}^\varepsilon, \mathbf{x}^\varepsilon, \varepsilon t)\, dt \le \int_0^T L(\dot{\mathbf{y}}, \mathbf{y}, \varepsilon t)\, dt \tag{1.11}
$$

for each $T > 0$ and each curve $\mathbf{y}(\cdot)$ with $\mathbf{x}^\varepsilon(0) = \mathbf{y}(0)$, $\mathbf{x}^\varepsilon(T) = y(T)$.

Define $\mathbf{p}^\varepsilon := D_qL(\dot{\mathbf{x}}^\varepsilon, \mathbf{x}^\varepsilon, \varepsilon t)$, and rescale:

$$
\mathbf{x}_\varepsilon(t) := \varepsilon \mathbf{x}^\varepsilon\left(\frac{t}{\varepsilon}\right), \ \mathbf{p}_\varepsilon(t) := \mathbf{p}^\varepsilon\left(\frac{t}{\varepsilon}\right). \tag{1.12}
$$

Then

$$
\begin{cases} \dot{\mathbf{x}}_\varepsilon = D_pH\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right) \\ \dot{\mathbf{p}}_\varepsilon = -\frac{1}{\varepsilon}D_xH\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right). \end{cases} \tag{1.13}
$$

Unlike the system (1.10), the time variable is now changing on only an $O(1)$–scale.

The foregoing is the setting for the classical physical principle of the *adiabatic invariance of the action.* (See for instance Arnold [A] or Goldstein [Gd]; Crawford [Cr] provides several worked examples.) The idea is to switch variables according to (1.3), (1.4), where we now *redefine*

$$
u(P, x, t) := P \cdot x + v(P, x, t); \tag{1.14}
$$

and *for each time* t, v solves the cell problem

$$
\begin{cases} H(P + D_xv, x, t) = \bar{H}(P, t) \quad \text{in } \mathbb{R}^n \\ x \mapsto v \ \text{ is } \ \mathbb{T}^n\text{-periodic.} \end{cases} \tag{1.15}
$$

The notation $\bar{H}(P,t)$ thus means the effective Hamiltonian had by "averaging" H in the spatial variables only, t held fixed. This is an entirely different procedure from averaging in space and time together, as in (1.6).

We now introduce the very strong assumption that we can invert (1.3), (1.4), (1.14), to solve for X, P as smooth functions of x, p, t. We can then define the *action-angle variables*

$$\begin{cases} \mathbf{P}_\varepsilon(t) := P\left(\mathbf{p}_\varepsilon(t), \frac{\mathbf{x}_\varepsilon(t)}{\varepsilon}, t\right) \\ \mathbf{X}_\varepsilon(t) := X\left(\mathbf{p}_\varepsilon(t), \frac{\mathbf{x}_\varepsilon(t)}{\varepsilon}, t\right). \end{cases} \tag{1.16}$$

The principle of adiabatic invariance of the action suggests that

$$\mathbf{P}_\varepsilon(\cdot) \to P, \tag{1.16}$$

uniformly on compact subsets of $[0,\infty)$, for some constant vector $P \in \mathbb{R}^n$. Now this is certainly not really true in general: We usually cannot even define $\mathbf{P}_\varepsilon(\cdot)$, since we cannot invert (1.4) to solve for $P = P(p,x,t)$. We will however see in §4 that we can in fact associate a vector P with the rescaled dynamics (1.13), thereby establishing at least a weak form of the adiabatic invariance in this setting. The key hypothesis, we repeat, is that the curve $\mathbf{x}(\cdot)$ is a minimizer of the action.

Remark. We can further interpret all this in light of homogenization theory for nonlinear first–order PDE, as in Lions–Papanicolaou–Varadhan [L-P-V]. If we look at this initial-value problem for the Hamilton–Jacobi PDE with rapidly oscillating coefficients

$$\begin{cases} u^\varepsilon_t + H\left(D_x u^\varepsilon, \frac{x}{\varepsilon}, t\right) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u^\varepsilon = g & \text{on } \mathbb{R}^n \times \{t=0\} \end{cases} \tag{1.17}$$

and let $\varepsilon \to 0$, the solution u^ε converges to the solution u of

$$\begin{cases} u_t + \bar{H}(D_x u, t) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t=0\}. \end{cases} \tag{1.18}$$

Now the function u^ε can be computed in terms of characteristic ODE, which for the case at hand are precisely (1.13). We can consequently informally state that

> *the homogenization of* (1.17) *into* (1.18) *is a consequence of the adiabatic invariance of the action for minimizing trajectories,*

since the characteristics of the viscosity solution of (1.17) are in fact minimizers, at least on an $O(1)$-time scale. A similar comment holds for PDE of the form $u^\varepsilon_t + H\left(D_x u^\varepsilon, \frac{x}{\varepsilon}, x, t\right)$. □

1.3. Geometric correction to the angle.

The asymptotics as $\varepsilon \to 0$ of the angle variables $\mathbf{X}_\varepsilon(\cdot)$ are more subtle. A calculation of Hannay [H] and Berry [B] suggests that

$$\dot{\mathbf{X}}_\varepsilon(t) - \frac{1}{\varepsilon} D_P \bar{H}(\mathbf{P}_\varepsilon(t), t) \to DK(P), \tag{1.19}$$

with a "corrector" $K(P)$ computed in terms of v: see §5.3.

We provide a formal interpretation of K in light of the effective Hamiltonians introduced before. For this we first modify the space-time averaging problem (1.6), now to find for each $\delta > 0$ a unique real number $\lambda = \hat{H}(P, \delta)$ for which the cell problem

$$\begin{cases} \delta w_t^\delta + H(P + D_x w^\delta, x, t) = \hat{H}(P, \delta) & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w \ \text{ is } \mathbb{T}^{n+1}\text{–periodic} \end{cases} \tag{1.20}$$

has a viscosity solution. Then, as we will show in §6,

$$\lim_{\delta \to 0} \hat{H}(P, \delta) = \int_0^1 \bar{H}(P, t)\, dt. \tag{1.21}$$

We furthermore establish some inequalities for difference quotients of $\hat{H}$, which in turn suggest the formula

$$\frac{\partial \hat{H}}{\partial \delta}(P, 0) = K(P). \tag{1.22}$$

In this formal sense the Berry–Hannay phase factor is "hidden" within the effective Hamiltonian $\hat{H}$.

2. Averaging in space and time.

In this section we discuss the construction and properties of the effective Hamiltonian $\hat{H}$. Our results are mostly not new and involve ideas found for instance in Fathi [F1]. However we have found no precise reference in the literature.

2.1. The Hamiltonian and Lagrangian.

Hypotheses on the Hamiltonian. Suppose hereafter that the smooth function $H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $H = H(p,x,t)$, satisfies these conditions:

(i) *periodicity in x and t*:

$$\left\{ \begin{array}{l} \text{For each } p \in \mathbb{R}^n, \text{ the mapping} \\ (x,t) \mapsto H(p,x,t) \text{ is } \mathbb{T}^{n+1}\text{-periodic.} \end{array} \right. \tag{2.1}$$

(ii) *strict convexity*:

$$\left\{ \begin{array}{l} \text{There exist constants } \Gamma, \gamma > 0 \text{ such that} \\ \gamma|\xi|^2 \le \sum_{i,j=1}^n \frac{\partial^2 H}{\partial p_i \partial p_j}(p,x,t)\xi_i\xi_j \le \Gamma|\xi|^2 \\ \text{for each } p, x, \xi \in \mathbb{R}^n,\ t \in \mathbb{R}. \end{array} \right. \tag{2.2}$$

(iii) *gradient bound in x*:

$$\left\{ \begin{array}{l} \text{There exists a constant } C \text{ such that} \\ \qquad |D_x H(p,x,t)| \le C(1+|p|) \\ \text{for all } p, x \in \mathbb{R}^n,\ t \in \mathbb{R}. \end{array} \right. \tag{2.3}$$

The Lagrangian. We define the corresponding *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $L = L(q,x,t)$, by

$$L(q,x,t) := \sup_p (p \cdot q - H(p,x,t)).$$

Then

$$(x,t) \mapsto L(q,x,t) \text{ is } \mathbb{T}^{n+1}\text{-periodic,} \tag{2.4}$$

and

$$\left\{ \begin{array}{l} \text{there exist constants } \Gamma, \gamma > 0 \text{ such that} \\ \gamma|\xi|^2 \le \sum_{i,j=1}^n \frac{\partial^2 L}{\partial q_i \partial q_j}(q,x,t)\xi_i\xi_j \le \Gamma|\xi|^2 \\ \text{for each } q, x, \xi \in \mathbb{R}^n,\ t \in \mathbb{R}. \end{array} \right. \tag{2.5}$$

2.2. Averaging in x and t.

We prove first that the cell problem (1.6) is solvable, in an appropriately weak sense.

Theorem 2.1. (i) *For each* $P \in \mathbb{R}^n$ *there exists a unique real number, denoted* $\hat{H}(P)$, *such that the* cell problem

$$\begin{cases} w_t + H(P + D_x w, x, t) = \hat{H}(P) & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w \quad \text{is } \mathbb{T}^{n+1}\text{–periodic} \end{cases} \tag{2.6}$$

has a Lipschitz continuous viscosity solution w.

(ii) *In addition, there exists a constant* α *such that*

$$D^2 w \le \alpha I \quad \text{in } \mathbb{R}^n \times \mathbb{R} \tag{2.7}$$

in the distribution sense.

Assertion (2.7) means that w is semiconcave jointly in the variables x and t. We call $\hat{H} : \mathbb{R}^n \to \mathbb{R}$ so defined the *effective* or *averaged Hamiltonian.*

Remarks Observe carefully that $\hat{H}$ results from our "averaging" H with respect to both the space variables x and the time variable t. As noted before, we write

$$\bar{H}(\cdot, t)$$

to denote the averaged Hamiltonian with respect to the space variable x only, for each fixed time t. One of our objectives is understanding the relationships between $\hat{H}$ and $\{\bar{H}(\cdot, t)\}_{0 \le t \le 1}$. □

Proof. 1. *Existence.* Let $X = C(\mathbb{T}^n)$ denote the space of $\mathbb{T}^n$-periodic, continuous functions, with the max-norm. We hereafter fix $P \in \mathbb{R}^n$. Given $g \in X$, we consider then the initial–value problem

$$\begin{cases} v_t + H(P + D_x v, x, t) = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ \qquad\qquad\qquad\qquad v = g & \text{on } \mathbb{R}^n \times \{t = 0\}, \end{cases} \tag{2.8}$$

which has the unique, $\mathbb{T}^n$-periodic viscosity solution

$$v(x,t) = \inf \left\{ \int_0^t L(\dot{\mathbf{x}}, \mathbf{x}, s) - P \cdot \dot{\mathbf{x}} \, ds + g(\mathbf{x}(0)) \right\}, \tag{2.9}$$

the infimum taken over all Lipschitz continuous functions $\mathbf{x} : [0,t] \to \mathbb{R}^n$ with $\mathbf{x}(t) = x$.

