ON DYNAMICS OF LAGRANGIAN TRAJECTORIES FOR HAMILTON–JACOBI EQUATIONS

KONSTANTIN KHANIN AND ANDREI SOBOLEVSKI

Abstract. Characteristic curves of a Hamilton–Jacobi equation can be seen as action minimizing trajectories of fluid particles. However this description is valid only for smooth solutions. For nonsmooth "viscosity" solutions, which give rise to discontinuous velocity fields, this picture holds only up to the moment when trajectories hit a shock and cease to minimize the Lagrangian action. In this paper we show that for any convex Hamiltonian, a viscous regularization allows to construct a nonsmooth flow that extends particle trajectories and determines dynamics inside the shock manifolds. This flow consists of integral curves of a particular velocity field, which is uniquely defined everywhere in the flow domain and is discontinuous on shock manifolds.

1. Introduction

1.1. **The Hamilton–Jacobi equation and viscosity solutions.** The evolutionary Hamilton–Jacobi equation

$$\frac{\partial \phi}{\partial t} + H(t, x, \nabla\phi) = 0, \tag{1.1}$$

appears in diverse mathematical models ranging from analytical mechanics to combinatorics, condensed matter, turbulence, and cosmology (see, e.g., a non-exhaustive set of references in [1]). In many of these applications the objects of interest are described by singularities of solutions, which inevitably appear for generic initial data after a finite time due to the nonlinearity of (1.1). Therefore one of the central issues both for theory and applications is to understand the behaviour of the system after singularities form.

A useful example to be borne in mind when thinking about (1.1)—and arguably the most widely known variant thereof—is the Riemann, or inviscid Burgers, equation. In the physics notation (the dot $\cdot$ for inner product and ∇ for spatial gradient) this equation has the form

$$\frac{\partial u}{\partial t} + u \cdot \nabla u = 0, \qquad u = \nabla\phi. \tag{1.2}$$

The first eq. (1.2) corresponds to the Hamiltonian $H(t, x, p) = |p|^2/2$. This equation may in turn be considered as a limit of vanishing viscosity of the Burgers equation

$$\frac{\partial u^\mu}{\partial t} + u^\mu \cdot \nabla u^\mu = \mu\nabla^2 u^\mu, \qquad u^\mu = \nabla\phi^\mu, \tag{1.3}$$

so solutions of (1.2) can be defined as limits of smooth solutions to (1.3) as the positive parameter μ goes to zero.

2000 *Mathematics Subject Classification.* Primary 35F21; Secondary 49L25, 76N10.

We acknowledge the support of the French Ministry for Science and Higher Education. The work of A.S. was partially supported by Laboratory for structural methods of data analysis in predictive modeling of the Moscow Iinstitute of Physics and Technology, RF Government grant 11.G34.31.0073, and by Agence Nationale de la Recherche, France, project ANR–07–BLAN–0235 OTARIE. A.S. gratefully acknowledges hospitality of Department of Mathematics, University of Toronto. K.K. acknowledges support of the NSERC Discover Grant.

The Burgers equation is in fact very special: it can be exactly mapped by the Cole–Hopf transformation into the linear heat equation and therefore explicitly integrated, which in turn allows to explicitly study the limit $\mu \downarrow 0$ [2]. Although in the case of general convex Hamiltonian the Hopf–Cole transformation is not available, the qualitative behaviour of solutions to a parabolic regularization of (1.1)

$$\frac{\partial \phi^\mu}{\partial t} + H(t, x, \nabla\phi^\mu) = \mu\nabla^2\phi^\mu \tag{1.4}$$

as viscosity vanishes is similar to that for the Burgers equation. The limit $\phi(t, x) = \lim_{\mu\downarrow 0} \phi^\mu(t, x)$ exists and is called the *entropy* (or *viscosity*) *solution*.

A theory of weak solutions for a general Hamilton–Jacobi equation, employing the regularization by infinitesimal viscosity, exists since the 1970s [3–5]. In the one-dimensional setting this theory is essentially equivalent to the earlier theory of hyperbolic conservation laws [2, 6–8]. The theory of weak solutions for the Hamilton–Jacobi equation is closely related to calculus of variations, and introduction of diffusion corresponds to stochastic control arguments [9]. The viewpoint of the present paper is somewhat complementary: the Hamilton–Jacobi equation is considered as a fluid dynamics model, and the main goal is to construct a flow of "fluid particles" inside the shocks of a weak solution. However it is convenient to start with the Lax–Oleinik variational principle, which provides a purely variational construction of the viscosity solution. Remarkably this construction does not use any explicit viscous regularization.

1.2. **The variational construction of viscosity solutions and shocks.** Assume that the Hamiltonian function $H(t, x, p)$ is smooth and strictly convex in the momentum variable p, i.e., is such that for all (t, x) the graph of $H(t, x, p)$ as a function of p lies above any tangent plane and contains no straight segments. This implies that the formula $v = \nabla_p H(t, x, p)$ establishes a one-to-one correspondence between values of velocity v and momentum p. Moreover, the Lagrangian function

$$L(t, x, v) = \max_p\, [p \cdot v - H(t, x, p)], \tag{1.5}$$

under the above hypotheses is smooth and strictly convex in v. (Note that L may be not finite everywhere: e.g., the relativistic Hamiltonian $H(t, x, p) = \sqrt{1 + |p|^2}$ corresponds to the Lagrangian $L(t, x, v)$ that is defined for $|v| \le 1$ as $-\sqrt{1 - |v|^2}$ and takes value $+\infty$ elsewhere. This does not happen if in addition one assumes that the Hamiltonian H grows superlinearly in $|p|$.)

The relation between the Lagrangian and the Hamiltonian is symmetric: they are Legendre–Fenchel conjugate (1.5) to one another. This relation can also be expressed in the form of the Young inequality:

$$L(t, x, v) + H(t, x, p) \ge v \cdot p, \tag{1.6}$$

which holds for all v and p and turns into equality whenever $v = \nabla_p H(t, x, p)$ or equivalently $p = \nabla_v L(t, x, v)$. The two maps $p \mapsto \nabla_p H(t, x, p)$ and $v \mapsto \nabla_v L(t, x, v)$ are thus inverse to each other; we will call them the *Legendre transforms* at (t, x) of p and of v. (Usually the term "Legendre transform" refers to the relation between the conjugate functions H and L; here we follow the usage that is adopted by A. Fathi in his works on weak KAM theory [10] and is more convenient in the present context.)

Note that if $H(t, x, p) = |p|^2/2$, then $L(t, x, v) = |v|^2/2$ and the Legendre transform reduces to the identity $v = p$, blurring the distinction between velocities and momenta. This is another special feature of the (inviscid) Burgers equations.

Now assume that $\phi(t, x)$ is a strong solution of the inviscid equation (1.1), i.e., a C^2 function that satisfies the equation in the classical sense. For an arbitrary differentiable trajectory $\gamma(t)$ the full time derivative of ϕ along γ is given by

$$\frac{\mathrm{d}\phi(t,\gamma)}{\mathrm{d}t} = \frac{\partial \phi}{\partial t} + \dot\gamma \cdot \nabla\phi = \dot\gamma \cdot \nabla\phi - H(t,\gamma,\nabla\phi) \le L(t,\gamma,\dot\gamma), \tag{1.7}$$

where at the last step the Young inequality (1.6) is used. This implies a bound for the mechanical action corresponding to the trajectory γ:

$$\phi(t_2,\gamma(t_2)) \le \phi(t_1,\gamma(t_1)) + \int_{t_1}^{t_2} L(s,\gamma(s),\dot\gamma(s))\,\mathrm{d}s. \tag{1.8}$$

Equality in (1.7) is only achieved if $\dot\gamma$ is the Legendre transform of $\nabla\phi$ at every point $(t,\gamma(t))$:

$$\dot\gamma(t) = \nabla_p H(t,\gamma,\nabla\phi(t,\gamma)). \tag{1.9}$$

Therefore the bound (1.8) is achieved for trajectories satisfying Hamilton's canonical equations, with momentum given for the trajectory γ by $p_\gamma(t) := \nabla\phi(t,\gamma(t))$. (The second canonical equation, $\dot p = -\nabla_x H$, follows from (1.1) and (1.9) for a C^2 solution ϕ because

$$\dot p_\gamma(t) = \frac{\partial \nabla\phi}{\partial t} + \dot\gamma\cdot(\nabla\otimes\nabla\phi) = -\nabla_x H(t,\gamma,\nabla\phi) - \nabla_p H\cdot(\nabla\otimes\nabla\phi) + \dot\gamma\cdot(\nabla\otimes\nabla\phi), \tag{1.10}$$

where the last two terms cancel.)

This is a manifestation of the variational *principle of the least action*: Hamiltonian trajectories $(\gamma(t), p_\gamma(t))$ are (locally) action minimizing. In particular, if the initial condition

$$\phi(t=0,y) = \phi_0(y), \tag{1.11}$$

is a fixed smooth function, the identity

$$\phi(t,x) = \phi_0(\gamma(0)) + \int_0^t L(s,\gamma(s),\dot\gamma(s))\,\mathrm{d}s \tag{1.12}$$

holds for a Euler–Lagrange trajectory γ such that $\gamma(t) = x$ and $p_\gamma(0) = \nabla\phi_0(\gamma(0))$.

