University Of Helsinki
Department Of Mathematics and Statistics

Hamilton-Jacobi Equations

Master's Thesis

Nikita Skourat

supervised by
Pr. Kari Astala

May 31, 2017

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty

Matemaattis-luonnontieteellinen

Laitos — Institution — Department

Matematiikan ja tilastotieteen laitos

Tekijä — Författare — Author

Nikita Skourat

Työn nimi — Arbetets titel — Title

Hamilton-Jacobi Equations

Oppiaine — Läroämne — Subject

Matematiikka

Työn laji — Arbetets art — Level

Pro gradu-tutkielma

Aika — Datum — Month and year

Toukokuu 2017

Sivumäärä — Sidoantal — Number of pages

71 s.

Tiivistelmä — Referat — Abstract

This thesis presents the theory of Hamilton-Jacobi equations. It is first shown how the equation is derived from the Lagrangian mechanics, then the traditional methods for searching for the solution are presented, where the Hopf-Lax formula along with the appropriate notion of the weak solution is defined. Later the flaws of this approach are remarked and the new notion of viscosity solutions is introduced in connection with Hamilton-Jacobi equation. The important properties of the viscosity solution, such as consistency with the classical solution and the stability are proved. The introduction into the control theory is presented, in which the Hamilton-Jacobi-Bellman equation is introduced along with the existence theorem. Finally multiple numerical methods are introduced and aligned with the theory of viscosity solutions.

The knowledge of the theory of partial differential analysis, calculus and real analysis will be helpful.

Avainsanat — Nyckelord — Keywords

Hamilton-Jacobi Equation, PDE, Classical Mechanics, Viscosity Solutions

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpulan tiedekirjasto

Muita tietoja — Övriga uppgifter — Additional information

Contents

Introduction

The topic of this thesis is the Hamilton-Jacobi equation. This equation is an example of a partial differential equations that arises in many different areas of physics and mathematics, and hence has an utmost importance in science. Originally, it was derived by William Rowen Hamilton in his essays [15] and [16] as a formulation of classical mechanics for a system of moving particles. Hamilton extended the Lagrangian formulation, using the calculus of variations, thus deriving the following equation:

$$\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial q_1}, ..., \frac{\partial S}{\partial q_d}, q_1, ..., q_d\right) = 0.$$

Here H was a Hamiltonian - a descriptor of the system, defined by Hamilton, and q_i are the generalized coordinates. The solution to the equation the classical action $S = \int_0^t (U + T) ds$, also called the principal function, where U is a potential and T is a kinetic energy of the system. However, mathematically Hamilton's arguments contained some flaws. For example, Hamilton assumed that the equation would hold only in case of conservative systems (satisfying the canonical Hamilton's equations). Jacobi showed later in his papers, that such an assumption is not necessary and also provided mathematically meaningful adjustments to the equation. That's why the modern name for the equation is *Hamilton-Jacobi equation*.

Later, another version of the derivation of the equation was presented. It showed that the principal function S can be seen as a generating function for a canonical transformation of coordinates. In this thesis we'll follow this approach.

In mathematics, the similar equation appeared in control theory in 50s, when Richard Bellman developed a dynamic programming method in [2] with his colleagues. This so-called Hamilton-Jacobi-Bellman equation defines a broader class of equations, than the version from the classical mechanics.

Mathematicians have been working on solving this equations for many years. It appeared, that this equation in general has no classical solutions, hence first, some notions of weak solution was developed by Hopf ([17]) and Lax([21]). Their investigations led to the important Hopf-Lax formula, that can be used for calculating the solution or studying its properties.

Finally, in the 80s, the new notion of viscosity solution was developed for the Hamilton-Jacobi equation and extended to a broader class of partial differential equations. The first ideas about this weaker notion of solutions appeared in Evans' paper [10] by applying the vanishing viscosity method to the equation, however the exact definition was given later by Crandall and Lions in [9] along with some very important properties.