Define then the nonlinear operator $T : X \to X$ by

$$T[g](x) := v(x,1) - \min_{\mathbb{T}^n} v(\cdot, 1) \qquad (x \in \mathbb{T}^n). \tag{2.10}$$

2. We next develop some estimates for $v(\cdot,1)$, in terms of $||g||_{L^\infty}$. First of all, we can take $\mathbf{x}(t) \equiv x$ in (2.9), to deduce

$$v(x,1) \leq \int_0^1 L(0,x,s)\,ds + g(x) \tag{2.11}$$

for $x \in \mathbb{T}^n$. On the other hand, since

$$L(\dot{\mathbf{x}},\mathbf{x},s) \geq L(0,\mathbf{x},s) + D_q L(0,\mathbf{x},s)\cdot\dot{\mathbf{x}} + \frac{\gamma}{2}|\dot{\mathbf{x}}|^2,$$

we have

$$v(x,1) \geq -C - ||g||_{L^\infty}.$$

This and (2.11) imply

$$|v(x,1)| \leq C(1+||g||_{L^\infty}) \quad \text{for all } x \in \mathbb{T}^n. \tag{2.12}$$

3. Next we show $v(\cdot,1)$ is Lipschitz continuous. According to (2.9), we have

$$v(x,1) = \int_0^1 L(\dot{\mathbf{x}},\mathbf{x},s) - P\cdot\dot{\mathbf{x}}\,ds + g(\mathbf{x}(0)), \tag{2.13}$$

for some Lipschitz continuous curve $\mathbf{x} : [0,1] \to \mathbb{R}^n$ with $\mathbf{x}(1) = x$. We estimate the essential supremum of $|\dot{\mathbf{x}}|$ by first noting from the foregoing that $\int_0^1 L(\dot{\mathbf{x}},\mathbf{x},s) - P\cdot\dot{\mathbf{x}}\,ds$ is bounded in terms of $||g||_{L^\infty}$. Since L grows quadratically in the variable q, we deduce that $|\dot{\mathbf{x}}(t_0)| \leq M$ at some time $0 < t_0 < 1$, where M denotes a constant computable in terms of known quantities. Therefore the momentum

$$\mathbf{p}(t_0) := D_q L(\dot{\mathbf{x}}(t_0),\mathbf{x}(t_0),t_0)$$

is bounded. The Hamiltonian equation for all $p, x \in \mathbb{R}^n$, $t \in \mathbb{R}$. estimate (2.3), and Gronwall's inequality now imply that $\mathbf{p}$, and thus $|\dot{\mathbf{x}}|$, are essentially bounded on $[0,1]$, with estimates depending only upon $||g||_{L^\infty}$.

Now for any other point y, we have

$$v(y,1) = \inf\left\{\int_0^1 L(\dot{\mathbf{y}},\mathbf{y},s) - P\cdot\dot{\mathbf{y}}\,ds + g(\mathbf{y}(0))\right\}, \tag{2.14}$$

the infimum taken over Lipschitz curves $\mathbf{y}(\cdot)$ with $\mathbf{y}(1) = y$. Set

$$\mathbf{y}(s) := \mathbf{x}(s) + s(y-x) \qquad (0 \leq s \leq 1)$$

in (2.14), and write $w := y - x$. Then

$$\begin{aligned} v(y,1) - v(x,1) &\leq \int_0^1 L(\dot{\mathbf{x}}+w,\mathbf{x}+sw,s) - L(\dot{\mathbf{x}},\mathbf{x},s) + P\cdot w\,ds \\ &\leq C|w| = C|x-y|. \end{aligned} \tag{2.15}$$

The same estimate holds with the points x and y interchanged. Therefore

$$|v(x,1) - v(y,1)| \leq C|x-y| \quad \text{for all } x,y \in \mathbb{T}^n. \tag{2.16}$$

for some constant C depending on $||g||_{L^\infty}$.

4. Next we show that the operator T defined in Step 1 has a fixed point. First of all, standard comparison results for viscosity solutions show that the time-one map $g \mapsto w(\cdot,1)$ is a contraction in the max-norm. Consequently

$$\|T[g] - T[\hat{g}]\| \leq 2\|g - \hat{g}\|.$$

Therefore $T : X \to X$ is continuous. In light of estimate (2.16), T is a compact mapping as well.

Finally, fix $0 \leq \sigma \leq 1$ and suppose g^σ is a fixed point of the operator $\sigma T[\cdot]$; that is,

$$\sigma \left(v^\sigma(x,1) - \min_{\mathbb{T}^n} v^\sigma(\cdot,1) \right) = g^\sigma(x) \qquad (x \in \mathbb{T}^n), \tag{2.17}$$

where

$$\begin{cases} v_t^\sigma + H(P + D_x v^\sigma, x, t) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ v^\sigma = g^\sigma & \text{on } \mathbb{R}^n \times \{t=0\}. \end{cases} \tag{2.18}$$

According to (2.17),

$$g^\sigma \geq 0, \quad g^\sigma(x_0) = 0 \qquad \text{for some point } x_0 \in \mathbb{T}^n. \tag{2.19}$$

The solution of (2.18) is

$$v^\sigma(x,t) = \inf \left\{ \int_0^t L(\dot{\mathbf{x}}, \mathbf{x}, s) - P \cdot \dot{\mathbf{x}}\, ds + g^\sigma(\mathbf{x}(0)) \right\},$$

the infimum taken over all Lipschitz continuous functions $\mathbf{x} : [0,t] \to \mathbb{R}^n$ with $\mathbf{x}(t) = x$. We take $\mathbf{x}(s) := \left(1 - \frac{s}{t}\right) x_0 + \frac{s}{t} x$ for $0 \leq s \leq t$, to deduce

$$\begin{aligned} v^\sigma(x,t) &\leq \int_0^t L\left(\frac{x - x_0}{t}, (1 - \frac{s}{t})x_0 + \frac{s}{t}x, s \right) - P \cdot \frac{(x-x_0)}{t}\, ds \\ &\leq C\left(1 + \frac{|x - x_0|^2}{t} \right) \end{aligned} \tag{2.20}$$

for $x \in \mathbb{T}^n$, $0 \leq t \leq 1$. On the other hand, since $g^\sigma \geq 0$ we see that

$$v^\sigma(x,t) \geq \inf \left\{ \int_0^t L(\dot{\mathbf{x}}, \mathbf{x}, s) - P \cdot \dot{\mathbf{x}}\, ds \right\}.$$

Recalling the strict convexity of L, we deduce

$$v^\sigma(x,t) \geq -Ct$$

for $x \in \mathbb{T}^n$, $0 \leq t \leq 1$. Combining this with (2.20) yields

$$|v^\sigma(x,t)| \leq \frac{C}{t} \qquad (x \in \mathbb{R}^n,\ 0 \leq t \leq 1). \tag{2.21}$$

Consequently (2.21) in turn implies

$$\max_{\mathbb{T}^n} |g^\sigma| \leq C \qquad (0 \leq \sigma \leq 1). \tag{2.22}$$

5. In summary, the operator $T : X \to X$ is continuous, compact, and we have the *a priori* estimate (2.22) for any fixed point of $\sigma T[\cdot]$, where $0 \leq \sigma \leq 1$. Consequently Schaeffer's Theorem (cf. [E2]) implies that T possesses at least one fixed point:

$$T[g] = g.$$

This means that $v(x,1) - \min_{\mathbb{T}^n} v(\cdot,1) = g(x)$ for each point $x \in \mathbb{T}^n$, where

$$\begin{cases} v_t + H(P + D_x v, x, t) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ v = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

Finally define

$$\begin{cases} \hat{H}(P) := -\min_{\mathbb{T}^n} v(\cdot,1), \\ w(x,t) := v(x,t) + t\hat{H}(P). \end{cases}$$

Then

$$w_t + H(P + D_x w, x, t) = \hat{H}(P) \quad \text{in } \mathbb{R}^n \times (0,\infty)$$

and $(x,t) \mapsto w$ is $\mathbb{T}^{n+1}$-periodic.

6. *Uniqueness.* Uniqueness of the constant $\lambda = \hat{H}(P)$ in (2.7) follows from an argument due to Lions–Papanicolaou–Varadhan[L-P-V]. Suppose

$$\begin{cases} w_t + H(P + D_x w, x, t) = \lambda & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w \ \text{ is } \mathbb{T}^{n+1}\text{–periodic} \end{cases}$$

and

$$\begin{cases} v_t + H(P + D_x v, x, t) = \mu & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto v \ \text{ is } \mathbb{T}^{n+1}\text{–periodic.} \end{cases}$$

We may assume $\lambda > \mu$, and also—upon adding if necessary a constant to v—that

$$v > w \quad \text{in } \mathbb{R}^n \times \mathbb{R}. \tag{2.23}$$

But for sufficiently small $\epsilon > 0$,

$$w_t + H(P + D_x w, x, t) + \epsilon w > v_t + H(P + D_x v, x, t) + \epsilon v \quad \text{in } \mathbb{R}^n \times \mathbb{R},$$

in the sense of viscosity solutions. The comparison principle for viscosity solutions however then implies

$$w \geq v \quad \text{in } \mathbb{R}^n \times \mathbb{R},$$

in contradiction to (2.23).

7. *Semiconcavity in x and t.* Finally we show semiconcavity in space and time. By periodicity, it suffices to consider the time $t = 1$. Recall

$$v(x,1) = \int_0^1 L(\dot{\mathbf{x}}, \mathbf{x}, s) - P \cdot \dot{\mathbf{x}}\, ds + g(\mathbf{x}(0)),$$

for some Lipschitz curve $\mathbf{x}(\cdot)$ with $\mathbf{x}(1) = x$. Take any vector z and real number h, with $|h| < 1$. Then

$$v(x+z, 1+h) \leq \int_0^{1+h} L(\dot{\mathbf{y}}, \mathbf{y}, s) - P \cdot \dot{\mathbf{y}}\, ds + g(\mathbf{x}(0)),$$

for $\mathbf{y}(s) := \mathbf{x}(\frac{s}{1+h}) + \frac{s}{1+h} z$. Similarly,

$$v(x-z, 1-h) \leq \int_0^{1-h} L(\dot{\mathbf{z}}, \mathbf{z}, s) - P \cdot \dot{\mathbf{z}}\, ds + g(\mathbf{x}(0)),$$

for $\mathbf{z}(s) := \mathbf{x}(\frac{s}{1-h}) - \frac{s}{1-h} z$. Then

$$\begin{aligned}
v(x+z,&1+h) - 2v(x,1) + v(x-z,1-h) \\
&\leq \int_0^{1+h} L(\frac{1}{1+h}\dot{\mathbf{x}}(\frac{s}{1+h}) + \frac{1}{1+h}z, \mathbf{x}(\frac{s}{1+h}) + \frac{s}{1+h}z, s)\, ds \\
&\quad - 2\int_0^1 L(\dot{\mathbf{x}}(s), \mathbf{x}(s), s)\, ds \\
&\quad + \int_0^{1-h} L(\frac{1}{1-h}\dot{\mathbf{x}}(\frac{s}{1-h}) - \frac{1}{1-h}z, \mathbf{x}(\frac{s}{1-h}) - \frac{s}{1-h}z, s)\, ds \\
&= \int_0^1 (1+h)L(\frac{1}{1+h}\dot{\mathbf{x}}(s) + \frac{1}{1+h}z, \mathbf{x}(s) + sz, (1+h)s) \\
&\quad - 2L(\dot{\mathbf{x}}(s), \mathbf{x}(s), s) + (1-h)L(\frac{1}{1-h}\dot{\mathbf{x}}(s) - \frac{1}{1-h}z, \mathbf{x}(s) - sz, (1-h)s)\, ds.
\end{aligned}$$