However the least action principle has wider validity: in fact it can be used to *construct* the viscosity solution corresponding to the initial data (1.11):

$$\phi(t,x) = \min_{\gamma\colon \gamma(t)=x} \Big(\phi_0(\gamma(0)) + \int_0^t L(s,\gamma(s),\dot\gamma(s))\,\mathrm{d}s\Big). \tag{1.13}$$

This is the celebrated Lax–Oleinik formula (see, e.g., [11] or [10]), which reduces a PDE problem (1.1), (1.11) to the variational problem (1.13) where minimization is extended to all sufficiently smooth (in fact absolutely continuous) curves γ such that $\gamma(t) = x$.

At those points (t,x) where the function ϕ defined by (1.13) is smooth in x, the minimizing trajectory is unique. In this case, the minimizer can be embedded in a smooth family of minimizing trajectories whose endpoints at time 0 and t are continuously distributed about $\gamma(0)$ and $\gamma(t) = x$ (a convenient reference is [17, Section 6.4], although this fact is classical). A piece of initial data ϕ_0 gets continuously deformed according to (1.7) along this bundle of trajectories into a piece of smooth solution ϕ to (1.1) defined in a neighbourhood of x at time t. Of course the Hamilton–Jacobi equation is satisfied by ϕ in strong sense at all points where it is differentiable.

But the crucial feature of (1.13) is that generally there will be points (t,x) with several minimizers γ_i that start at different locations $\gamma_i(0)$ and bring the same value of action to $x = \gamma_i(t)$. Just as above, each of these Hamiltonian trajectories will be responsible for a separate smooth "piece" of solution. Thus for locations x' close to x the function ϕ will be represented as a pointwise minimum of these smooth pieces ϕ_i:

$$\phi(t, x') = \min_i \phi_i(t, x'). \tag{1.14}$$

As all γ_i have the same terminal value of action, all the pieces intersect at (t, x): $\phi_1(t, x) = \phi_2(t, x) = \cdots = \phi(t, x)$. Thus the neighbourhood of x at time t is partitioned into domains where ϕ coincides with each of the smooth functions ϕ_i and satisfies the Hamilton–Jacobi equation (1.1) in the strong sense. These domains are separated by surfaces of various dimensions where two, or possibly three or more, pieces ϕ_i intersect and their pointwise minimum ϕ is not differentiable. Such surfaces are called *shock manifolds* or simply *shocks*. Note that a function ϕ defined by the Lax–Oleinik formula is continuous everywhere, including the shocks; it is its gradient that suffers a discontinuity.

In general, there are infinitely many continuous functions that match the initial condition (1.11) and in the complement of the shock surfaces are differentiable and satisfy the Hamilton–Jacobi equation (1.1), just as ϕ does. What distinguishes the function ϕ defined by the variational construction (1.13) from all these "weak solutions", and grants it with important physical meaning, is that ϕ appears in the limit of vanishing viscosity for the regularized equation (1.4) with the initial condition (1.11) (see, e.g., [4]). For a smooth Hamiltonian it can be proved that in a viscosity solution minimizers can only merge with shocks but never leave them.

Now observe that in a solution ϕ given by the Lax–Oleĭnik formula (1.13) a minimizer that has come to a shock cannot be continued any longer as a minimizing trajectory: wherever it might go, there will be other trajectories originated at $t = 0$ that will bring smaller values of action to the same location. Hence for the purpose of the least action description (1.13), Hamiltonian trajectories become irrelevant as soon as they hit shocks. The set of trajectories which survive as minimizers until time $t > 0$ is decreasing with t, but at all times it is sufficiently large to cover the whole continuum of final positions.

1.3. **The fluid dynamics picture.** Let us now adopt an alternative viewpoint and consider the Hamilton–Jacobi equation as a fluid dynamics model, assuming that Hamiltonian trajectories (1.9) are described by material "particles" transported by the velocity field $u(t, x)$, which is the Legendre transform of the momenta field $p(t, x) = \nabla\phi(t, x)$. From this new perspective it is no longer natural to accept that particles annihilate once they reach a shock. Can therefore something be said about the dynamics of those particles that got into the shock, notwithstanding the fact that their trajectories cease to minimize the action? The difficulty in such an approach is related to the discontinuous nature of the velocity field u, which makes it impossible to construct classical solutions to the transport equation $\dot\gamma(t) = u(t, \gamma)$.

In dimension $d = 1$ the answer to the question above is readily available. Shocks at each fixed t are isolated points in the x space and as soon as a trajectory merges with one of them, it continues to move with the shock at all later times. This description is related to C. Dafermos' theory of generalized characteristics [12] which, in fact, can be extended to a much more general situation of nonconvex Hamiltonians and systems of conservation laws. However, in several space dimensions shock manifolds are extended surfaces of different codimension, and dynamics of trajectories inside shocks is by no means trivial.

Our interest in this problem is related to the earlier work by I. Bogaevsky where the problem was solved for the case of the Burgers equation (1.3) [13, 14] using the following approach. Consider the differential equation

$$\dot\gamma^\mu(t) = u^\mu(t, \gamma^\mu), \qquad \gamma^\mu(0) = y. \tag{1.15}$$

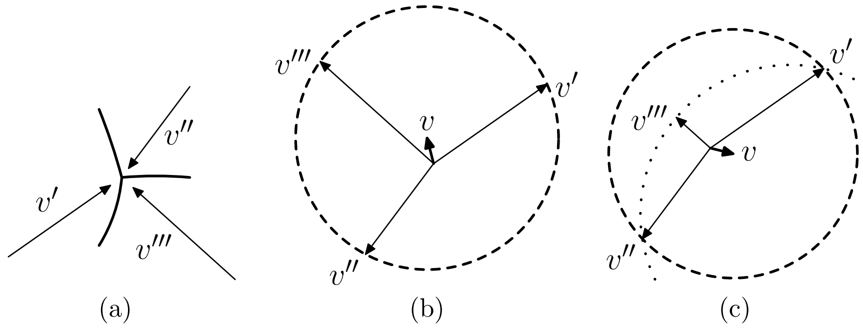

FIGURE 1. Bogaevsky's construction [13, 14] of the effective velocity v at a triple shock point in dimension $d = 2$: (a) the local structure of the flow (v', v'', and v''' are the limiting values of velocity when the triple point is approached from three different domains of smooth flow); (b) the effective velocity v is the center of the smallest circle containing the three limiting velocities; (c) the smallest circle (dashed) is not necessarily the circumscribed one (dotted), so the effective velocity may be determined by a proper subset of limiting velocities (here, v' and v'').

Since u^μ for $\mu > 0$ is a smooth vector field, this equation defines a family of particle trajectories that form a smooth flow. The next step is to take the limit of this flow as $\mu \downarrow 0$. It turns out that this limit exists as a non-differentiable continuous flow, for which the forward derivative $\dot\gamma(t+0) = \lim_{\tau\downarrow 0}[\gamma(t+\tau)-\gamma(t)]/\tau$ is defined everywhere. If $\gamma(t)$ is located outside shocks, this derivative coincides with $u(t, \gamma(t))$. Otherwise the effective velocity $\dot\gamma(t + 0)$ is determined by the extremal values of velocities $u_i = \nabla\phi_i$, and there is an interesting explicit representation for it: $\dot\gamma(t+0)$ coincides with the center of the smallest ball that contains all u_i (fig. 1). The limiting flow turns out to be *coalescing* (and therefore not time-reversible): once any two trajectories intersect, they stay together for all later times.

Moreover, it turns out that pieces of the shock manifold may be classified into *restraining* and *nonrestraining* depending on whether trajectories stay on them or leave them along pieces of shock manifold of lower codimension[1]. Shocks of codimension one are always restraining; in particular, such are all shocks in the one-dimensional case. Interestingly, this classification, introduced for the first time by Bogaevsky in [13] ("acute" and "obtuse" superdifferentials of ϕ) seems to have been overlooked by physicists despite its clear physical significance [15].

The proofs of all these facts in [13, 14] were based on specific properties of the quadratic Hamiltonian and cannot be extended to the general setting of a convex Hamiltonian. In this work we propose a different approach to the vanishing viscosity limit in the general setting. This approach, which is based on the fundamental uniqueness of the possible limiting behaviour of γ^μ, leads to results on existence, uniqueness, and explicit representation of the limit velocities.

It should be remarked that equation (1.9), which relates the velocity $\dot\gamma$ of a trajectory to the gradient $\nabla\phi(t,\gamma)$ of the solution, can be seen as defining a generalization of the gradient flow of the function ϕ [10, 14]. Such a flow coincides with the conventional gradient flow when $H(p) = |p|^2/2$ and ϕ is smooth. The case of a concave (or semiconcave) nonsmooth ϕ can be handled using the differential inequality that goes back to the work of H. Brézis [16]. A similar approach was also used by P. Cannarsa and C. Sinestrari in the context of propagation of singularities

[1]In the example of fig. 1, case (c) corresponds to a nonrestraining triple point in $d = 2$, which trajectories leave through the shock line that divides domains of smooth flow with limiting velocities v' and v''; see fig. 3 from [13] and the discussion therein.

for the eikonal equation with quadratic Hamiltonian [17, Lemma 5.6.2]. However the classification of singluarities into “restraining” and “nonrestraining” seems to have been unknown before the work [13] even in the quadratic case.