The purpose of this thesis is to present the theory of Hamilton-Jacobi equations from multiple sides. In the first chapter we will show how this equation is derived by using the canonical transformation argument. In the second chapter the traditional approach to the existence and uniqueness theory is presented along with Hopf-Lax formula. And finally, in the third chapter we'll introduce the viscosity solutions from scratch: we'll first show how the application of the method of vanishing viscosity to the Hamilton-Jacobi equation leads to the definition, and then define it strictly and show, that such weak solutions are consistent and stable. Also we give a short introduction into the control theory and dynamic programming, thus also deriving the Hamilton-Jacobi-Bellman equation.
This thesis uses Chapters 3.3 and 10 from [11] as a skeleton. The prior knowledge of theory of partial differentiable equations, real analysis is assumed.

1 Hamilton-Jacobi Equations. Derivation and Basic Properties

The purpose of this thesis is to study the initial value problem for the following Hamilton-Jacobi equation:

$$\begin{cases} u_t + H(D_x u, x) = 0 \\ u(x,0) = g. \end{cases} \tag{1.1}$$

Here $u : \mathbb{R}^d \times [0,\infty) \to \mathbb{R}, u = u(x,t)$ is an unknown function, $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, H = H(p,x)$ is a Hamiltonian (we'll define it later) and $g : \mathbb{R}^d \to \mathbb{R}, g = g(x)$ is the initial condition. We denote by $D_x u$ a spatial gradient, i.e. $D_x u = (\frac{\partial u}{\partial x_1}, ..., \frac{\partial u}{\partial x_d})$.
In this section we'll show how this equation is derived from Lagrangian mechanics. We'll first derive the formulation of Lagrangian mechanics, using the principle of least action, then we'll show how Hamilton's ODE are derived and finally we'll show how Hamilton-Jacobi equation comes in place.

Remark. *The notation and definitions presented in this paper may be different from the ones in mechanics books. It was chosen as it better suits the mathematical PDE theory, that we're studying in this thesis. You can read more about Hamilton's principle and Hamilton's ODEs from the physical perspective in §2 and §40 in [20], or Chapter 9 in [1].*

1.1 Hamilton's Principle

Hamiltonian mechanics is one of the several formulations of classical mechanics, that uses generalised coordinates $x = (x_1, ..., x_d)$ as one of the parameters. We know from the Newtonian formulation, that to define the position of a system of N moving particles[1] in three-dimensional Euclidian space, we need to specify N vectors, which requires a total pf $3N$ coordinates. This number of coordinates is also called the number of *degrees of freedom*. The generalised coordinates are some number of independent quantities, that are needed to describe the motion of system of particles ($3N$ in the example above). Note, that these quantities must not necessarily be Cartesian coordinates and the optimal choice often depends on the given problem.
In this paper, we are looking at the motion law for a system of particles with d degrees of freedom, which is mathematically described as a path $x : [0, \infty] \to \mathbb{R}^d$, i.e. $x_j = x_j(t), j = 1...d$. Our first goal is to derive the Euler-Lagrange equations for this law. Before we state the Hamilton's principle and its consequences, let's give some definitions first:

Definition 1.1. *The* Lagrangian *of the system of moving particles is a definite smooth function $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, denoted as $L = L(q, x)$, that completely characterizes the conditions of the movement.*

Lagrangian is the central notion of the so-called Lagrangian mechanics. In general, there's no single expression to calculate the Lagrangian, and any function, that generates the physically correct motion equations, can be taken as a Lagrangian. There're however some wide classes of problems, where the Lagrangian can be easily constructed. For example, in non-relativistic case, the Lagrangian for the moving system of particles can be defined as $L = T - V$, where T is the total kinetic energy of the system and V is the potential energy. More about it can be found in [24].

Definition 1.2. *The* action functional *is the following integral:*

$$I[x(\cdot)] = \int_0^t L(\dot{x}(s), x(s))\mathrm{d}s, \tag{1.2}$$

defined for functions $x(s) = (x_1(s), ...x_d(s))$, belonging to the admissible class $\mathcal{A}$:

$$\mathcal{A} = \{u : [0, t] \to \mathbb{R}^d, u \in C^2([0, t])|u(0) = y, u(t) = x\} \tag{1.3}$$

[1]In this paper we mean by a particle a physical body whose dimensions may be neglected in describing its motion.