This last expression equals

$$
\begin{aligned}
&\int_0^1 L(\dot{\mathbf{x}}(s) - \frac{h}{1+h}\dot{\mathbf{x}}(s) + \frac{1}{1+h}z, \mathbf{x}(s) + sz, (1+h)s) \\
&\qquad - 2L(\dot{\mathbf{x}}(s), \mathbf{x}(s), s) + L(\dot{\mathbf{x}}(s) + \frac{h}{1-h}\dot{\mathbf{x}}(s) - \frac{1}{1-h}z, \mathbf{x}(s) - sz, (1-h)s)\, ds \\
&+ h\int_0^1 L(\dot{\mathbf{x}}(s) - \frac{h}{1+h}\dot{\mathbf{x}}(s) + \frac{1}{1+h}z, \mathbf{x}(s) + sz, (1+h)s) \\
&\qquad - L(\dot{\mathbf{x}}(s) + \frac{h}{1-h}\dot{\mathbf{x}}(s) - \frac{1}{1-h}z, \mathbf{x}(s) - sz, (1-h)s)\, ds \\
&\leq C(|z|^2 + h^2).
\end{aligned}
$$

Thus v, and so also the function w, are semiconcave in x, t. □

Remarks. Much of this proof is based upon ideas in Fathi [F-1], whom we thank for help. Our hypothesis (2.3), rewritten in terms of the Lagrangian, says that

$$|D_x L(q, x, t)| \leq C(1 + |D_q L(q, x, t)|)$$

for all $q, x \in \mathbb{R}^n$, $t \in \mathbb{R}$. As Fathi pointed out to us, this is a standard condition to ensure that the minimization problems (e.g. (2.9)) have Lipschitz continuous minimizers. We could replace (2.3) by the assumption that

$$|H_t(p, x, t)| \leq C(1 + |p|^2) \tag{2.27}$$

for all $p, x \in \mathbb{R}^n$, $t \in \mathbb{R}$. Since

$$\frac{d}{dt}H(\mathbf{p}, \mathbf{x}, t) = H_t(\mathbf{p}, \mathbf{x}, t),$$

estimate (2.27), the strict convexity of H, and Gronwall's inequality again provide Lipschitz bounds.

See Jauslin–Kreiss–Moser [J-K-M] and Weinan E [EW2] for a special case of $n = 1$. The approach in these papers is to consider the corresponding conservation law. The arguments above are also strongly related to those in the general theory of nonlinear "additive eigenvalue" problems; see for instance Nussbaum [N], Chou–Duffin [C-D], etc. Similar ideas are in Concordel [C1]. □

We define then the *effective Lagrangian*

$$\hat{L}(Q) := \sup_P (P \cdot Q - \hat{H}(P)) \tag{2.28}$$

for $Q \in \mathbb{R}^n$. The functions $\hat{H}$ and $\hat{L}$ are both convex and finite-valued, with superlinear growth:

$$\lim_{|P| \to \infty} \frac{\hat{H}(P)}{|P|} = +\infty, \quad \lim_{|Q| \to \infty} \frac{\hat{L}(Q)}{|Q|} = +\infty.$$

The convexity of $\hat{H}$ can be shown using for instance the method in [E3].

3. Fast variations in time.

This section, which extends slightly some methods of [E-G1], uses our weak solution of the time-dependent corrector problem (2.7) to study the structure of a measure on phase space that records the long time asymptotics of minimizing trajectories.

Definition. A Lipschitz continuous curve $\mathbf{x} : [0, \infty) \to \mathbb{R}^n$ is a (one-sided) *absolute minimizer* if

$$\int_0^T L(\dot{\mathbf{x}}, \mathbf{x}, t)\, dt \le \int_0^T L(\dot{\mathbf{y}}, \mathbf{y}, t)\, dt \tag{3.1}$$

for each time $T > 0$ and each Lipschitz curve $\mathbf{y}(\cdot)$ such that $\mathbf{x}(0) = \mathbf{y}(0),\ \mathbf{x}(T) = \mathbf{y}(T)$.

Remark. Note that we are simply assuming $\mathbf{x}(\cdot)$ to be Lipschitz continuous globally in time. If the Hamiltonian does not depend upon t, the invariance of the Hamiltonian implies a global bound on $|\dot{\mathbf{x}}|$. □

Given an absolute minimizer, we set

$$\mathbf{p}(t) := D_q L(\dot{\mathbf{x}}(t), \mathbf{x}(t), t); \tag{3.2}$$

so that

$$\begin{cases} \dot{\mathbf{x}} = D_p H(\mathbf{p}, \mathbf{x}, t) \\ \dot{\mathbf{p}} = -D_x H(\mathbf{p}, \mathbf{x}, t) \end{cases} \tag{3.3}$$

for $t \ge 0$. Next rescale:

$$\mathbf{x}_\varepsilon(t) := \varepsilon \mathbf{x}\left(\frac{t}{\varepsilon}\right),\ \mathbf{p}_\varepsilon(t) := \mathbf{p}\left(\frac{t}{\varepsilon}\right).$$

Then

$$\begin{cases} \dot{\mathbf{x}}_\varepsilon = D_p H\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right) \\ \dot{\mathbf{p}}_\varepsilon = -\frac{1}{\varepsilon} D_x H\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right). \end{cases} \tag{3.4}$$

Next we introduce measures on phase space to record oscillations of these dynamics as $\varepsilon \to 0$:

Theorem 3.1. *There exists a sequence $\varepsilon_k \to 0$ and for a.e. $t > 0$ a Radon probability measure μ_t on $\mathbb{R}^n \times \mathbb{T}^{n+1}$ such that*

$$\Phi\left(\mathbf{p}_{\varepsilon_k}(t), \frac{\mathbf{x}_{\varepsilon_k}(t)}{\varepsilon_k}, \frac{t}{\varepsilon_k}\right) \rightharpoonup \bar{\Phi}(t) = \int_{\mathbb{R}^n} \int_{\mathbb{T}^{n+1}} \Phi(p, x, s)\, d\mu_t \tag{3.5}$$

for each bounded, continuous function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $\Phi = \Phi(p, x, s)$, for which $(x, s) \mapsto \Phi$ is $\mathbb{T}^{n+1}$–periodic.

This assertion follows as in [E-G1].

Lemma 3.2. (i) *The support of the measures* $\{\mu_t\}_{t\geq 0}$ *is bounded, uniformly in* t.

(ii) *For each* C^1 *function* Φ *as above,*

$$\int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} \Phi_s + \{H,\Phi\}\, d\mu_t = 0 \tag{3.6}$$

for a.e. $t \geq 0$, *where*

$$\{H,\Phi\} := D_pH\cdot D_x\Phi - D_xH\cdot D_p\Phi$$

is the Poisson bracket.

Proof. Statement (i) is a consequence of the bounds on $|\dot{\mathbf{x}}|$, and thus $|\mathbf{p}_{\varepsilon_k}|$. Furthermore,

$$\begin{aligned}\frac{d}{dt}\Phi\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right) &= D_p\Phi\cdot\dot{\mathbf{p}}_\varepsilon + D_x\Phi\cdot\frac{\dot{\mathbf{x}}_\varepsilon}{\varepsilon} + \frac{1}{\varepsilon}\Phi_s\\ &= \frac{1}{\varepsilon}(\Phi_s + \{H,\Phi\}).\end{aligned}$$

Take $\zeta : [0,T]\to\mathbb{R}$ to be smooth, with compact support. Then

$$\int_0^T (\Phi_s+\{H,\Phi\})\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right)\zeta\, dt = -\int_0^T \varepsilon\dot{\zeta}\Phi\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right) dt.$$

Send $\varepsilon = \varepsilon_k \to 0$. □

We may assume, passing if necessary to a further subsequence, that

$$\mathbf{x}_{\varepsilon_k} \to \mathbf{X} \tag{3.7}$$

uniformly on compact subsets of $[0,\infty)$, where $\mathbf{X}(\cdot)$ is a Lipschitz continuous curve, $\mathbf{X}(0) = 0$. Then $\dot{\mathbf{X}}(t) = \mathbf{Q}(t)$, for

$$\mathbf{Q}(t) := \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} D_pH(p,x,s)\, d\mu_t. \tag{3.8}$$

Theorem 3.3. (i) *For a.e.* $t\geq 0$ *we have*

$$\hat{L}(\mathbf{Q}(t)) = \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} L(D_pH(p,x,s),x,s)\, d\mu_t. \tag{3.9}$$

(ii) *Furthermore, there exists a vector* $P\in\mathbb{R}^n$ *such that*

$$P\in\partial\hat{L}(\mathbf{Q}(t)),\ \ \mathbf{Q}(t)\in\partial\hat{H}(P) \qquad \textit{for a.e. } t\geq 0. \tag{3.10}$$

Remarks. (i) We call P the *action vector* for the rescaled trajectories $\{\mathbf{x}_\varepsilon(\cdot)\}_{\varepsilon>0}$.

(ii) The second assertion above can be rewritten

$$\begin{cases} \dot{\mathbf{X}} \in \partial \hat{H}(\mathbf{P}) \\ \dot{\mathbf{P}} = 0 \end{cases} \quad \text{for a.e. } t \geq 0,$$

and this formulation should be compared with (1.2). □

Outline of proof. 1. Firstly, define

$$S_\varepsilon(x,y,t) := \inf \left\{ \int_0^t L\left(\dot{\mathbf{x}}, \frac{\mathbf{x}}{\varepsilon}, \frac{s}{\varepsilon}\right) ds \mid \mathbf{x}(t) = x,\ \mathbf{x}(0) = y \right\},$$

for $x, y \in \mathbb{R}^n, t > 0$. Then

$$S_\varepsilon(x,y,t) \to t\hat{L}\left(\frac{x-y}{t}\right) \quad \text{as } \varepsilon \to 0,$$

uniformly on compact subsets of $\mathbb{R}^n \times \mathbb{R}^n \times (0,\infty)$, as shown by Lions–Papanicolaou–Varadhan [L-P-V, §IV]. (See also Weinan E [EW1], Braides–Defranceschi [B-D, §16.2].)

Let $y_\varepsilon := \mathbf{x}_\varepsilon(0) = \varepsilon \mathbf{x}(0) \to 0$. Thus

$$S_{\varepsilon_k}(x, y_{\varepsilon_k}, t) \to t\hat{L}\left(\frac{x}{t}\right) \qquad (x \in \mathbb{R}^n,\ t > 0),$$

uniformly on compact subsets. But

$$S_\varepsilon(\mathbf{x}_\varepsilon(t), y_\varepsilon, t) = \int_0^t L\left(\dot{\mathbf{x}}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{s}{\varepsilon}\right) ds,$$

since the curve $\mathbf{x}_\varepsilon(\cdot)$ is an absolute minimizer. Consequently

$$S_{\varepsilon_k}(\mathbf{x}_{\varepsilon_k}(t), y_{\varepsilon_k}, t) \to t\hat{L}\left(\frac{\mathbf{X}(t)}{t}\right).$$

2. Differentiating in time, it then follows that

$$L\left(\dot{\mathbf{x}}_{\varepsilon_k}, \frac{\mathbf{x}_{\varepsilon_k}}{\varepsilon_k}, \frac{t}{\varepsilon_k}\right) \rightharpoonup \frac{d}{dt}\left(t\hat{L}\left(\frac{\mathbf{X}}{t}\right)\right),$$

the half-arrow denoting weak convergence. Now

$$\frac{d}{dt}\left(t\hat{L}\left(\frac{\mathbf{X}}{t}\right)\right) \in \hat{L}\left(\frac{\mathbf{X}}{t}\right) + \partial\hat{L}\left(\frac{\mathbf{X}}{t}\right)\left(\dot{\mathbf{X}} - \frac{\mathbf{X}}{t}\right) \leq \hat{L}(\dot{\mathbf{X}}),$$

by convexity. Since $\dot{\mathbf{x}}_\varepsilon = D_pH\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right)$, we deduce that

$$\int_{\mathbb{R}^n} \int_{\mathbb{T}^{n+1}} L(D_pH, x, s)\, d\mu_t \leq \hat{L}(\dot{\mathbf{X}}(t))$$

for a.e. $t > 0$.