We conclude with a few observations concerning our approach.

Seen as a family of continuous maps of variational origin from initial coordinates y to current coordinates x, the flow γ_y is clearly relevant for optimal transportation problems [23, 24]. An interesting problem suggested by B. Khesin is to study the extremal properties of this flow. Indeed it is known from [26] that before the first shock formation the flow γ_y is and action minimizing flow of diffeomorphisms, while the first shock formation time t^* marks a conjugate point in the corresponding variational problem. According to the suggested view, the flow constructed above may be seen as a kind of saddle-point, rather than minimum, for a suitable transport optimization problem.

Another natural context to place our construction in is that of differential inclusions (see, e.g., [27]). The flow consructed in this paper may be seen as a solution of differential inclusion

$$\dot{\gamma} \in \nabla_p H(t, \gamma, \mathrm{Pr}_p \partial\phi(t, \gamma)), \tag{1.16}$$

where Pr_p is the p projection of the superdifferential $\partial\phi$. In comparison with standard constructions of the theory of differential inclusions the flow γ_y solves (1.16) in a stronger sense: the forward derivative $\dot{\gamma}_y(t + 0)$ exists everywhere.

1.4. **Outline of the paper.**

In Section 2 we develop a local theory for Lagrangian particles in a gradient flow defined by a viscosity solution ϕ. Here we introduce the notions of *admissible velocity* and *admissible momentum* at a shock, which are central to our approach. The admissible velocity at each point turns out to be the unique solution to a particular convex minimization problem, which extends the construction of the center of the smallest Euclidean ball (cf fig. 1) to the general convex case.

In Section 3 we show that the limit of a flow regularized with small viscosity is tangent to the field of admissible velocities. This establishes an existence theorem for integral curves of this field.

The issue of uniqueness of the limiting flow and a perturbative approach that allows to determine the higher time derivatives of limiting trajectories are discussed in Section 4.

In the course of this work we benefitted from valuable discussions with Jérémie Bec, Patrick Bernard, Ilya Bogaevsky, Yann Brenier, Philippe Choquard, Michael Dabkowski, Uriel Frisch, and Boris Khesin. It is a pleasure to recognize their help as well as the unique environment of the Observatoire de la Côte d’Azur, where this work has started and advanced.

2. Viscosity solutions and admissible gradient vector fields

2.1. **Superdifferentials of viscosity solutions.**

Let ϕ be a viscosity solution to the Hamilton–Jacobi equation (1.1) with initial data (1.11). We shall use the following standard facts, for which we refer the reader again to the recent and very useful exposition in [17, Section 6.4], although many of these facts date from 50 and more years ago: (i) the function ϕ is locally uniformly semiconcave in (t, x) variables; (ii) if there is a single minimizer coming to (t, x), then ϕ is differentiable at this point, C^2 smooth in some its neighbourhood, and

$$\phi(t + \tau, x + \xi) = \phi(t, x) + \frac{\partial\phi}{\partial t}\,\tau + \nabla\phi \cdot \xi + o(|\tau| + |\xi|) \tag{2.1}$$

$$= \phi(t, x) - H(t, x, \nabla\phi)\,\tau + \nabla\phi \cdot \xi + o(|\tau| + |\xi|); \tag{2.2}$$

(iii) if ϕ is not differentiable at (t, x) and there is a finite number of minimizers γ_i such that $\gamma_i(t) = x$, then each of them corresponds to a different smooth branch ϕ_i of solution defined in the neighbourhood of (t, x). Then the Lax–Oleinik formula implies that

$$\phi(t+\tau, x+\xi) = \min_i \phi_i(t+\tau, x+\xi) \tag{2.3}$$

$$= \phi(t, x) + \min_i(-H_i\tau + p_i \cdot \xi) + o(|\tau| + |\xi|), \tag{2.4}$$

where $p_i := \nabla\phi_i(t, x)$ and $H_i := H(t, x, p_i)$.

In the latter case neither of expressions $-H_i\tau + p_i \cdot \xi$ provides a valid linear approximation to the difference $\phi(t+\tau, x+\xi) - \phi(t, x)$ at all points, but they all *majorize* this difference up to a remainder that is either linear or higher-order, depending on τ and ξ. Evidently, so does the linear form $-H\tau + p \cdot \xi$ for any convex combination

$$p = \sum_i \lambda_i p_i, \qquad H = \sum_i \lambda_i H_i \tag{2.5}$$

with $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$. In convex analysis these convex combinations are called *supergradients* of ϕ at (t, x) and the whole collecton of them, which is a convex polytope with vertices $(-H_i, p_i)$, is called the *superdifferential* of ϕ [17,18]. We use for the superdifferential the notation $\partial\phi(t, x)$.

To avoid a possible misunderstanding it should be noted that, although uniqueness of a minimizer coming to (t, x) implies differentiability in x of a nonsmooth solution ϕ to the Hamilton–Jacobi equation at t and earlier times, it does *not* imply its differentiability at any $t + \tau > t$. Such points (t, x) where the differentiability cannot be extended to an open neighbourhood in spacetime are called *preshocks* [1] and correspond to conjugate points of a corresponding variational problem; a classification of all the possible combinations of shocks and preshocks in dimensions $d = 2$ and $d = 3$ is provided in [19]. Note that at a preshock the linearization of ϕ_i does not fully anticipate the shocks at times $t + \tau$ for any $\tau > 0$. However the differentiability is recovered as $\tau \downarrow 0$, because the corresponding superdifferential shrinks to the gradient of ϕ at (t, x).

Under a viscous regularization ϕ^μ of the solution ϕ, a shock point (t, x) is "smeared" over a small area where $(\partial\phi^\mu/\partial t, \nabla\phi^\mu)$ takes on all values from relative interior of $\partial\phi(t, x)$. Thus, intuitively, $\partial\phi(t, x)$ is a set of values taken by the spacetime gradient of the nonsmooth function ϕ in an infinitesimal neighbourhood of (t, x).

Completing the gradient field with superdifferentials at points where ϕ is not smooth recovers, in a weaker sense, the continuity of the map $(t, x) \mapsto \partial\phi(t, x)$. Indeed, suppose (t_n, x_n) converges to (t, x) and the sequence $(-H_n, p_n) \in \partial\phi(t_n, x_n)$ has a limit point $(-H, p)$. By definition of superdifferential,

$$\phi(t_n + \tau, x_n + \xi) - \phi(t_n, x_n) \leq -H_n\,\tau + p_n \cdot \xi + o(|\tau| + |\xi|); \tag{2.6}$$

passing here to the limit and using continuity of ϕ, we see that $(-H, p) \in \partial\phi(t, x)$. Therefore the superdifferential $\partial\phi(t, x)$ contains all the limit points of superdifferentials $\partial\phi(t_n, x_n)$ as (t_n, x_n) converges to (t, x).

To make this argument rigorous, some control is needed over the remainder term in (2.6). This is easy for convex or concave functions [18], for which such inequalities hold without remainders. A wider function class, which contains viscosity solutions of Hamilton–Jacobi equations and in which such control is still possible, is formed by semiconvex or semiconcave functions [17]. We refer a reader interested in proofs of this and other convex analytic results used in this paper to monographs [17,18].

2.2. **Admissible velocities and admissible momenta.** This section will describe a procedure that gives a unique possible velocity and momentum at each point (t, x). The construction is based solely on the convexity of Hamiltonian in the momentum variable. We thus set the stage for the next section where the connection with the viscous regularization is established.

If (t, x) is a regular point, i.e., not a point of shock, then the velocity $u(t, x)$ and the momentum $p(t, x)$ are naturally defined. Suppose now that (t, x) is a point of shock formed by intersection of smooth branches ϕ_i, $i \in \mathfrak{I}$. For a particle starting from a shock point (t, x) its possible velocity v must correspond to one of the the "available" momenta, i.e., to a momentum p that belongs to the convex hull of the momenta $p_i = \nabla\phi_i(t, x)$, $i \in \mathfrak{I}$, or equivalently to the p-projection of the superdifferential $\partial\phi(t, x)$.

In fact more can be said. For an infinitesimal positive τ, when the particle has already left its original location with velocity v, not all branches ϕ_i will be relevant for the solution ϕ at a point $(t+\tau, x+v\tau)$ as $\tau \downarrow 0$, but only those that contribute to the linear approximation of the solution, i.e., the minimum in $\min_{i\in\mathfrak{I}}(-H_i + p_i \cdot v)$ (cf. (2.4) with $\xi = v\tau$). The branches not contributing to the minimum can be discarded.

Denote the set of relevant indices

$$I(v) := \{j \in \mathfrak{I}\colon\ -H_j + p_j \cdot v = \min_{i\in\mathfrak{I}}(-H_i + p_i \cdot v)\}; \tag{2.7}$$

it is nonempty because the minimum is attained due to convexity of $H(t, x, \cdot)$.

We can now postulate that any possible velocity v of a Lagrangian particle inside a shock satisfies the following condition. This postulate is justified by Lemma 2 in §3.