The Hamilton's principle in mechanics states, that the particle moves between points y and x in a way, that the action integral takes the least possible value. This principle is also called *the principle of least action*.
According to this, to find the law of motion of the particle, we have the following calculus of variation problem: we want to find a C^2-curve, that minimizes the action functional among all admissible curves $w \in \mathcal{A}$:

$$I[x(\cdot)] = \min_{w \in \mathcal{A}} \int_0^t L(\dot{w}(s), w(s))\, ds, \tag{1.4}$$

To approach this minimization problem, we can solve the corresponding Euler-Lagrange equations of the variational problem. In our case it's a system of d second order ordinary differential equations.

Theorem 1.1. *Assume there exists a function $x \in \mathcal{A}$, that solves the minimization problem (1.4). Then this function satisfies the following Euler-Lagrange equation:*

$$-\frac{d}{ds}(D_q L(\dot{x}(s), x(s))) + D_x L(\dot{x}(s), x(s)) = 0 \tag{1.5}$$

Proof. Let's choose a smooth function $v : [0, t] \to \mathbb{R}^d$, such that $v(0) = v(t) = 0$. For $\tau \in \mathbb{R}$ we define a function:

$$w(\cdot) := x(\cdot) + \tau v(\cdot) \tag{1.6}$$

Clearly, $w \in \mathcal{A}$ and by definition of x, we have that $I[x(\cdot)] \leq I[w(\cdot)]$. Hence, the real valued function i, defined as:

$$i(\tau) := I[x(\cdot) + \tau v(\cdot)], \tag{1.7}$$

attains a minimum at $\tau = 0$. Hence, if $i'(\tau)$ exists, we have that $i'(\tau) = 0$. Let's calculate this derivative explicitly.

$$i(\tau) = \int_0^t L(\dot{x}(s) + \tau \dot{v}(s), x(s) + \tau v(s))\, ds,$$

Using Leibniz integral rule:

$$\begin{aligned} i'(\tau) = \int_0^t &(D_q L(\dot{x}(s) + \tau \dot{v}(s), x(s) + \tau v(s)) \cdot \dot{v}(s) \\ &+ D_x L(\dot{x}(s) + \tau \dot{v}(s), x(s) + \tau v(s)) \cdot v(s))\, ds. \end{aligned}$$

Next, let $\tau = 0$:

$$0 = i'(0) = \int_0^t (D_q L(\dot{x}(s), x(s)) \cdot \dot{v}(s) + D_x L(\dot{x}(s), x(s)) \cdot v(s))\, ds \tag{1.8}$$

Finally, notice that we can simplify the first part of the integrand with integration by parts, taking into account the that v vanishes at the integration limits:

$$\begin{aligned}\int_0^t D_q L(\dot{x}(s),x(s)) \cdot \dot{v}(s)\,ds &= \sum_{i=1}^{d} \int_0^t (D_{q_i} L(\dot{x}(s), x(s))\dot{v}_i(s)\,ds \\ &= \sum_{i=1}^{d} \int_0^t (-\frac{d}{ds} D_{q_i} L(\dot{x}(s), x(s)) v_i(s)\,ds \\ &= \int_0^t (-\frac{d}{ds} D_q L(\dot{x}(s), x(s)) \cdot v(s)\,ds\end{aligned}$$

Now passing this result back to (1.8), we obtain:

$$0 = \int_0^t (-\frac{d}{ds}(D_q L(\dot{x}(s), x(s)) + D_x L(\dot{x}(s), x(s))) \cdot v(s)\,ds \tag{1.9}$$

Due to the fact, that v was an arbitrary smooth function, we must have, that:

$$-\frac{d}{ds}(D_q L(\dot{x}(s), x(s))) + D_x L(\dot{x}(s), x(s)) = 0$$

□

Euler-Lagrange equations with the corresponding initial values constitute the formulation of the classical mechanics, that uses generalized coordinates and generalized velocities as parameters.