On the other hand, we have

$$\int_0^T \hat{L}(\dot{\mathbf{X}})\,dt \le \liminf \int_0^T L\left(\dot{\mathbf{x}}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, \frac{t}{\varepsilon}\right)\,dt,$$

and so

$$\hat{L}(\dot{\mathbf{X}}(t)) \le \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} L(D_pH, x, s)\,d\mu_t$$

for a.e. t.

3. In particular,

$$\frac{d}{dt}\left(t\hat{L}\left(\frac{\mathbf{X}(t)}{t}\right)\right) = \hat{L}(\dot{\mathbf{X}}(t)) = \hat{L}(\mathbf{Q}(t)) \;\text{ a.e.;}$$

and consequently

$$\begin{aligned}\frac{1}{T}\int_0^T \hat{L}(\mathbf{Q}(t))\,dt &= \frac{1}{T}\int_0^T \frac{d}{dt}\left(t\hat{L}\left(\frac{\mathbf{X}(t)}{t}\right)\right)\,dt \\ &= \hat{L}\left(\frac{\mathbf{X}(T)}{T}\right) \\ &= \hat{L}\left(\frac{1}{T}\int_0^T \mathbf{Q}(t)\,dt\right).\end{aligned}$$

This identity, valid for each time $T > 0$, implies that $\{\mathbf{Q}(t)\}_{t\ge 0}$ lies a supporting domain of $\hat{L}$. This means that

$$P \in \partial\hat{L}(\mathbf{Q}(t)) \quad \text{for a.e. } t \ge 0$$

for some vector $P \in \mathbb{R}^n$. Equivalently, $\mathbf{Q}(t) \in \partial\hat{H}(P)$ for a.e. $t \ge 0$. □

Henceforth we consider any one of the measures $\mu = \mu_t$, and denote by ρ the *projection* of μ onto the x, s-variables. We also write $Q = \mathbf{Q}(t)$, and take $P \in \partial\hat{L}(Q)$, as above.

Next let $w = w(P, x, t)$ be any viscosity solution of the cell PDE

$$\text{(3.11)} \qquad \begin{cases} w_t + H(P + D_xw, x, t) = \hat{H}(P) \;\text{ in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w \;\text{ is } \mathbb{T}^{n+1}\text{–periodic,} \end{cases}$$

and, as before, set

$$\text{(3.12)} \qquad u(P, x, t) := x \cdot P + w(P, x, t).$$

Theorem 3.4. *(i) The function u is differentiable in (x,t) ρ–a.e.*

(ii) We have

$$p = D_x u(P,x,t) \qquad \mu\text{–a.e.}$$

Outline of proof. We follow [E-G1], to which the interested reader may refer for more details in the case that H, L do not depend on time.

Set $u^\varepsilon := \eta_\varepsilon * u$, where η_ε is a radial convolution kernel in the variables x,t. Since the PDE $u_t + H(D_x u, x, t) = \hat{H}(P)$ holds pointwise a.e., we have

$$\beta_\varepsilon(x,t) + u_t^\varepsilon(x,t) + H(D_x u^\varepsilon(x,t), x, t) \le \hat{H}(P) + C\varepsilon \tag{3.13}$$

for each $(x,t) \in \mathbb{T}^{n+1}$, where

$$\beta_\varepsilon(x,t) := \frac{\gamma}{2}\int_{\mathbb{R}^{n+1}} \eta_\varepsilon(x-y,t-s)|D_x u(y,s) - D_x u^\varepsilon(x,t)|^2\,dyds.$$

Using the strict convexity of H with respect to the variable p, we deduce

$$\begin{aligned}&\frac{\gamma}{2}\int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} |D_x u^\varepsilon(x,s) - p|^2\,d\mu \\ &\quad\le \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} H(D_x u^\varepsilon(x,s),x,s) - H(p,x,s) - D_pH(p,x,s)\cdot(D_x u^\varepsilon(x,s) - p)\,d\mu.\end{aligned}$$

Now $D_x u^\varepsilon = P + D_x w^\varepsilon, u_t^\varepsilon = w_t^\varepsilon$, where $w^\varepsilon = \eta_\varepsilon * w$ is periodic. Furthermore

$$\int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} w_t^\varepsilon + D_pH\cdot D_x w^\varepsilon\,d\mu = 0,$$

according to (3.6). This observation and (3.13) imply

$$\begin{aligned}&\frac{\gamma}{2}\int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} |D_x u^\varepsilon - p|^2\,d\mu + \int_{\mathbb{T}^{n+1}}\beta_\varepsilon\,d\rho \\ &\qquad\qquad \le \hat{H}(P) - \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} H + D_pH\cdot(P-p)\,d\mu + C\varepsilon.\end{aligned}$$

Observe that $\hat{L}(Q) + \hat{H}(P) = P\cdot Q$, $L(D_pH,x,s) + H(p,x,s) = D_pH\cdot p$, and recall $Q = \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} D_pH\,d\mu$. Substituting above, we find

$$\begin{aligned}&\frac{\gamma}{2}\int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} |D_x u^\varepsilon - p|^2\,d\mu + \int_{\mathbb{T}^{n+1}}\beta_\varepsilon\,d\rho \\ &\qquad\qquad \le -\hat{L}(Q) + \int_{\mathbb{R}^n}\int_{\mathbb{T}^{n+1}} L(D_pH,x,s)\,d\mu + C\varepsilon = C\varepsilon,\end{aligned}$$

according to Theorem 3.3.

Since u is semiconcave, this estimate proves that u is differentiable in x at ρ a.e. point. Now define for ρ–every point $(x,t) \in \mathbb{T}^{n+1}$

$$u_t(x,t) := \hat{H}(P) - H(D_x u(x,t), x, t). \tag{3.14}$$

We will now demonstrate that u is differentiable in t at (x,t), with derivative given above. Indeed,

$$\begin{aligned}
\int_{\mathbb{R}^{n+1}} \eta_\varepsilon(x-y,t-s)(u_t(y,s)-u_t(x,t))^2\,dyds \\
&\le \int_{\mathbb{R}^{n+1}} \eta_\varepsilon(x-y,t-s)|H(D_x u(y,s),y,s)-H(D_x u(x,t),x,t)|^2\,dyds \\
&\le C\int_{\mathbb{R}^{n+1}} \eta_\varepsilon(x-y,t-s)|D_x u(y,s)-D_x u(x,t)|^2\,dyds + C\varepsilon^2.
\end{aligned}$$

The last term in (3.14) is $o(1)$ as $\varepsilon \to 0$. But then u_t is approximately continuous at (x,t). Since u is semiconcave, this means that u is differentiable in t there. □

Remark. In general

$$\int_{\mathbb{R}^n} \int_{\mathbb{T}^{n+1}} H(p,x,s)\,d\mu \neq \hat{H}(P), \tag{3.15}$$

in contrast to the autonomous case: see [E-G1] and also Dias Carneiro [DC]. The corresponding formula with the Lagrangians is true, however: remember (3.9). □

4. Slow variations in time.

In this section we modify some of the foregoing calculations to cover the situation that the time dependence of H is slowly varying.

Definition. Fix $\varepsilon > 0$. A Lipschitz continuous curve $\mathbf{x}^\varepsilon(\cdot) : [0,\infty) \to \mathbb{R}^n$ is an *ε-absolute minimizer* provided

$$\int_0^T L(\dot{\mathbf{x}}^\varepsilon, \mathbf{x}^\varepsilon, \varepsilon t)\,dt \le \int_0^T L(\dot{\mathbf{y}}, \mathbf{y}, \varepsilon t)\,dt \tag{4.1}$$

for each $T > 0$ and each Lipschitz continuous curve $\mathbf{y}(\cdot)$ such that $\mathbf{x}^\varepsilon(0) = \mathbf{y}(0)$, $\mathbf{x}^\varepsilon(T) = \mathbf{y}(T)$.

Note carefully the slow variations in the time variables here. Given $\mathbf{x}^\varepsilon(\cdot)$ as above, we define

$$\mathbf{p}^\varepsilon(t) = D_q L(\dot{\mathbf{x}}^\varepsilon, \mathbf{x}^\varepsilon, \varepsilon t); \tag{4.2}$$

so that

$$
\left\{
\begin{aligned}
\dot{\mathbf{x}}^\varepsilon &= D_p H(\mathbf{p}^\varepsilon, \mathbf{x}^\varepsilon, \varepsilon t) \\
\dot{\mathbf{p}}^\varepsilon &= -D_x H(\mathbf{p}^\varepsilon, \mathbf{x}^\varepsilon, \varepsilon t).
\end{aligned}
\right. \tag{4.3}
$$

In this setting, the Hamiltonian changes only by an amount $O(1)$ during time periods of length $O(\frac{1}{\varepsilon})$. We therefore rescale, to effect an $O(1)$–change in the Hamiltonian on an $O(1)$–time scale:

$$
\mathbf{x}_\varepsilon(t) := \varepsilon \mathbf{x}^\varepsilon\left(\frac{t}{\varepsilon}\right), \ \mathbf{p}_\varepsilon(t) = \mathbf{p}^\varepsilon\left(\frac{t}{\varepsilon}\right).
$$

Then

$$
\left\{
\begin{aligned}
\dot{\mathbf{x}}_\varepsilon &= D_p H\left(\mathbf{p}_\varepsilon, \tfrac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right) \\
\dot{\mathbf{p}}_\varepsilon &= -\tfrac{1}{\varepsilon} D_x H\left(\mathbf{p}_\varepsilon, \tfrac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right).
\end{aligned}
\right. \tag{4.4}
$$

Remark. Since $\frac{d}{dt} H\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right) = H_t\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right)$, we have the bounds

$$
\max_{0 \le t \le T} |\mathbf{p}_\varepsilon| \le C \quad \text{for each time } T > 0.
$$

□

Theorem 4.1. *There exists a sequence $\varepsilon_k \to 0$ and for a.e. $t > 0$ a Radon probability measure ν_t on $\mathbb{R}^n \times \mathbb{T}^n$ such that*

$$
\Phi\left(\mathbf{p}_{\varepsilon_k}(t), \frac{\mathbf{x}_\varepsilon(t)}{\varepsilon_k}, t\right) \rightharpoonup \bar{\Phi}(t) = \int_{\mathbb{R}^n}\int_{\mathbb{T}^n} \Phi(p, x, t)\, d\nu_t \tag{4.5}
$$

for each bounded continuous function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $\Phi = \Phi(p, x, t)$, for which $x \mapsto \Phi$ is $\mathbb{T}^n$-periodic.