Admissibility condition. *A velocity v^* is said to be* admissible *at (t, x) if the corresponding momentum $p^* = \nabla_v L(t, x, v^*)$ belongs to the convex hull of momenta p_i with $i \in I(v^*)$:*

$$p^* \in \operatorname{conv}\{p_j : j \in I(v^*)\}. \tag{2.8}$$

This value of momentum p^* is also called admissible *at (t, x).*

Observe that since the index set $I(v)$ depends on v, this condition can be viewed as a kind of self-consistency requirement.

Equivalently, one can write

$$v^* \in \nabla_p H(t, x, \operatorname{conv}\{p_j : j \in I(v^*)\}). \tag{2.9}$$

Note that, in contrast with theory of generalized characteristics for Hamilton–Jacobi equations [17, Definition 5.5.1], there is no convex hull taken in (2.9) *after* the (nonlinear) map $\nabla_p H(t, x, \cdot)$ is apllied to the superdifferential of ϕ at (t, x), even though the resulting set in the velocity space is generally non-convex.

It is known [20, Example 3.5] that with an extra convex hull operation mentioned above, uniqueness may fail for the velocity of a generalized characteristic. In contrast, the above definition allows to fix the velocity v^* uniquely:

Theorem 1 (uniqueness of admissible velocity)**.** *Let ϕ be a viscosity solution to the Cauchy problem* (1.1)*,* (1.11)*. Then at any (t, x) there exists a unique admissible velocity $v^* = v^*(t, x)$, which is the unique point of the global minimum for the function*

$$\hat{L}(v) := L(t, x, v) - \min_{i\in\mathfrak{I}}(-H_i + p_i \cdot v). \tag{2.10}$$

Proof. Recall that $L(t,x,v)$ is a strictly convex function of v because of assumptions formulated in §1. Rewriting

$$L_i(v) := L(t,x,v) + H_i - p_i \cdot v, \qquad \hat{L}(v) = \max_{i \in \mathfrak{I}} L_i(v), \tag{2.11}$$

we see that $\hat{L}(v)$ is a pointwise maximum of strictly convex functions and therefore is strictly convex itself. Furthermore, because the Hamiltonian $H(t,x,p)$ is assumed to be finite for all p, its conjugate Lagrangian $L(t,x,v)$ grows faster than any linear function as $|v|$ increases, and thus all its level sets are bounded. Therefore $\hat{L}(v)$ attains its minimum at a unique value of velocity v^*.

To simplify the presentation of ideas we start with an (elementary) proof of the theorem in the particular case when $I(v^*)$ is finite. We show first that the point of minimum v^* satisfies the admissibility condition (2.8). Indeed,

$$\nabla_v L_i(v^*) = \nabla_v L(t,x,v^*) - p_i = p^* - p_i. \tag{2.12}$$

Suppose that p^* does not belong to the convex hull of p_j, $j \in I(v^*)$. Then there exists a vector h such that $(p^* - p_j) \cdot h < 0$ for all $j \in I(v^*)$. It follows that $L_j(v^* + \epsilon h) < L_j(v^*)$ for all $j \in I(v^*)$ if $\epsilon > 0$ is sufficiently small. Hence, $\hat{L}(v^* + \epsilon h) < \hat{L}(v^*)$ for sufficiently small ϵ, which contradicts to our assumption that v^* is a point of minimum. This contradiction proves that v^* is admissible.

To prove uniqueness we show that if $\hat{v}$ is admissible then it is a (necessarily unique) point of global minimum for the strictly convex function $\hat{L}$. Using the strict convexity of L_j, we obtain

$$L_j(\hat{v} + h) = L(t,x,\hat{v} + h) + H_j - p_j \cdot (\hat{v} + h) \tag{2.13}$$

$$> L_j(\hat{v}) + \nabla_v L(t,x,\hat{v}) \cdot h - p_j \cdot h = L_j(\hat{v}) + (\hat{p} - p_j) \cdot h, \tag{2.14}$$

where $\hat{p}$ is the Legendre transform of $\hat{v}$. Since $\hat{v}$ is admissible, $\hat{p} = \sum_j \lambda_j p_j$, where all $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$. Hence, $\sum_j \lambda_j (\hat{p} - p_j) \cdot h = [(\sum_j \lambda_j)\hat{p} - \sum_j \lambda_j p_j] \cdot h = [\hat{p} - \hat{p}] \cdot h = 0$. It follows that $(\hat{p} - p_j) \cdot h > 0$ for at least one $j \in I(\hat{v})$. This implies that $\hat{L}(\hat{v} + h) > \hat{L}(\hat{v})$, which means that $\hat{v}$ is a point of global minimum for $\hat{L}$.

In the general case of an arbitrary $I(v^*)$, only the first argument, namely admissibility of the global minimum v^*, needs modification. We shall us the following result of Clarke based on earlier work of Ioffe and Levin [21, theorem 2.8.2 and corollary 1]). Let $\hat{L}(v) = \max_{i \in \mathfrak{I}} L_i(v)$, where $\mathfrak{I}$ is a compact topological space, and suppose that all functions $L_i(\cdot)$ are convex and Lipschitz with the same constant and $I(v)$ is the set of i's for which the maximum is attained; then the subdifferential $\partial \hat{L}(v)$ is a weakly-$*$ closed convex hull of the union of $\partial L_i(v)$ for $i \in I(v)$. (To justufy compactness of $\mathfrak{I}$, observe that one can use the values of momenta p_i instead of the abstract indices i and that the set of minimizers coming to (t,x) is closed and their momenta are bounded.) Now take into account that in our case $\partial L_i(v) = \nabla_v L(t,x,v) - p_i$ and that v^* is the point of minimum, i.e., that

$$0 \in \partial \hat{L}(v^*) = \operatorname{conv}\{\, \nabla_v L(t,x,v^*) - p_i \colon i \in I(v^*)\,\} = \{\, p^* - p_i \colon i \in I(v^*)\,\};$$

this coincides with the admissibility condition $p^* \in \operatorname{conv}\{\, p_i \colon i \in I(v^*)\,\}$ (2.8). □

Thus the admissibility property, first formulated above in the hardly manageable combinatorial form (2.8), turns out to be the optimality condition for a nice convex minimization problem (2.10). In particular, if ϕ is differentiable at (t,x), then $\hat{L}(v) = L(t,x,v) + H(t,x,\nabla\phi) - \nabla\phi \cdot v$ and the minimum in (2.10) is achieved at the Legendre transform of $\nabla\phi$. We thus recover Hamilton's equation (1.9).

The following reformulation will clarify the connection between admissibility and the original construction for the Burgers equation proposed by Bogaevsky in [13,14]. Let $v_i = \nabla_p H(t,x,p_i)$ be the velocity corresponding to the limit momentum p_i

and observe that $p_i = \nabla_v L(t,x,v_i)$. The Legendre duality implies that $H_i = H(t,x,p_i) = p_i \cdot v_i - L(t,x,v_i)$ and therefore (2.11) assumes the form

(2.15)
$$\hat{L}(v) = \max_{i\in\mathfrak{I}} [L(t,x,v) - L(t,x,v_i) - \nabla_v L(t,x,v_i)\cdot(v-v_i)] = \max_{i\in\mathfrak{I}} D_L^{t,x}(v \mid v_i).$$

The quantity in square brackets is known as the *Bregman divergence* $D_L^{t,x}(v \mid v_i)$ of vector v with respect to v_i, a non-symmetric measure of separation of vectors with respect to the convex function $L(t,x,\cdot)$ [22]. Theorem 1 terefore means that the admissible velocity is the center of the the smallest "Bregman sphere" containing all v_i, $i \in \mathfrak{I}$. When $L(t,x,v) = |v|^2/2$, the Bregman divergence reduces to (half) the squared distance between the two vectors. Therefore the admissible velocity v^* exactly conicides with the centre of smallest ball containing all v_i, and we recover the result of [13, 14].

Finally, let us discuss the "physical" meaning of the function $\hat{L}$. Consider an infinitesimal movement from (t,x) with velocity v. It follows from the least action principle that $\phi(t,x) + L(t,x,v)\,\mathrm{d}t - \phi(t+\mathrm{d}t, x+v\,\mathrm{d}t) \geq 0$. It is easy to see that to the linear order in $\mathrm{d}t$

$$\phi(t,x) + L(t,x,v)\,\mathrm{d}t - \phi(t+\mathrm{d}t, x+v\,\mathrm{d}t) = \hat{L}(v)\,\mathrm{d}t. \tag{2.16}$$

Hence the unique admissible velocity v^* minimizes the rate of growth of the difference in action between the true minimizers and trajectories of particles on shocks. In other words, the trajectory inside a shock cannot be a minimizer but it does its best to keep its surplus action growing as slowly as possible.

3. The vanishing viscosity limit for velocities

In the preceding section we constructed a canonical vector field of admissible velocities $v^*(t,x) = \nabla_p H(t,x,p^*(t,x))$ that corresponds to a given viscosity solution ϕ of the Cauchy problem (1.1), (1.11). Notice that in general this vector field is discontinuous on the shock manifold.

To see how the vector field of admissible velocities arises for Lagrangian particles inside shocks, consider the vanishing viscosity limit for a flow corresponding to the parabolic regularization

$$\frac{\partial \phi^\mu}{\partial t} + H(t,x,\nabla\phi^\mu) = \mu\nabla^2\phi^\mu, \qquad \mu > 0, \tag{3.1}$$

of the Hamilton–Jacobi equation (1.1).