Lemma 4.2. (i). *The support of the measures $\{\nu_t\}_{t\ge 0}$ is bounded, uniformly in t.*

(ii). *For each C^1 function Φ as above*

$$
\int_{\mathbb{R}^n}\int_{\mathbb{T}^n} \{H, \Phi\}\, d\nu_t = 0 \tag{4.6}
$$

for a.e. $t \ge 0$.

The proof is similar to that of Lemma 3.2 before.

We may assume

$$
\mathbf{x}_{\varepsilon_k} \to \mathbf{X} \tag{4.7}
$$

uniformly on compact subsets of $[0, \infty)$. Consequently $\dot{\mathbf{X}}(t) = \mathbf{Q}(t)$, for

$$
\mathbf{Q}(t) := \int_{\mathbb{R}^n}\int_{\mathbb{T}^n} D_p H(p, x, t)\, d\nu_t. \tag{4.8}
$$

Theorem 4.3. (i) *We have*

$$\bar{L}(\mathbf{Q}(t), t) = \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} L(D_p H, x, t)\, d\nu_t. \tag{4.9}$$

(ii) *In addition, there exists $P \in \mathbb{R}^n$ such that*

$$P \in \partial \bar{L}(\mathbf{Q}(t), t), \;\; \mathbf{Q}(t) \in \partial \bar{H}(P, t) \quad \textit{for a.e. } t. \tag{4.10}$$

Remarks. (i) P is again an action vector, and the nontrivial point is that P does not depend on t. We will in §5.2 below interpret assertion (ii) as a weak formulation of the physical principle of adiabatic invariance of the action.

(ii) The second statement above can be rewritten

$$\begin{cases} \dot{\mathbf{X}} \in \partial \bar{H}(\mathbf{P}, t) \\ \dot{\mathbf{P}} = 0 \end{cases} \quad \text{for a.e. } t \geq 0.$$

Again, compare this with (1.2). □

Outline of proof. 1. Write

$$S_\varepsilon(x, y, t) := \inf \left\{ \int_0^t L\left(\dot{\mathbf{x}}, \frac{\mathbf{x}}{\varepsilon}, s\right) ds \mid \mathbf{x}(t) = x, \;\mathbf{x}(0) = y \right\},$$

for $x, y \in \mathbb{R}^n, t > 0$. Then

$$S_\varepsilon(x, y, t) \to \min \left\{ \int_0^t \bar{L}(\dot{\mathbf{Y}}, s)\, ds \mid \mathbf{Y}(t) = x, \;\mathbf{Y}(0) = y \right\} \quad \text{as } \varepsilon \to 0,$$

uniformly on compact subsets of $\mathbb{R}^n \times \mathbb{R}^n \times (0, \infty)$. This follows as in Lions–Papanicolaou–Varadhan [L-P-V, §IV], Weinan E [EW1], etc.

Let $y_\varepsilon := \mathbf{x}_\varepsilon(0) = \varepsilon \mathbf{x}(0) \to 0$. Then

$$S_\varepsilon(\mathbf{x}_\varepsilon(t), y_\varepsilon, t) = \int_0^t L\left(\dot{\mathbf{x}}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, s\right) ds,$$

since the curve $\mathbf{x}_\varepsilon(\cdot)$ is an absolute minimizer. Consequently

$$S_{\varepsilon_k}(\mathbf{x}_{\varepsilon_k}(t), y_{\varepsilon_k}, t) \to \min \left\{ \int_0^t \bar{L}(\dot{\mathbf{Y}}, s)\, ds \mid \mathbf{Y}(t) = \mathbf{X}(t), \;\mathbf{Y}(0) = 0 \right\}$$

for each time $t \geq 0$.

2. The term on the left hand side above is less than or equal to $\int_0^t \bar{L}(\dot{\mathbf{X}}, s)\, ds$. Consequently the weak limit of

$$L\left(\dot{\mathbf{x}}_{\varepsilon_k}, \frac{\mathbf{x}_{\varepsilon_k}}{\varepsilon_k}, t\right) - \bar{L}(\dot{\mathbf{X}}, t)$$

is less than or equal to zero. It follows that

$$\int_{\mathbb{R}^n}\int_{\mathbb{T}^n} L(D_pH, x, t)\, d\nu_t \leq \bar{L}(\dot{\mathbf{X}}(t), t)$$

for a.e. $t > 0$.

On the other hand, we have

$$\int_0^T \bar{L}(\dot{\mathbf{X}}, t)\, dt \leq \liminf_{\varepsilon\to 0} \int_0^T L\left(\dot{\mathbf{x}}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right)\, dt,$$

and so

$$\bar{L}(\dot{\mathbf{X}}(t), t) \leq \int_{\mathbb{R}^n}\int_{\mathbb{T}^n} L(D_pH, x, t)\, d\nu_t$$

for a.e. t.

3. In particular, for each time $T > 0$

$$\int_0^T \bar{L}(\dot{\mathbf{X}}, t)\, dt \leq \int_0^T \bar{L}(\dot{\mathbf{Y}}, t)\, dt,$$

for each Lipschitz continuous curve $\mathbf{Y}$ satisfying $\mathbf{Y}(0) = \mathbf{X}(0) = 0$, $\mathbf{Y}(T) = \mathbf{X}(T)$.

It follows from the Pontryagin Maximum Principle (see, for instance, Clarke [Cl, page 169]) that there exists a Lipschitz continous function $\mathbf{P}(\cdot)$ such that

$$\begin{cases} \dot{\mathbf{X}} \in \partial\bar{H}(\mathbf{P}, t) \\ \dot{\mathbf{P}} = 0 \end{cases} \quad \text{for a.e. } t \geq 0.$$

But then $\mathbf{P}(\cdot) \equiv P$ for some $P \in \mathbb{R}^n$, with

$$\mathbf{Q}(t) = \dot{\mathbf{X}}(t) \in \partial\bar{H}(P, t) \quad \text{for a.e. } t \geq 0.$$

Equivalently, $P \in \partial\bar{L}(\mathbf{Q}(t), t)$ for a.e. $t \geq 0$. □

5. Calculations for integrable systems.

The remainder of our paper we largely devote to some interesting heuristic calculations that indicate connections between our $\hat{H}$ and $\{\bar{H}(\cdot, t)\}_{0\leq t\leq 1}$ and certain averaging effects.

We begin by providing here, for the reader's convenience and for later reference, some formal calculations for integrable Hamiltonians. We henceforth loosely interpret "integrable" to mean those with Hamiltonians for which the cell problem

$$\begin{cases} H(P + D_xv, x, t) = \bar{H}(P, t) & \text{in } \mathbb{R}^n \\ x \mapsto v \ \text{ is } \mathbb{T}^n\text{-periodic} \end{cases} \tag{5.1}$$

has a smooth solution $v = v(P, x, t)$ for each time $0 \leq t \leq 1$, and for which the relations

$$
\text{(5.2)} \qquad \begin{cases} p = D_x u(P, x, t) \\ X = D_P u(P, x, t) \end{cases}
$$

are smoothly, globally invertible, to provide us the change of variables formulas

$$
\text{(5.3)} \qquad \begin{cases} X = X(p, x, t), \;\; P = P(p, x, t) \\ x = x(P, X, t), \;\; p = p(P, X, t). \end{cases}
$$

We hereafter write

$$
u = P \cdot x + v.
$$

Note carefully that this differs from the earlier definition (3.12).

In particular, if we were to take the autonomous Hamiltonian $H^0 := H(\cdot, \cdot, t_0)$ (for some fixed time t_0) and considered the rescaled dynamics governed by H^0:

$$
\begin{cases} \dot{\mathbf{x}}_\varepsilon = D_p H^0 \left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon} \right) \\ \dot{\mathbf{p}}_\varepsilon = -D_x H^0 \left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon} \right), \end{cases}
$$

we could then change variables according to (5.3) (with t_0 in place of t), to convert to the trivial dynamics

$$
\begin{cases} \dot{\mathbf{X}}_\varepsilon = D\bar{H}^0(\mathbf{P}_\varepsilon) \\ \dot{\mathbf{P}}_\varepsilon = 0. \end{cases}
$$

A more interesting issue is understanding to what extent this procedure works for the slowly time–dependent Hamiltonian $H^\varepsilon = H(\cdot, \cdot, \varepsilon t)$. Remember from §4 that the rescaled dynamics in this situation read

$$
\text{(5.4)} \qquad \begin{cases} \dot{\mathbf{x}}_\varepsilon = D_p H \left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t \right) \\ \dot{\mathbf{p}}_\varepsilon = -D_x H \left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t \right). \end{cases}
$$

Let us also define

$$
\text{(5.5)} \qquad \begin{cases} \mathbf{X}_\varepsilon(t) := X \left(\mathbf{p}_\varepsilon(t), \frac{\mathbf{x}_\varepsilon(t)}{\varepsilon}, t \right) \\ \mathbf{P}_\varepsilon(t) := P \left(\mathbf{p}_\varepsilon(t), \frac{\mathbf{x}_\varepsilon(t)}{\varepsilon}, t \right). \end{cases}
$$

5.1. An assumption about spatial averaging.

We propose to study the limiting behavior of $\mathbf{X}_\varepsilon, \mathbf{P}_\varepsilon$ as $\varepsilon \to 0$. First, as demonstrated below, $\dot{\mathbf{P}}_\varepsilon = O(1)$ is bounded; and so, passing if necessary to a subsequence, we have

$$\mathbf{P}_\varepsilon \to \mathbf{P} \text{ uniformly on compact subsets of } [0,\infty), \tag{5.6}$$

for some Lipschitz continuous function $\mathbf{P} : [0,\infty) \to \mathbb{R}^n$.

On the other hand, $\dot{\mathbf{X}}_\varepsilon = O\left(\frac{1}{\varepsilon}\right)$ is in general unbounded. Consequently, we expect only

$$\mathbf{X}_\varepsilon \rightharpoonup \mathbf{X} \tag{5.7}$$

weakly as $\varepsilon \to 0$; and we may further expect that the weak limit (5.7) entails some sort of spatial averaging.

We will for the remainder of §5 assume that the images of the function $\mathbf{X}_\varepsilon$ became uniformly distributed with respect to Lebesgue measure, as $\varepsilon \to 0$. This seems to be a standard assumption in much of the physical and mathematical literature, and means precisely that

$$\Phi\left(\frac{\mathbf{x}_\varepsilon(t)}{\varepsilon}\right) = \Phi(x(\mathbf{P}_\varepsilon(t), \mathbf{X}_\varepsilon(t), t)) \rightharpoonup \int_{\mathbb{T}^n} \Phi(x(\mathbf{P}(t), X, t))\, dX \tag{5.8}$$

for each continuous, $\mathbb{T}^n$–periodic function Φ. We transform this hypothesis into a statement about the Young measure ν_t, introduced in §4.2. Indeed, setting $\varepsilon = \varepsilon_k \to 0$, the left hand side of (5.8) converges weakly to

$$\int_{\mathbb{T}^n} \Phi(x)\, d\sigma_t,$$

where we define σ_t to be the projection of ν_t onto the x-variables. Since $X = D_P u$, we are therefore making the *assumption of spatial averaging*:

$$d\sigma_t = \det(D^2_{xP} u)\, dx \tag{5.9}$$

for each time $0 \le t \le 1$ and for $u = u(\mathbf{P}(t), x, t)$.