For sufficiently smooth initial data $\phi_0(y) = \phi^\mu(t=0,y)$ the partial differential equation (3.1) has a globally defined strong solution ϕ^μ, which is locally Lipschitz with a constant independent of μ. Moreover, ϕ^μ converges as $\mu \downarrow 0$ to the unique viscosity solution ϕ corresponding to the same initial data. Proofs of these facts may be found, e.g., in [4], where they are established for $\phi_0 \in C^{2,\alpha}$.

Consider now the transport equation

$$\dot{\gamma}^\mu(t) = \nabla_p H(t,\gamma^\mu,\nabla\phi^\mu(t,\gamma^\mu)), \qquad \gamma^\mu(0) = y. \tag{3.2}$$

For $\mu > 0$ this equation has a unique solution which continuously depends on the initial location y. Fix a point (t_0,x_0) with $t_0 > 0$ and pick trajectories γ^μ for all sufficiently small $\mu > 0$ such that $\gamma^\mu(t_0) \to x_0$ as $\mu \downarrow 0$. The uniform Lipschitz property of solutions ϕ^μ implies that the curves γ^μ are uniformly bounded and equicontinuous on some interval containing t_0. Hence there exists a curve $\bar{\gamma}$ and a sequence $\mu_i \downarrow 0$ such that $\lim_{\mu_i\downarrow 0}\gamma^{\mu_i} = \bar{\gamma}$ uniformly in t on that interval. Note that all γ^{μ_i} and $\bar{\gamma}$ are also Lipschitz with a constant independent of μ and that $\bar{\gamma}(t_0) = x_0$.

Let furthermore $\bar v$ be a limit point of the "forward velocity" of the curve $\bar\gamma$ at (t_0, x_0), i.e., let for some sequence $\tau_k \downarrow 0$

$$\bar v = \lim_{\tau_k \downarrow 0} \frac{1}{\tau_k}[\bar\gamma(t_0+\tau_k) - \bar\gamma(t_0)]. \tag{3.3}$$

We cannot conclude *a priori* that the curve $\bar\gamma$ or the velocity $\bar v$ are uniquely defined. However it turns out that $\bar v$ must satisfy the admissibility condition with respect to the solution ϕ and therefore it coincides with the unique admissible velocity v^*.

Theorem 2 (admissibility of limit velocities)**.** *The momentum $\bar p = \nabla_v L(t_0, x_0, \bar v)$ corresponding to a limit velocity $\bar v$ in* (3.3) *is admissible at (t_0, x_0), i.e., it satisfies $\bar p \in \operatorname{conv}\{p_i : i \in I(\bar v)\}$.*

Proof. Our general strategy in what follows is a proof by contradiction: assume that $\bar v$ is not admissible and show that it then cannot be a limit velocity.

We first set up some notation regarding geometry of the closed convex set $\partial\phi(t_0, x_0)$. Denote by (s, p) the space-time co-tangent coordinates, with s a scalar dual to the time subspace and p a vector dual to the d-dimensional configuration subspace. Let $\Lambda_{\bar v}$ be the hyperplane supporting the convex compact set $\partial\phi(t_0, x_0)$ from below with the slope corresponding to the velocity $\bar v$:

$$\Lambda_{\bar v} = \{(s, p) : s = -p\cdot\bar v + \min_{i\in\mathfrak{I}}(p_i\cdot\bar v - H_i)\} \tag{3.4}$$

and let $\bar S$ be the intersection of $\Lambda_{\bar v}$ and of the superdifferential $\partial\phi(t_0, x_0)$, i.e., the face of $\partial\phi(t_0, x_0)$ spanned by vertices $(-H_i, p_i)$ with indices in $I(\bar v)$ [cf. equation (2.7)]. In this geometric setting the admissibility condition for momentum (2.8) can be formulated in the following way: an admissible momentum must belong to the p-projection of the set $\bar S$. Denote also $\bar\Phi := \min_{i\in\mathfrak{I}}(p_i\cdot\bar v - H_i)$.

Lemma 3. *If $\bar p$ does not belong to the p-projection of $\bar S$, then*

$$M := \min_{(s,p)\in\partial\phi(t_0,x_0)} [s + p\cdot\bar v - \bar\Phi + (p-\bar p)\cdot(\nabla_p H(t_0, x_0, p) - \bar v)] > 0, \tag{3.5}$$

Proof. It should be noted that M in (3.5) is an auxiliary quantity, which plays role in the subsequent proof but has no geometric meaning by itself. It can be seen as a sum of two parts, each of which is nonnegative for reasons related to the convexity of the Hamiltonian H and the superdifferential $\partial\phi(t_0, x_0)$, and which cannot simultaneously vanish.

Denote $\bar s = -\bar p\cdot\bar v + \bar\Phi$ and observe that the point $(\bar s, \bar p)$ cannot belong to $\partial\phi(t_0, x_0)$ because $(\bar s, \bar p) \in \Lambda_{\bar v}$ but the p-projection of the face $\bar S = \Lambda_{\bar v} \cap \partial\phi(t_0, x_0)$ does not contain $\bar p$.

Monotonicity of the gradient $\nabla_p H(t_0, x_0, p)$ of the convex function H implies that for $p \neq \bar p$

$$(p - \bar p)\cdot(\nabla_p H(t_0, x_0, p) - \bar v) > 0. \tag{3.6}$$

Indeed, from the strict convexity of $H(t_0, x_0, \cdot)$ in momentum it follows that

$$\begin{aligned} H(t_0, x_0, p) &> H(t_0, x_0, \bar p) + (p - \bar p)\cdot\nabla_p H(t_0, x_0, \bar p), \\ H(t_0, x_0, \bar p) &> H(t_0, x_0, p) + (\bar p - p)\cdot\nabla_p H(t_0, x_0, p) \end{aligned}$$

whenever $p \neq \bar p$ and in particular when $(s, p) \in \bar S$. Adding these two inequalities and taking into account that $\nabla_p H(t_0, x_0, \bar p) = \bar v$, we get (3.6).

Furthermore, as $\Lambda_{\bar v}$ supports $\partial\phi(t_0, x_0)$ from below, for all $(s, p) \in \partial\phi(t_0, x_0)$ we have

$$s + p\cdot\bar v - \bar\Phi \geq 0 \tag{3.7}$$

with equality only when $(s,p) \in \Lambda_{\bar v}$. Thus the function of (s,p) in the square brackets in (3.5) is strictly positive on $\partial\phi(t_0,x_0)$. Indeed, if (s,p) belongs to the face $\bar S$, then (3.6) is positive, and otherwise (3.7) is positive. □

In the rest of the proof it will be convenient to use a different rearrangment of the expression in square brackets in (3.5):

$$[\mathrm{I}] - [\mathrm{II}] := [s + (p-\bar p)\cdot\nabla_p H(t_0,x_0,p)] - [\bar\Phi - \bar p\cdot\bar v]. \tag{3.8}$$

Next we provide a precise meaning to the intuitive idea that for (t,x) sufficiently close to (t_0,x_0) and μ sufficiently small, the values of the function ϕ^μ and its derivatives are close to those for the linearization

$$\phi(t_0,x_0) + \min_{i\in\mathfrak{I}}\,[p_i\cdot(x-x_0) - (t-t_0)\,H_i] \tag{3.9}$$

of the viscosity solution ϕ near (t_0,x_0).

For $\epsilon > 0$ let V_ϵ be the ϵ-neighbourhood of $\partial\phi(t_0,x_0)$. Choose $\epsilon < M/(6+3|\bar v|)$ so small that $(\bar s,\bar p) \notin V_\epsilon$ and

$$\min_{(s,p)\in V_\epsilon}([\mathrm{I}] - [\mathrm{II}]) \geq 2M/3 > 0. \tag{3.10}$$

Using the upper semicontinuity of the superdifferential (see e.g. [17, Proposition 3.3.4] or [18, Corollary 24.5.1], where a similar result is proved for convex functions), choose $R = R(\epsilon) > 0$ and $T = T(\epsilon) > 0$ such that for all $(t,x) \in \mathcal{D}_{T,R} := \{(t,x)\colon 0 \leq t-t_0 \leq T,\ |x-x_0| \leq R\}$ the superdifferential $\partial\phi(t,x)$ is contained in the set $V_{\epsilon/2}$.

Reducing T, R if necessary and using the Lipschitz property of ϕ, ϕ^μ (which implies boundedness of momenta) and continuity of $\nabla_p H(t,x,p)$ in (t,x) variables, we can assume in addition that for all $(t,x) \in \mathcal{D}_{T,R}$

$$|(p-\bar p)\cdot\nabla_p H(t,x,p) - (p-\bar p)\cdot\nabla_p H(t_0,x_0,p)| < \epsilon \tag{3.11}$$

Denote $\Gamma(t) := x_0 + \bar v(t-t_0)$. Reducing T once again, we can guarantee that for all $t_0 \leq t \leq t_0 + T$ both $|\Gamma(t) - \Gamma(t_0)| < qR$ and $|\bar\gamma(t) - \bar\gamma(t_0)| < qR$ with any $0 < q < 1$ (this margin is needed because we will approximate $\bar\gamma$ by γ^μ, which must also belong to $\mathcal{D}_{T,R}$) and that, moreover, $\partial\phi(t,\Gamma(t))$ is contained in $\bar S_{\epsilon/2}$, the $\epsilon/2$-neighbourhood of $\bar S$. The latter is possible because all limit points of $\partial\phi(t,\Gamma(t))$ as $t \downarrow t_0$ belong to the face of $\partial\phi(t_0,x_0)$ that corresponds to the direction $\bar v$, i.e., to $\bar S$. A proof of this result, which refines the upper semicontinutity property of superdifferentials mentioned above, can be found e.g., in the context of convex functions in [18, Theorem 24.6]; its generalization to the semiconcave case is evident.

In what follows we will refer to the values of μ_i from the sequence that determines $\bar\gamma$, but will drop the index i to simplfy the notation. Choose $\bar\mu = \bar\mu(\epsilon)$ sufficiently small so that the following three conditions hold:

$$(t,\bar\gamma(t)) \in \mathcal{D}_{T,R} \quad \text{for } \mu < \bar\mu(\epsilon)$$

(this is indeed possible because $(t,\bar\gamma(t)) \in \mathcal{D}_{T,qR}$ with $q < 1$),

$$\Big(\frac{\partial\phi^\mu}{\partial t}(t,x), \nabla\phi^\mu(t,x)\Big) \in V_\epsilon \tag{3.12}$$

everywhere in $\mathcal{D}_{T,R}$, and

$$\Big(\frac{\partial\phi^\mu}{\partial t}(t,\Gamma(t)), \nabla\phi^\mu(t,\Gamma(t))\Big) \in \bar S_\epsilon \tag{3.13}$$

for $t_0 < t < t_0 + T$. The latter two conditions hold because convergence of semiconcave functions ϕ^μ to ϕ implies that limit points of their derivatives belong to $\partial\phi(t,x) \subset V_{\epsilon/2}$ (in particular, $\partial\phi(t,\Gamma(t)) \subset \bar S_{\epsilon/2}$ along the trajectory Γ).

We are now set for the concluding argument. Assume that $\bar v$ is not an admissible velocity and therefore the correspondent momentum $\bar p$ does not belong to the p-projection of $\bar S$. We are going to show that in this case, although trajectories γ^μ may occasionally pass close to the trajectory $\Gamma(t) = x_0 + \bar v(t - t_0)$, any possible limiting value of velocity of the limit trajectory $\bar\gamma$ as $\tau = t - t_0 \downarrow 0$ differs from $\bar v$ by a positive constant. The central argument is provided by the following lemma.

Lemma 4. *Under conditions of Lemma 3 fix arbitrary positive $\tau < T/3$ and $\delta < M/[6(L + |\bar p|)]$, where L is the common spatial Lipschitz constant of ϕ^μ in $\mathcal{D}_{T,R}$ for $0 < \mu < \bar\mu$. Define the cone $K_\delta := \{(t, x) \in \mathcal{D}_{T,R} \colon |x - \Gamma(t)| < \delta(t - t_0)\}$ and suppose that $(t_0 + \tau, \bar\gamma(t_0 + \tau)) \in K_\delta$. Then $(t, \gamma^\mu(t)) \notin K_\delta$ for all $\mu < \bar\mu$ and t such that $3\tau < t - t_0 < T$.*

Proof. The full time derivative of the function $(t, x) \mapsto \phi^\mu(t, x) - \bar p \cdot x$ along γ^μ is given by

$$\begin{aligned}(3.14)\quad \frac{d}{dt}[\phi^\mu(t, \gamma^\mu(t)) - \bar p \cdot \gamma^\mu(t)] &= \frac{\partial\phi^\mu}{\partial t}(t, \gamma^\mu) + (\nabla\phi^\mu(t, \gamma^\mu) - \bar p) \cdot \dot\gamma^\mu \\ &= \frac{\partial\phi^\mu}{\partial t}(t, \gamma^\mu) + (\nabla\phi^\mu(t, \gamma^\mu) - \bar p) \cdot \nabla_p H(t, \gamma^\mu, \nabla\phi^\mu(t, \gamma^\mu)) \\ &\ge \frac{\partial\phi^\mu}{\partial t}(t, \gamma^\mu) + (\nabla\phi^\mu(t, \gamma^\mu) - \bar p) \cdot \nabla_p H(t_0, x_0, \nabla\phi^\mu(t, \gamma^\mu)) - \epsilon,\end{aligned}$$

where the last inequality follows from (3.11). Integrating this from $t_0 + \tau$ to t we get

$$\begin{aligned}\text{[I]}\quad &\phi^\mu(t, \gamma^\mu(t)) - \bar p \cdot \gamma^\mu(t) - \phi^\mu(t_0 + \tau, \gamma^\mu(t_0 + \tau)) + \bar p \cdot \gamma^\mu(t_0 + \tau) \\ &\ge \int_{t_0+\tau}^{t} \Big[\frac{\partial\phi^\mu}{\partial t}(t', \gamma^\mu) + (\nabla\phi^\mu(t', \gamma^\mu) - \bar p) \cdot \nabla_p H(t_0, x_0, \nabla\phi^\mu(t', \gamma^\mu))\Big]\, dt' \\ &\qquad - \epsilon(t - t_0 - \tau)\end{aligned}$$

On the other hand,

$$\begin{aligned}(3.15)\quad \frac{d}{dt}[\phi^\mu(t, \Gamma(t)) - \bar p \cdot \Gamma(t)] &= \frac{\partial\phi^\mu}{\partial t}(t, \Gamma(t)) + (\nabla\phi^\mu(t, \Gamma(t)) - \bar p) \cdot \bar v \\ &\le \bar\Phi - \bar p \cdot \bar v + \epsilon\,(1 + |\bar v|),\end{aligned}$$

where we took into account (3.13) and the fact that $s + p \cdot \bar v = \bar\Phi$ for all $(s, p) \in \bar S$ (cf. (3.7)). It follows that

$$\begin{aligned}\text{[II]}\quad &\phi^\mu(t, \Gamma(t)) - \bar p \cdot \Gamma(t) - \phi^\mu(t_0 + \tau, \Gamma(t_0 + \tau)) + \bar p \cdot \Gamma(t_0 + \tau) \\ &\le \int_{t_0+\tau}^{t} (\bar\Phi - \bar p \cdot \bar v)\, dt' + \epsilon\,(1 + |\bar v|)(t - t_0 - \tau).\end{aligned}$$

Subtracting [II] from [I], using (3.8), (3.10), (3.12) and observing that the Lipschitz property of ϕ^μ (and correspondingly that of $x \mapsto \phi^\mu(t, x) - \bar p \cdot x$, with the constant $L + |\bar p|$) implies that

$$\begin{aligned}(3.16)\quad &\big|\phi^\mu(t_0 + \tau, \Gamma(t_0 + \tau)) - \bar p \cdot \Gamma(t_0 + \tau) \\ &\quad - \phi^\mu(t_0 + \tau, \gamma^\mu(t_0 + \tau)) + \bar p \cdot \gamma^\mu(t_0 + \tau)\big| \le (L + |\bar p|)\delta\tau,\end{aligned}$$

we get

$$\begin{aligned}(3.17)\quad &\phi^\mu(t, \gamma^\mu(t)) - \bar p \cdot \gamma^\mu(t) - \phi^\mu(t, \Gamma(t)) + \bar p \cdot \Gamma(t) \\ &\ge [\tfrac{2}{3}M - \epsilon(2 + |\bar v|)](t - t_0 - \tau) - (L + |\bar p|)\delta\tau.\end{aligned}$$

Using again the Lipschitz property and the inequality $\epsilon \le M/(6+3|\bar{v}|)$, we get

$$|\gamma^\mu(t)-\Gamma(t)| \ge \frac{\frac{2}{3}M - \epsilon(2+|\bar{v}|)}{L+|\bar{p}|}(t-t_0-\tau)-\delta\tau \ge \frac{M}{3(L+|\bar{p}|)}(t-t_0-\tau)-\delta\tau. \tag{3.18}$$

Since $\delta < M/[6(L+|p|)]$, this means that $\gamma^\mu(t)$ stays outside K_δ for $t-t_0 > 3\tau$. □

We can now conclude the proof. Suppose that there is a sequence $t_i \downarrow t_0$ such that $(t_i, \bar{\gamma}(t_i)) \in K_\delta$. Then for all sufficiently small μ Lemma 4 implies that $(t, \gamma^\mu(t)) \notin K_\delta$ when $t > t_0 + 3(t_i - t_0)$ for all i, which means in turn that $(t, \bar{\gamma}(t))$ also cannot belong to K_δ for such t. As $t_i \downarrow t_0$, the trajectory $\bar{\gamma}$ has to stay outside K_δ for all $t_0 < t < t_0 + T$, a contradiction with what has been assumed. This proves Theorem 2. □

4. Integral curves of the field of admissible velocities

4.1. **The issue of existence and uniqueness for integral curves.** We have seen that limit trajectories $\bar{\gamma}$ of solutions γ^μ to the transport equation (3.2) are tangent in forward time to the discontinuous field of admissible velocities v^*, i.e., that $\dot{\bar{\gamma}}(t+0) = v^*(t,x)$ for any t and for any limit trajectory $\bar{\gamma}$ passing through some $x = \bar{\gamma}(t)$. This however does not imply that limit trajectories are unique.