Remark. We pause to check that

$$\omega := \det(D^2_{xP} u)$$

is indeed flow invariant, for each fixed time $0 \le t \le 1$ and $P = \mathbf{P}(t)$, where we write

$$D^2_{xP} u := \begin{pmatrix} u_{x_1 P_1} \dots u_{x_1 P_n} \\ \vdots \quad \ddots \quad \vdots \\ u_{x_n P_1} \dots u_{x_n P_n} \end{pmatrix}, \quad D^2_{Px} u := (D^2_{xP} u)^T.$$

That is, we claim

$$\operatorname{div}(\omega D_p H(D_x u, x, t)) = 0 \qquad \text{for each time } 0 \le t \le 1, \tag{5.10}$$

the divergence taken in the x–variables. This accords with (4.6).

To confirm this, let us first differentiate (5.1) in P:

$$(D^2_{xP} u)^T D_p H = D_p H D^2_{xP} u = D\bar{H}.$$

Left multiply by the cofactor matrix $\operatorname{cof}(D^2_{xP}u)$ and recall the matrix identity $(\operatorname{cof} A)A^T = (\det A)I$. Then

$$\omega D_p H = \operatorname{cof}(D^2_{xP}u) D\bar{H},$$

and so

$$\operatorname{div}(\omega D_p H) = \operatorname{div}(\operatorname{cof}(D^2_{xP}u) D\bar{H}) = 0.$$

The last equality holds since $\bar{H}$ does not depend on x and the columns of $\operatorname{cof}(D^2_{xP}u)$ are divergence-free in x. (See for instance [E2, p. 440].) □

5.2. Adiabatic invariance of the "action" variables.

A first consequence of the averaging hypotheses (5.9) is the adiabatic invariance of the action:

Lemma 5.1. *We have*

$$\dot{\mathbf{P}}_\varepsilon \rightharpoonup 0 \quad \textit{as } \varepsilon \to 0,$$

and so

$$\mathbf{P}_\varepsilon \to P \quad \textit{uniformly on compact subsets of } [0, \infty), \tag{5.11}$$

for some constant vector $P \in \mathbb{R}^n$.

Proof. 1. We compute

$$\dot{\mathbf{P}}_\varepsilon = D_p P\, \dot{\mathbf{p}}_\varepsilon + D_x P \frac{\dot{\mathbf{x}}_\varepsilon}{\varepsilon} + P_t, \tag{5.12}$$

for $P = (P^1, \dots, P^n)$ and

$$D_p P := \begin{pmatrix} P^1_{p_1} \dots P^1_{p_n} \\ \vdots \quad \ddots \quad \vdots \\ P^n_{p_1} \dots P^n_{p_n} \end{pmatrix}, \quad D_x P := \begin{pmatrix} P^1_{x_1} \dots P^1_{x_n}, \\ \vdots \quad \ddots \quad \vdots \\ P^n_{x_1} \dots P^n_{x_n} \end{pmatrix}, \quad P_t = (P^1_t, \dots, P^n_t).$$

Now (5.2) implies

$$
\left\{ \begin{array}{l} I = D^2_{xP} u D_p P \\ O = D^2_{xP} u D_x P + D^2_x u, \end{array} \right. \tag{5.13}
$$

where $D^2_{xP}u$ is defined as before and

$$
D^2_x u := \begin{pmatrix} u_{x_1 x_1} \dots u_{x_1 x_n} \\ \vdots \quad \ddots \quad \vdots \\ u_{x_n x_1} \dots u_{x_n x_n} \end{pmatrix}.
$$

Furthermore, since $H(D_x u, x, t) \equiv \bar{H}(P, t)$, we have

$$
D_p H D^2_x u + D_x H = 0. \tag{5.14}
$$

We combine (5.12)–(5.14), to discover

$$
\begin{aligned}
\dot{\mathbf{P}}_\varepsilon &= (D^2_{xP}u)^{-1} \left(\dot{\mathbf{p}}_\varepsilon - D^2_x u \frac{\dot{\mathbf{x}}_\varepsilon}{\varepsilon} \right) + P_t \quad \text{by (5.13)} \\
&= - \frac{(D^2_{xP}u)^{-1}}{\varepsilon} (D_x H + D_p H D^2_x u) + P_t \quad \text{by (5.4) and the symmetry of } D^2_x u \\
&= P_t \quad \text{by (5.14).}
\end{aligned}
$$

Therefore

$$
\dot{\mathbf{P}}_{\varepsilon_k}(t) = P_t \left(\mathbf{p}_{\varepsilon_k}(t), \frac{\mathbf{x}_{\varepsilon_k}}{\varepsilon_k}(t), t \right) \rightharpoonup \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} P_t(p, x, t)\, d\nu_t. \tag{5.15}
$$

2. We demonstrate next that the right hand side of (5.15) in fact equals zero. For this, we first deduce from (5.2) that

$$
O = D^2_{xP} u P_t + D^2_{xt} u,
$$

and thus

$$
P_t = -(D^2_{xP}u)^{-1} D^2_{xt} u = - \frac{\operatorname{cof}(D^2_{xP}u)^T}{\det(D^2_{xP}u)} D^2_{xt} u.
$$

This and our averaging hypotheses (5.9) together imply

$$
\begin{aligned}
\int_{\mathbb{R}^n} \int_{\mathbb{T}^n} P_t \, d\nu_t &= \int_{\mathbb{T}^n} P_t(D_x u, x, t)\, d\sigma_t \\
&= - \int_{\mathbb{T}^n} \operatorname{cof}(D^2_{xP}u)^T D^2_{xt} u \, dx \\
&= \int_{\mathbb{T}^n} \operatorname{div}(\operatorname{cof}(D^2_{xP}u)^T) u_t \, dx = 0,
\end{aligned}
$$

for $u = u(\mathbf{P}(t), x, t)$. The integration-by-parts in the penultimate step is valid since $u_t = v_t$ is periodic in x. $\square$

Remarks. (i) The computations in this proof in particular reproduce the classical assertion

$$\{P^k, H\} = 0 \quad \text{for } k = 1, \dots, n \tag{5.16}$$

for each time t. Here $P = (P^1, \dots, P^n)$, $P = P(p, x, t)$.

(ii) The foregoing calculation of course strongly depends upon our having smooth functions u, the averaging hypothesis (5.9), etc. It is therefore worth emphasizing again that Theorem 4.3,(ii) is valid in general, without these restrictions, and so can perhaps best be understood as a very weak form of the adiabatic invariance principle. $\square$

5.3. Asymptotics of the "angle" variables.

We turn next to the limiting behavior of the angle variable $\mathbf{X}_\varepsilon(\cdot)$.

Lemma 5.2. *We have*

$$\dot{\mathbf{X}}_\varepsilon(t) = \frac{1}{\varepsilon} D_P \bar{H}(\mathbf{P}_\varepsilon, t) + X_t\left(\mathbf{p}_\varepsilon, \frac{\mathbf{x}_\varepsilon}{\varepsilon}, t\right), \tag{5.17}$$

and so

$$\dot{\mathbf{X}}_{\varepsilon_k}(t) - \frac{1}{\varepsilon_k} D_P \bar{H}(\mathbf{P}_{\varepsilon_k}(t), t) \rightharpoonup \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} X_t(p, x, t)\, d\nu_t. \tag{5.18}$$

Proof. According to (5.5)

$$\dot{\mathbf{X}}_\varepsilon = D_p X \dot{\mathbf{p}}_\varepsilon + D_x X \frac{\dot{\mathbf{x}}_\varepsilon}{\varepsilon} + X_t. \tag{5.19}$$

Also, owing to (5.2),

$$\begin{cases} D_p X = D_P^2 u D_p P \\ D_x X = D_P^2 u D_x P + (D_{xP} u)^T. \end{cases} \tag{5.20}$$

We insert (5.20) into (5.19), and recall as well the ODE (5.4):

$$\begin{aligned} \dot{\mathbf{X}}_\varepsilon &= \frac{D_P^2 u}{\varepsilon}(-D_p P D_x H + D_x P D_p H) \\ &\quad + \frac{1}{\varepsilon}(D_{xP} u)^T D_p H + X_t \\ &= \frac{1}{\varepsilon} D_p H D_{xP} u + X_t, \end{aligned} \tag{5.21}$$

since $\{P, H\} = 0$, according to (5.16).

Next, differentiate (5.1):

$$D_p H D_{xP} u = D_P \bar{H},$$

and employ this identity in (5.21). □

The $O(1)$–term

$$\int_{\mathbb{R}^n} \int_{\mathbb{T}^n} X_t(p, x, t)\, d\nu_t = \int_{\mathbb{T}^n} X_t(D_x u, x, t)\, d\sigma_t \tag{5.22}$$

on right hand side of (5.18) is a correction, recording by how much $\dot{X}_\varepsilon$ fails to equal $\frac{1}{\varepsilon} D\bar{H}(\mathbf{P}^\varepsilon, t)$ for small ε. Over the time interval $0 \leq t \leq 1$, the total correction is

$$\int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} X_t(p, x, t)\, d\nu_t dt.$$

Following Hannay [H] and Berry [B], let us next show

Theorem 5.3. *We have*

$$\int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} X_t(p, x, t)\, d\nu_t dt = D_P K(P), \tag{5.23}$$

for

$$K(P) := \int_0^1 \int_{\mathbb{T}^n} u_t(P, x, t)\, d\sigma_t dt. \tag{5.24}$$

Remark. The expression (5.24) is *Hannay's formula.* It is usually displayed in the equivalent form

$$K(P) = -\int_0^1 \int_{\mathbb{T}^n} p(P, X, t) \cdot x_t(P, X, t)\, dX dt. \tag{5.25}$$

□

Proof. 1. Define

$$w(P, X, t) := u(P, x(P, X, t), t).$$

Then

$$w_t = D_x u \cdot x_t + u_t = p \cdot x_t + u_t,$$

and thus

$$D_P(w_t - p \cdot x_t) = D^2_{Pt} u + D^2_{xt} u D_P x. \tag{5.26}$$

for

$$D^2_{Pt}u := (u_{P_1t}, \ldots, u_{P_nt}), \quad D^2_{xt}u := (u_{x_1t}, \ldots, u_{x_nt}).$$

Now $X = D_P u(P, x, t) = D_P u(P(p, x, t), x, t)$, and consequently

$$X_t = D^2_P u P_t + D^2_{Pt} u. \tag{5.27}$$

2. Since $X = D_P u(P, x(P, X, t), t)$, we also have

$$0 = D^2_P u + (D^2_{xP})^T u D_P x.$$

Likewise, $p = D_x u(P, x, t) = D_x u(P(p, x, t), x, t)$ and so

$$0 = D^2_{xP} u P_t + D^2_{xt} u.$$

Therefore

$$D^2_P u P_t = D^2_{xt} u D_P x$$

and so (5.26), (5.27) imply

$$D_P(w_t - p \cdot x_t) = X_t.$$

Consequently, since $t \mapsto w$ is $[0, 1]$-periodic,

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} X_t(p, x, t)\, d\nu_t dt &= \int_0^1 \int_{\mathbb{T}^n} X_t(p(P, X, t), x(P, X, t), t)\, dX dt \\ &= -D_P \left(\int_0^1 \int_{\mathbb{T}^n} p(P, X, t) \cdot x_t(P, X, t)\, dX dt \right). \end{aligned}$$

3. Finally, we observe that

$$p(P, X, t) \cdot x_t(P, X, t) = D_x u(P, x(P, X, t), t) \cdot x_t = \frac{d}{dt}(u(P, x, t)) - u_t.$$

Therefore

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{T}^n} X_t(p, x, t)\, d\nu_t dt &= D_P \left(\int_0^1 \int_{\mathbb{T}^n} u_t(P, x(P, X, t), t)\, dX dt \right) \\ &= D_P \left(\int_0^1 \int_{\mathbb{T}^n} u_t(P, x, t)\, d\sigma_t dt \right) = D_P K(P). \end{aligned}$$

□

6. Averaging and the Berry–Hannay correction.

We devote this section to working out some more connections between averaging in the spatial variables (for each fixed time) versus averaging in spatial and temporal variables together. These relations appear in certain asymptotic limits.