There are in fact two different uniqueness problems: that for limit trajectories as $\mu \to 0$ for the viscous regularization (3.2), repeated here for convenience:

$$\dot{\gamma}^\mu(t) = \nabla_p H(t, \gamma^\mu, \nabla\phi^\mu(t, \gamma^\mu)), \tag{4.1}$$

and that for integral curves of the differential equation

$$\dot{\gamma}(t+0) = v^*(t, \gamma). \tag{4.2}$$

Since any limit trajectory of (4.1) is an integral curve of (4.2) according to Theorem 2, uniqueness for integral curves would imply uniqueness for limit trajectories. However, it is *a priori* possbile that more than one integral curve passes through the same singular point, but the vanishing viscosity regularization selects only one among these curves as a limit trajectory.

Uniqueness of limit trajectories can be established in the case when the Hamiltonian is quadratic in the momentum variable. This follows from a particular differential inequality for the squared separation between two close trajectories, first established in [16], which allows to control the expansion of the distance in terms of the semiconcavity constant of the solution ϕ. Indeed, two limit trajectories passing through the same point (t,x) cannot diverge by a finite distance in finite time, because their viscous regularizations must stay arbitrarily close to one another over this time interval, provided these regularizations are close enough at time t. Hence, as observed in [13, 14], the limit flow $\bar{\gamma}$ is defined uniquely and is therefore *coalescing*: once two trajectories intersect, they stay together at all later times.

However an analogue of the key differential inequality is not known for arbitrary convex Hamiltonians. Observe also that the argument just outlined bypasses the issue of integral curves altogether. It is therefore interesting to consider the existence and uniqueness issues for the differential equation (4.2) irrespective of viscous regularization. Here we will restrict ourselves to formal arguments based on rather generous regularity assumptions and aiming to convey the intuition of what is going on.

Let the shock manifold of ϕ be locally finite, i.e., suppose that at each shock point (t_0, x_0) there is a finite number k of minimizers connecting that point with the initial data. This implies that in a neighbourhood of (t_0, x_0) the solution ϕ may be represented locally as a pointwise minimum of a finite number k of C^2 smooth branches ϕ_i, each of which satisfies the Hamilton–Jacobi equation classically, and that locally the shock manifold is composed of C^1 smooth pieces of different dimensions. (The case of *preshocks*, introduced on p. 7, provides an exception to the condition of smoothness and should be considered separately.)

One can show that on each smooth piece of the shock manifold the spacetime field $(1, v^*)$ determined by admissible velocities is a Lipschitz vector field tangent to the piece. The usual ODE arguments then show that the flow generated by the vector field v^* is uniquely defined on smooth pieces of the shock manifold, as well as in the bulk where the solution ϕ is smooth.

In Section 2.2 it was shown that at a shock point (t_0, x_0) not all of the intersecting branches ϕ_i of solution are relevant for the integral curve γ at times $t > t_0$, but only those with $i \in I(v^*)$, i.e., those that are relevant in the first-order (linear) approximation to both the solution ϕ and the integral curve γ. Denote the corresponding index set with $\mathfrak{I}_1 := I(v^*)$.

Uniqueness of integral curves can only fail at shock connections: there must be at least two pieces of shock manifold that have a common point (t_0, x_0) and share the same tangent spacetime direction $(1, v^*)$ but at later times carry two disjoint trajectories both issued from x_0 at time t_0 with velocity v^*. Note that this is not possible if $|\mathfrak{I}_1| \leq d + 1$, where d is the spatial dimension, and the velocities v_i are in general position: indeed, in this situation removal of any branch ϕ_i with $i \in \mathfrak{I}_1$ would change the admissible velocity v^*.

In fact a (formal) perturbative analysis of an integral curve γ in higher orders of approximation reveals a nested sequence of finite index sets $\mathfrak{I}_1 \supseteq \mathfrak{I}_2 \supseteq \dots$ such that $\mathfrak{I}_k$ lists branches relevant for the integral curve in kth order, and the intersection $\mathfrak{I} = \cap_{k \geq 1} \mathfrak{I}_k$ is not empty (i.e., the sequence stabilizes). In particular if $|\mathfrak{I}| \leq d + 1$, then the integral curve γ is defined uniquely.

In what follows we illustrate this procedure in the second order and obtain $\mathfrak{I}_2$.

4.2. **Admissibility in the second order of perturbation theory.** Take an integral curve γ such that $\gamma(t_0) = x_0$ and assume it to be twice differentiable at t_0:

$$\gamma(t) = x_0 + (t - t_0)v^* + \frac{(t - t_0)^2}{2} a + o\big((t - t_0)^2\big), \tag{4.3}$$

where v^* is the vector of admissible velocity at (t_0, x_0) and a is the yet unknown acceleration of γ at t_0. At times $t = t_0 + \tau$ with sufficiently small $\tau > 0$ the point $\gamma(t)$ lies at intersection of a possibly smaller set of branches ϕ_i, which all have the same value at $(t, \gamma(t))$. The first two time derivatives of this common value along γ can be expressed as follows.

Using the Hamilton–Jacobi equation (1.1) and denoting $p_i^\gamma(t) = \nabla\phi_i(t, \gamma(t))$, for the first time derivative we get

$$\dot{\phi}_i(t, \gamma(t)) = \partial_t\phi_i(t, \gamma(t)) + \dot{\gamma}(t) \cdot \nabla\phi_i(t, \gamma(t)) = \dot{\gamma}(t) \cdot p_i^\gamma(t) - H(t, \gamma(t), p_i^\gamma(t)). \tag{4.4}$$

In particular

$$\dot{\phi}_i(t_0, x_0) = v^* \cdot p_i - H_i, \tag{4.5}$$

where $p_i = p_i^\gamma(t_0)$ and $H_i = H(t_0, x_0, p_i)$ as above. Using the Legendre duality (see (1.6) and discussion thereafter), we can modify expression (4.4) as follows:

$$\begin{aligned}\dot{\phi}_i(t, \bar{\gamma}(t)) &= \dot{\gamma}(t) \cdot p_i^\gamma(t) - H(t, \gamma(t), p_i^\gamma(t)) \\ &= (\dot{\gamma}(t) - v_i^\gamma(t)) \cdot \nabla_v L(t, \gamma(t), v_i^\gamma(t)) + L(t, \gamma(t), v_i^\gamma(t)),\end{aligned} \tag{4.6}$$

where $v_i^\gamma(t) = \nabla_p H(t, \gamma, \nabla\phi_i(t, \gamma(t)))$ and $p_i^\gamma(t) = \nabla_v L(t, \gamma(t), v_i^\gamma(t)) = \nabla\phi_i(t, \gamma(t))$ are values of velocity and momentum that correspond to the gradient $p_i^\gamma(t)$ along the curve γ. Recalling the expression for Bregman divergence (2.15)

$$D_L^{t,x}(v^* \mid v) = L(t,x,v^*) - L(t,x,v) - (v^* - v)\cdot\nabla_v L(t,x,v), \tag{4.7}$$

we can now express the time derivative $\dot\phi_i(t,\gamma(t))$ in the form

$$\dot\phi_i(t,\gamma(t)) = L(t,\gamma(t),\dot\gamma(t)) - D_L^{t,\gamma(t)}(\dot\gamma(t) \mid v_i^\gamma(t)). \tag{4.8}$$

Observe that the difference between $\phi_i(t,\gamma(t))$ and the mechanical action along the curve $\gamma(\cdot)$ decreases as the (negative) integral over (t_0,t) of the Bregman divergence $D_L^{\cdot,\bar\gamma(\cdot)}(\dot\gamma \mid v_i^\gamma)$. Of course subtracting the common quantity from the values of branches $\phi_i(t,\gamma(t))$ for all i does not change the mutual order of these values. We notice that the bigger is the Bregman divergence $D_L^{\cdot,\gamma(\cdot)}(\dot\gamma \mid v_i^\gamma)$, the faster decreases this difference: up to the second order in $t-t_0$, the value $\min_i \phi_i$ will be attained at the branch or branches for which $\dot\gamma(t)$ is the most distant (in the Bregman sense) from $v_i^\gamma(t)$.