To start, let us take $P \in \mathbb{R}^n$ and for each $0 < \delta \leq 1$ define

$$\hat{H}(P, \delta) \tag{6.1}$$

to be the unique real number for which the corrector problem

$$\begin{cases} \delta w_t^\delta + H(P + D_x w^\delta, x, t) = \hat{H}(P, \delta) & \text{in } \mathbb{R}^n \times \mathbb{R} \\ (x,t) \mapsto w^\delta \ \text{ is } \mathbb{T}^{n+1}\text{-periodic}, \end{cases} \tag{6.2}$$

has a solution $w^\delta = w^\delta(P, x, t)$. The existence and uniqueness of $\hat{H}(P, \delta)$ follows as in §2.2.

Remark. We can interpret the parameter δ as controlling the scaling in time. To see this, observe that (6.2) can be recast as

$$\begin{cases} \hat{w}_t^\delta + H(P + D_x \hat{w}^\delta, x, \delta t) = \hat{H}(P, \delta) & \text{in } \mathbb{R}^n \times \mathbb{R} \\ x \mapsto \hat{w}^\delta \text{ is } \mathbb{T}^n\text{-periodic}, \ t \mapsto \hat{w}^\delta \text{ is } [0, \frac{1}{\delta}]\text{-periodic}, \end{cases} \tag{6.3}$$

for $\hat{w}^\delta(P, x, t) := w^\delta(P, x, \delta t)$. □

We will be concerned with the behavior of $\hat{H}$ for small δ:

Lemma 6.1. *We have*

$$\lim_{\delta \to 0} \hat{H}(P, \delta) = \int_0^1 \bar{H}(P, t)\, dt \qquad (P \in \mathbb{R}^n). \tag{6.4}$$

Proof. 1. At a point (x_0, t_0) where w^δ attains its maximum,

$$H(P, x_0, t_0) \leq \hat{H}(P, \delta);$$

and likewise

$$H(P, x_1, t_1) \geq \hat{H}(P, \delta)$$

at a point (x_1, t_1) where w^δ attains its minimum. Thus

$$|\hat{H}(P, \delta)| \leq \max_{\mathbb{T}^{n+1}} |H(P, \cdot, \cdot)|.$$

We may therefore take $\delta_k \to 0$ such that

$$\hat{H}(P, \delta_k) \to \lambda,$$

for some real number λ.

2. Next, define

$$u^\delta(x,t) := \delta\left(w^\delta\left(\frac{x}{\delta},t\right) - w^\delta(0,0)\right); \tag{6.5}$$

then

$$u_t^\delta + H\left(P + D_x u^\delta, \frac{x}{\delta}, t\right) = \hat{H}(P,\delta) \quad \text{in } \mathbb{R}^n \times \mathbb{R}. \tag{6.6}$$

Furthermore, as in the proof of Theorem 2.1, we have

$$\max_{\mathbb{R}^n\times\mathbb{R}} |u^\delta|, |D_x u^\delta|, |u_t^\delta| \le C$$

for a constant C independent of δ. Taking $\delta = \delta_k$, and passing if needs be to a further subsequence, we have

$$u^{\delta_k} \to u \quad \text{uniformly on } \mathbb{R}^n \times \mathbb{R} \tag{6.7}$$

and

$$u_t + \bar{H}(P + D_x u, t) = \lambda \quad \text{in } \mathbb{R}^n \times \mathbb{R}. \tag{6.8}$$

On the other hand, write

$$\tilde{u} := \left(\int_0^1 \bar{H}(P,s)\,ds\right) t - \int_0^t \bar{H}(P,s)\,ds \tag{6.9}$$

and notice

$$\tilde{u}_t + \bar{H}(P + D_x\tilde{u}, t) = \int_0^1 \bar{H}(P,s)\,ds \quad \text{in } \mathbb{R}^n \times \mathbb{R}. \tag{6.10}$$

But according to a uniqueness argument similar to that in §2, there exists at most one constant λ such that (6.8) has a $\mathbb{T}^{n+1}$–periodic solution. Consequently $\lambda = \int_0^1 \bar{H}ds$. □

See Gomes [G1] for a different proof.

Notation. In view of (6.4), we hereafter write

$$\hat{H}(P,0) := \int_0^1 \bar{H}(P,s)\,ds. \tag{6.11}$$

We propose next to estimate the first-order behavior of $\hat{H}(P,\delta)$ for δ near zero. For this we reintroduce for each time $0 \le t \le 1$ the cell problem

$$\begin{cases} H(P + D_x v, x, t) = \bar{H}(P,t) & \text{in } \mathbb{R}^n \\ x \mapsto v \ \text{ is } \mathbb{T}^n\text{-periodic.} \end{cases} \tag{6.12}$$

Also, we define

$$v^\delta := w^\delta - \frac{1}{\delta}\int_0^t \hat{H}(P,0) - \bar{H}(P,s)\,ds. \tag{6.13}$$

Assume further that we have probability measures ρ^δ and $\{\sigma_t\}_{0\le t\le 1}$, as in §§3,4. Then, similarly to (3.6),

$$\int_{\mathbb{T}^{n+1}} \delta\Phi_t + D_pH(P+D_xw^\delta,x,t)\cdot D_x\Phi\,d\rho^\delta = 0 \tag{6.14}$$

for each C^1 function Φ periodic in x,t; and

$$\int_{\mathbb{T}^n} D_pH(P+D_xv,x,t)\cdot D_x\Phi\,d\sigma_t = 0 \tag{6.15}$$

for each $0\le t\le 1$ and each Φ periodic in x.

Lemma 6.2. *For each $\delta>0$, we have the bounds*

$$\int_0^1\int_{\mathbb{T}^n} v_t^\delta\,d\sigma_t dt \le \frac{\hat{H}(P,\delta)-\hat{H}(P,0)}{\delta} \le \int_{\mathbb{T}^{n+1}} v_t\,d\rho^\delta. \tag{6.16}$$

Proof. 1. From (6.1), (6.15) we see that

$$\delta v_t^\delta + H(P+D_xv^\delta,x,t) = \hat{H}(P,\delta) - \hat{H}(P,0) + \bar{H}(P,t), \tag{6.17}$$

and so

$$\delta v_t^\delta + H(P+D_xv^\delta,x,t) - H(P+D_xv,x,t) = \hat{H}(P,\delta) - \hat{H}(P,0) \tag{6.18}$$

in the viscosity sense.

Now convexity implies

$$H(P+D_xv^\delta,x,t) - H(P+D_xv,x,t) \begin{cases} \le D_pH(P+D_xv^\delta,x,t)\cdot(D_xv^\delta - D_xv) \\ \ge D_pH(P+D_xv,x,t)\cdot(D_xv^\delta - D_xv). \end{cases}$$

Consequently

$$\hat{H}(P,\delta)-\hat{H}(P,0) \le \delta(v^\delta - v)_t + D_pH(P+D_xw^\delta,x,t)\cdot D_x(v^\delta - v) + \delta v_t.$$

We integrate with respect to the measure ρ^δ and recall (6.14), thereby proving the second inequality in (6.16). Likewise

$$\hat{H}(P,\delta)-\hat{H}(P,0) \ge \delta v_t^\delta + D_pH(P+D_xv,x,t)\cdot D_x(v^\delta - v);$$

and we integrate with respect to σ_t, then with respect to t, recalling (6.15). □

Remark. The foregoing estimate suggests that we can recover the Hannay-Berry correction, discussed in §5, by differentiating $\hat{H}(P,\delta)$ with respect to δ. We propose therefore the formal relation

$$\frac{\partial \hat{H}}{\partial \delta}(P,0) = \int_0^1 \int_{\mathbb{T}^n} v_t \, d\sigma_t dt = \int_0^1 \int_{\mathbb{T}^n} u_t \, d\sigma_t dt = K(P), \tag{6.19}$$

although we are not able to make any precise statement as to the general validity of this identity. It follows of course from (6.16) if $\int_0^1 \int_{\mathbb{T}^n} v_t^\delta \, d\sigma_t dt, \int_{\mathbb{T}^{n+1}} v_t \, d\rho^\delta \to \int_0^1 \int_{\mathbb{T}^n} v_t \, d\sigma_t dt$ as $\delta \to 0$.

This formula is potentially interesting, since both $\hat{H}(P,\delta)$ and $\hat{H}(P,0) := \int_0^1 \bar{H}(P,s)\, ds$ exist in general, without the very strong assumptions of integrability and averaging introduced for this and the previous section. It therefore seems worthwhile to try to understand if and when $\frac{\partial \hat{H}}{\partial \delta}(P,0)$ exists. These heuristics provide, we hope, a bit of progress in the overall program of discovering how the effective Hamiltonians record information about dynamics. □

Appendix

In this appendix, we attempt to clarifly connections of our work with that of Mather, Fathi and others. Some key references are Mather [Mt1-4], Mather–Forni [M-F], Fathi [F1-4], Iturriaga [I], Mañé [Mn], Dias Carneiro [DC], and Weinan E [EW1,2]. See also Carlsson, Haurie, and Leizarowitz [C-H-L] and Maderna [M]. Contreras–Iturriaga [C-I] and Fathi [F5] provide lengthly and detailed lecture notes. For simplicity we discuss only autonomous Hamiltonians and Lagrangians.

Basic notation. The general setting for most of these papers is a complete, compact, connected manifold M, the tangent space of which is TM. Local coordinates on TM are written (x, v), for $x \in M$ and $v \in T_x M$. (In our work, we discuss only the case that $M = \mathbb{T}^n$, the n-dimensional flat torus, and $TM = \mathbb{T}^n \times \mathbb{R}^n$. The covering space is then $\mathbb{R}^n$. We usually write the velocity variable as "q" instead of "v", to emphasize the duality with the momentum variable "p".)

Given also is the *Lagrangian* $L : TM \to \mathbb{R}$, which is a smooth function that in local coordinates satisfies appropriate convexity and superlinearity hypotheses.

Classical action. To each trajectory $\mathbf{x}(\cdot)$ taking values in M and to each time $T > 0$ is associated the classical *action*

$$A[\mathbf{x}(\cdot)] := \int_0^T L(\mathbf{x}, \dot{\mathbf{x}})\, dt,$$

also sometimes denoted by S.

We may look for trajectories which minimize the action, subject to given and fixed endpoint conditions. The Euler-Lagrange equations in local coordinates then read

$$-\frac{d}{dt}\left(D_v L(\mathbf{x}, \dot{\mathbf{x}})\right) + D_x L(\mathbf{x}, \dot{\mathbf{x}}) = 0. \tag{7.1}$$

Generalized action. Instead of considering minimizing trajectories, Mather introduces measures invariant under the dynamics (7.1), that minimize the action. More precisely, consider a probability measure μ, supported in TM, that is invariant under the flow (7.1). The corresponding *action* of μ is

$$A[\mu] := \int_{TM} L(x, v)\, d\mu.$$

To replace the fixed endpoint constraints above, let us associate with each invariant measure μ an element $\rho(\mu)$ of the first homology group $H_1(M)$, by setting

$$< \rho(\mu), \omega > := \int_{TM} \omega\, d\mu$$

for each ω in the first cohomology group $H^1(M)$. (In our case that $M = \mathbb{T}^n$, we can identify $H_1(M) = H^1(M) = \mathbb{R}^n$ and $Q = \rho(\mu)$ is the rotation vector.)

Convex and concave functions. For each $\gamma \in H_1(M)$, Mather defines

$$\beta(\gamma) := \inf\{A[\mu] \mid \rho(\mu) = \gamma\};$$

that is, $\beta(\gamma)$ is the infimum of the action over all flow–invariant measures with fixed rotation number γ. (We identify γ with $Q \in \mathbb{R}^n$ and set $\beta(\gamma) = \bar{L}(Q)$. In other words, Mather's function β corresponds to the effective Lagrangian $\bar{L}$ in the sense of Lions-Papanicolaou-Varadhan.) Let us call a measure μ with $\rho(\mu) = \gamma$ and $A[\mu] = \beta(\gamma)$ a *minimizing* or *Mather measure.*

For $\omega \in H^1(M)$, Mather defines as well

$$A_\omega[\mu] := \int_{TM} L - \omega \, d\mu.$$

Since ω is a closed one form, the Euler-Lagrange equations for L and $L - \omega$ are the same. (When $M = \mathbb{T}^n$, we can represent ω as some vector $P \in \mathbb{R}^n$.) We may also consider for $\omega \in H^1(M)$ the convex function

$$\alpha(\omega) := -\inf\{A_\omega[\mu]\},$$

with no constraint on the rotation vector for μ. Then for each $\gamma \in H_1(M)$ there exists $\omega \in H^1(M)$ such that

$$\alpha(\omega) + \beta(\gamma) = < \gamma, \omega > . \tag{7.2}$$

(In the case $M = \mathbb{T}^n$, we set $\omega = P, \gamma = Q, \alpha(\omega) = \bar{H}(P), \beta(\gamma) = \bar{L}(Q)$; and then (7.2) asserts $\bar{H}(P) + \bar{L}(Q) = P \cdot Q$, which in turn means $P \in \partial\bar{L}(Q),\ Q \in \partial\bar{H}(P)$. This is just the duality between the effective Hamiltonian and Lagrangian.)

Weak KAM theory. In a sequence of papers [F1-4], recounted in the lecture notes [F5], A. Fathi has developed a weak form of KAM theory, which augments Mather's work. These papers largely intersect our work, and we spell out below some of the connections. We understand the term "weak KAM theory" to mean finding generalized, nonsmooth solutions in the large for appropriate Hamilton–Jacobi type PDE, for which usual KAM theory provides smooth solutions in the case of small, nonresonant perturbations.

In [F1] Fahti demonstrates the existence of a constant c and two Lipschitz continuous functions $u^\pm : M \to \mathbb{R}$, solving the PDE

$$H(D_x u^\pm, x) = c \quad \text{a.e.}, \tag{7.3}$$

H of course denoting the Hamiltonian derived from L. The constant c is $\bar{H}(0)$. (We note that in fact

$$H(D_x u^+, x) = c \quad \text{in the viscosity sense} \tag{7.4}$$

and

$$-H(D_x u^-, x) = -c \quad \text{in the viscosity sense,} \tag{7.5}$$

as follows from the variational formulas for $u^\pm$ in [F1]. Assertion (7.5) means that the usual inequalities in the definition of viscosity solutions are reversed: See for instance Chapter 10 of [E2].)

Define the Mather set $\tilde{\mathcal{M}}_0$ to be the closure of the union of the supports of all minimizing measures, that is, those flow invariant probability measures μ satisfying $\int_{TM} L(x,v)\,d\mu = -c$. Also let $\mathcal{M}_0$ denote the projection of $\tilde{\mathcal{M}}_0$ into M. Then Fathi notes that $u^\pm$ are differentiable on $\mathcal{M}_0$, and that $\tilde{\mathcal{M}}_0$ lies on a Lipschitz graph, as follows in part from earlier work of Mather and of Mañé. Dias Carneiro [DC] proved that the support of any minimizing measure is contained in a level set of the energy. Our paper [E-G1] recast and reproved these statements using PDE techniques, the main difference being that we constructed our version of Mather sets in phase space, i.e. in the variables x and p.

In [F2,3] Fathi establishes many more properties of $u^\pm$, and in particular shows how to construct "Peierls barriers" from the collection of functions $u^\pm$. In [F4] he demonstrates the convergence as $t \to \infty$ of solutions of $v_t + H(D_x v, x) = 0$ to a stationary solution of (7.3).

References

[A] V. Arnold, *Mathematical Methods of Classical Mechanics*, Springer, 1978.

[B] M. Berry, *Classical adiabatic angles and quantal adiabatic phase*, J. Physics A **18** (1985), 15–27.

[B-D] A. Braides and A. Defranceschi, *Homogenization of Multiple Integrals*, Oxford Univ. Press, 1998.

[C-H-L] D. Carlsson, A . Haurie, and A. Leizarowitz, *Infinite horizon optimal control: deterministic and stochastic systems*, Springer, 1991.

[DC] M. J. Dias Carneiro, *On minimizing measures of the action of autonomous Lagrangians*, Nonlinearity **8** (1995), 1077–1085.

[C-D] W. Chou and R. J. Duffin, *An additive eigenvalue problem of physics related to linear programming*, Advances in Appl. Math **8** (1987), 486–498.

[Cl] F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, 1983.

[C1] M. Concordel, *Periodic homogenization of Hamilton-Jacobi equations I: additive eigenvalues and variational formula*, Indiana Univ. Math. J. **45** (1996), 1095–1117.

[C2] M. Concordel, *Periodic homogenisation of Hamilton-Jacobi equations II: eikonal equations*, Proc. Roy. Soc. Edinburgh **127** (1997), 665–689.

[C-I] G. Contreras and R. Iturriaga, *Global minimizers of autonomous Lagrangians*, Lecture notes available at http://www.cimat.mx/ gonzalo/, (2000).

[Cr] F. Crawford, *Elementary examples of adiabatic invariance*, Amer. J. Physics **58** (1990), 337–344.

[EW1] Weinan E, *A class of homogenization problems in the calculus of variations*, Comm Pure and Appl Math **44** (1991), 733–754.

[EW2] Weinan E, *Aubry–Mather theory and periodic solutions of the forced Burgers equation*, Comm Pure and Appl Math **52** (1999), 811–828.

[E1] L. C. Evans, *Weak Convergence Methods for Nonlinear Partial Differential Equations*, American Math Soc, 1990.

[E2] L. C. Evans, *Partial Differential Equations*, American Math Soc, 1998.

[E3] L. C. Evans, *Periodic homogenization of certain fully nonlinear PDE*, Proc Royal Society Edinburgh **120** (1992), 245–265.

[E-G1] L. C. Evans and D. Gomes, *Effective Hamiltonians and averaging for Hamiltonian dynamics I*, Archive Rational Mech and Analysis **157** (2001), 1–33.

[F1] A. Fathi, *Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens*, C. R. Acad. Sci. Paris Sr. I Math. **324** (1997), 1043–1046.

[F2] A. Fathi, *Solutions KAM faibles conjuguées et barrières de Peierls*, C. R. Acad. Sci. Paris Sr. I Math. **325** (1997), 649–652.

[F3] A. Fathi, *Orbites hétéroclines et ensemble de Peierls*, C. R. Acad. Sci. Paris Sr. I Math. **326** (1998), 1213–1216.

[F4] A. Fathi, *Sur la convergence du semi-groupe de Lax-Oleinik*, C. R. Acad. Sci. Paris Sr. I Math. **327** (1998), 267–270.

[F5] A. Fathi, *Weak KAM theory in Lagrangian Dynamics, Preliminary Version*, Lecture notes (2001).

[G1] D. Gomes, *Hamilton–Jacobi Equations, Viscosity Solutions and Asymptotics of Hamiltonian Systems*, Ph.D. Thesis, University of California, Berkeley (2000).

[G2] D. Gomes, *Viscosity solutions of Hamilton-Jacobi equations and asymptotics for Hamiltonian systems*, to appear in Calculus of Variations and PDE.

[G3] D. Gomes, *Regularity theory for Hamilton-Jacobi equations*, In preparation.

[Gd] H. Goldstein, *Classical mechanics (2nd ed.)*, Addison-Wesley, 1980.

[H] J. H. Hannay, *Angle variable holonomy in adiabatic excursion of an integrable Hamiltonian*, J. Physics A **18** (1985), 221–230.

[I] R. Iturriaga, *Minimizing measures for time–dependent Lagrangians*, Proc. London Math Soc **73** (1996), 216–240.

[J-K-M] H. R. Jauslin, H. O. Kreiss and J. Moser, *On the forced Burgers equation with periodic boundary conditions*, in Differential equations: La Pietra 1996 (Florence), Proc. Sympos. Pure Math **65** (1999), 133–153.

[L-P-V] P.-L. Lions, G. Papanicolaou, and S. R. S. Varadhan, *Homogenization of Hamilton–Jacobi equations*, unpublished, circa 1988.

[M] E. Maderna, *Symetries des systemes Lagrangiens*, IThesis, Ecole Normale Superieure de Lyon, 2000.

[Mn] R. Mañé, *Global Variational Methods in Conservative Dynamics*, Instituto de Matemática Pura e Aplicada, Rio de Janeiro.

[Mt1] J. Mather, *Minimal measures*, Comment. Math Helvetici **64** (1989), 375–394.

[Mt2] J. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems*, Math. Zeitschrift **207** (1991), 169–207.

[Mt3] J. Mather, *Differentiability of the minimal average action as a function of the rotation number*, Bol. Soc. Bras. Math **21** (1990), 59–70.

[Mt4] J. Mather, *Variational construction of connecting orbits*, Ann. Inst. Fourier, Grenoble **43** (1993), 1347–1386.

[M-F] J. Mather and G. Forni, *Action minimizing orbits in Hamiltonian systems*, Transition to Chaos in Classical and Quantum Mechanics, Lecture Notes in Math 1589 (S. Graffi, ed.), Sringer, 1994.

[N] R. Nussbaum, *Convergence of iterates of a nonlinear operator arising in statistical mechanics*, Nonlinearity **4** (1991), 1223–1239.

Department of Mathematics, University of California, Berkeley, CA 94720

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540