To obtain the second time derivative we differentiate the r.h.s. of (4.4) to get

$$\begin{aligned}\ddot\phi_i(t,\gamma(t)) &= \ddot\gamma(t)\cdot p_i^\gamma(t) + \dot\gamma(t)\cdot\dot p_i^\gamma(t) - v_i^\gamma(t)\cdot\dot p_i^\gamma(t)\\ &\quad - \Big[\frac{\partial}{\partial t}H(t,\gamma(t),p_i^\gamma(t)) + \dot\gamma(t)\cdot\nabla_x H(t,\gamma(t),p_i^\gamma(t))\Big].\end{aligned} \tag{4.9}$$

It is convenient again to consider the second time derivative not of ϕ_i itself, but of the difference between ϕ_i and the mechanical action of γ:

$$\begin{aligned}&\ddot\phi_i(t,\gamma(t)) - \frac{\mathrm{d}}{\mathrm{d}t}L(t,\gamma(t),\dot\gamma(t))\\ &\quad = \ddot\gamma(t)\cdot(p_i^\gamma(t) - p_*^\gamma(t)) + (\dot\gamma(t) - v_i^\gamma(t))\cdot\dot p_i^\gamma(t)\\ &\quad - \Big[\frac{\partial}{\partial t}H(t,\gamma(t),p_i^\gamma(t)) + \dot\gamma(t)\cdot\nabla_x H(t,\gamma(t),p_i^\gamma(t))\Big]\\ &\quad - \Big[\frac{\partial}{\partial t}L(t,\gamma(t),\dot\gamma(t)) + \dot\gamma(t)\cdot\nabla_x L(t,\gamma(t),\dot\gamma(t))\Big],\end{aligned} \tag{4.10}$$

where $p_*^\gamma(t) = \nabla_v L(t,\gamma(t),\dot\gamma(t))$ is the value of momentum corresponding to the velocity $\dot\gamma(t)$. In particular at time t_0 we have

$$\ddot\phi_i - \frac{\mathrm{d}L}{\mathrm{d}t} = a\cdot(p_i - p^*) + (v^* - v_i)\cdot f_i - ([H]_i + [L]_i), \tag{4.11}$$

where $p^* = p_*^\gamma(t_0)$ is the usual admissible momentum (cf. (2.8)), $v_i = v_i^\gamma(t_0)$, $f_i = \dot p_i^\gamma(t_0)$, and $[H]_i$, $[L]_i$ denote values of the two square brackets at $t = t_0$.

Consider now an integral curve γ that is determined by intersection of smooth branches ϕ_i for some $i \in \mathfrak{I}$. Two conditions must hold for small $t - t_0 > 0$ along this curve:

(i) the velocity $\dot\gamma(t)$ must be admissible at $(t,\gamma(t))$, i.e., be the center of the "Bregman sphere" containing all $v_i^\gamma(t)$ at its boundary;
(ii) values of the remaining branches at $(t,\gamma(t))$ must be greater than the common value of $\phi_i(t,\gamma(t))$.

Define the piecewise linear concave function

$$F(a) = \min_{i\in\mathfrak{I}}\Big(a\cdot(p_i - p^*) + (v^* - v_i)\cdot f_i - ([H]_i + [L]_i)\Big). \tag{4.12}$$

Note that the velocity v^* of the curve γ at time t_0 is known, and therefore the values of f_i are the same for any integral curve γ, so the function F can be defined without knowing the curve γ. It is easy to see that the set $\mathfrak{I}(a)$ of indices where minimum is attained in (4.12) consists of precisely those indices for which condition (ii) holds. This set plays the same role in the quadratic approximation as did the set $I^*(v)$ in the linear approximation.

Condition (i) then becomes an admissibility condition for the acceleration similar to (2.8). Geometrically, the admissible acceleration a is the value at time t_0 of the rate of change of the center of the smallest Bregman sphere containing all $v_i(t)$ for sufficiently small $t - t_0$; compare this description with the fact that the velocity $\dot{\gamma}(t)$ is given by this center itself. It is clear that depending on the rates $\dot{v}_i$ (or equivalently, the values $\dot{p}_i = f_i$) at time t_0, some of the velocities present at t_0 may "sink" into the interior of the Bregman sphere for small $\tau = t - t_0 > 0$, leaving its surface defined by a smaller set $\{v_i : i \in \mathfrak{I}_2\}$.

In a similar way one can define the index sets $\mathfrak{I}_3$, $\mathfrak{I}_4$, and so on. Notice that this decreasing sequence of index sets will stabilize, since their intersection is nonempty. We conjecture that the resulting set $\mathfrak{I} = \cap_{s \geq 1} \mathfrak{I}_s$ determines the smooth manifold to which the integral curve γ belongs and which determines it uniquely as the integral curve of the corresponding filed of admissible velocities.

References

[1] J. Bec and K. Khanin, *Burgers turbulence*, Physics Reports **447** (2007), no. 1–2, 1–66, available at `arXiv:0704.1611`. MR2318582 ↑1, 7

[2] E. Hopf, *The partial differential equation $u_t + uu_x = \mu u_{xx}$*, Comm. Pure Appl. Math. **3** (1950), 201–230. MR0047234 ↑2

[3] S. N. Kružkov, *Generalized solutions of Hamilton–Jacobi equations of eikonal type: Statement of the problems; existence, uniqueness and stability theorems; certain properties of the solutions*, Mat. Sb. **98(140)** (1975), no. 3(11), 450–493 (Russian). MR0404870 ↑2

[4] P.-L. Lions, *Generalized solutions of Hamilton–Jacobi equations*, Research Notes in Mathematics, vol. 69, Pitman (Advanced Publishing Program), Boston, Mass.–London, 1982. MR0667669 ↑2, 4, 10

[5] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.) **27**, no. 1, 1–67, available at `arXiv:math/9207212`. MR1118699 ↑2

[6] P. D. Lax, *Weak solutions of nonlinear hyperbolic equations and their numerical computaton*, Comm. Pure Appl. Math. **7** (1954), 159–193. MR0066040 ↑2

[7] ———, *Hyperbolic systems of conservation laws.* II, Comm. Pure Appl. Math. **10** (1957), 537–566. MR0093653 ↑2

[8] O. A. Oleĭnik, *On Cauchy's problem for nonlinear equations in a class of discontinuous functions*, Doklady Akad. Nauk SSSR (N.S.) **95**, no. 3, 451–455 (Russian). MR0064258 ↑2

[9] W. H. Felming and H. M. Soner, *Controlled Markov processes and viscosity solutions*, 2nd ed., Stochastic Modelling and Applied Probability, vol. 25, Springer, New York, 2006. MR2179357 ↑2

[10] A. Fathi, *Weak KAM theorem in Lagrangian dynamics*, Cambridge Univ. Press, Cambridge, to appear. ↑2, 3, 5

[11] E W., K. Khanin, A. Mazel, and Ya. Sinai, *Invariant measures for Burgers equation with stochastic forcing*, Ann. of Math. (2) **151** (2000), no. 3, 877–960, available at `arXiv:math/0005306`. MR1779561 ↑3

[12] C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*, 2nd ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 325, Springer, Berlin, 2005. MR2169977 ↑4

[13] I. A. Bogaevsky, *Matter evolution in Burgulence* (July 29, 2004), available at `arXiv:math-ph/0407073`. ↑4, 5, 6, 9, 10, 14

[14] I. A. Bogaevsky, *Discontinuous gradient differential equations and trajectories in the calculus of variations*, Mat. Sb. **197** (2006), no. 12, 11–42; English transl., Sbornik Math. **197** (2006), no. 12, 1723–1751. ↑4, 5, 9, 10, 14

[15] S. Gurbatov and S. Shandarin, 2005. Private communication. ↑5

[16] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematical Studies, vol. 5, North-Holland, Amsterdam, 1973 (French). MR0348562 ↑5, 14

[17] P. Cannarsa and C. Sinestrari, *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, Progress in Nonlinear Differential Equations and their Applications, vol. 58, Birkhäuser Boston, Inc., Boston, MA, 2004. MR2041617 ↑3, 6, 7, 8, 12

[18] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, vol. 28, Princeton Univ. Press, Princeton, NJ, 1970. MR0274683 ↑7, 12

[19] I. A. Bogaevsky, *Perestroikas of shocks and singularities of minimum functions*, Physica D: Nonlinear Phenomena **173** (Dec. 1, 2002), no. 1–2, 1–28, available at `arXiv:0204237`. MR1945478 ↑7

[20] P. Cannarsa and Y. Yu, *Singular dynamics for semiconcave functions*, J. European Math. Soc. **11** (2009), 999-1024. ↑8

[21] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Applied Mathematics, vol. 5, SIAM, 1990. ↑9

[22] L. M. Brègman, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming*, Ž. Vyčisl. Mat. i Mat. Fiz. **7** (1967), 620–631 (Russian); English transl., USSR Comp. Math. and Math. Phys. **7** (1967), 200–217. MR0215617 ↑10

[23] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161. MR1440931 ↑6

[24] C. Villani, *Optimal transport*: *Old and new*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer, Berlin, 2009. MR2459454 ↑6

[25] I. A. Bogaevsky, *Perestroikas of shocks in optimal control*, J. Math. Sciences **126** (2005), no. 4, 1229–1242. ↑

[26] B. A. Khesin and G. Misiołek, *Shock waves for the Burgers equation and curvatures of diffeomorphismgroups*, Analysis and Singularities. Part 2. Collected Papers, Tr. Mat. Inst. Steklova, vol. 259, Nauka, Moscow, 2007, pp. 77–85, available at `arXiv:math/0702196`. MR2433678 ↑6

[27] J.-P. Aubin and A. Cellina, *Differential inclusions*: *Set-valued maps and viability theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 264, Springer, Berlin, 1984. MR0755330 ↑6

Department of Mathematics, University of Toronto, Toronto, Ontario, Canada
E-mail address: `khanin@math.toronto.edu`

The Russian Academy of Sciences, Institute for Information Transmission Problems (Kharkevich Institute), Moscow, Russia

National Research University Higher School of Economics
E-mail address: `sobolevski@iitp.ru`