Remark. *Note, that not every solution to Euler-Lagrange equation is a minimizer of the action functional. The solution to the Euler-Lagrange equation is called a* critical point *of the corresponding functional. So, what the proven theorem says, is that every minimizer is a critical point of $I[x(\cdot)]$, but a critical point need not be a minimizer. This is similar to Fermat's theorem in calculus, which states that at any point, where the differential function attains a local extremum, its derivative is zero.*

Remark. *Theorem (1.1) doesn't prove the existence of the minimizer. However, under some assumptions, we can ensure the existence. According to the theory of calculus of variation, the minimizer exists, if the Lagrangian L is coercive and convex in q. More about this fact can be found in [11], Chapter 8.2.*

1.2 Hamilton's ODE

As the second step of the construction, we convert the system of Euler-Lagrange equations into Hamilton's equations. We assume that x is a C^2 function, that is a critical point of the action functional. We define first the generalized momentum as a function $p : [0, t] \to \mathbb{R}^d$, that will serve as the second parameter for the Hamiltonian formulation of mechanics:

$$p(s) := D_q L(\dot{x}(s), x(s)). \tag{1.10}$$

Definition 1.3. *Assume, that for all $x, p \in \mathbb{R}^d$ the following equation:*

$$p = D_q L(q, x) \tag{1.11}$$

can be uniquely solved for q as a smooth function of p, x, $q = q(p, x)$.
The Hamiltonian, *associated with the Lagrangian L is the function $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, defined as:*

$$H(p, x) := p \cdot q(p, x) - L(q(p, x), x), \tag{1.12}$$

where q is defined implicitly by (1.11)

Example 1.1. *Remember what we told before, that for a moving particle Lagrangian is $L = T - V$, or, after expanding:*

$$L(q, x) = \frac{1}{2} m |q|^2 - V(x).$$

Now, since we are using the Cartesian coordinates, we know, that in this case the momentum is $p = mq$, so we can calculate the Hamiltonian by the above definition:

$$H(p, x) = \frac{1}{m} p \cdot p - L(q, x) = \frac{1}{m}|p|^2 - \frac{1}{2m}|p|^2 + V(x) = \frac{1}{2m}|p|^2 + V(x)$$

The first term in the sum is just the kinetic energy, written via the momentum. So in this case the Hamiltonian equals the total energy of the particle. However, this is not always the case, and it happens only if the Cartesian coordinates can be written in terms of the generalized coordinates without time dependence and vice versa. Otherwise the expression gets more complicated.

Now that we've defined the Hamiltonian and generalized momentum in terms of Lagrangian and generalized velocity, we are ready to derive the Hamilton's ODEs.

Theorem 1.2. *The functions x and p (defined as in (1.10)), satisfy Hamilton's ODE system:*

$$\begin{cases} \dot{p}(s) = -D_x H(p(s), x(s)) \\ \dot{x}(s) = D_p H(p(s), x(s)) \end{cases} \tag{1.13}$$

for $0 \leq s \leq t$. Moreover, the mapping $s \mapsto H(p(s), x(s)$ is constant.

This is the system of $2d$ first-order differential equations.

Proof. Remember, in the definition (1.3), the generalized velocity q is uniquely defined by the equation $p = D_q L(q(p, x), x)$. Then from (1.10) it follows that $\dot{x}(s) = q(p(s), x(s))$. We calculate the corresponding gradients of H from its definition (1.12):

$$\begin{aligned} D_x H(p, x) &= p \cdot D_x q(p, x) - D_q L(q(p, x), x) \cdot D_x q(p, x) - D_x L(q(p, x), x) \\ &= (p - D_q L(q(p, x), x)) \cdot D_x q(p, x) - D_x L(q(p, x), x) \\ &\overset{(1.11)}{=} -D_x L(q(p, x), x) = -D_x L(\dot{x}(s), x(s)) \end{aligned}$$

Now if we remember the Euler-Lagrange equation (1.5) and the above, we get:

$$\begin{aligned} D_x H(p(s), x(s)) = -D_x L(\dot{x}(s), x(s)) = \\ = -\frac{d}{ds} D_q L(\dot{x}(s), x(s)) = -\dot{p}(s) \end{aligned}$$

Thus we obtains the first equation from (1.13). The second one is obtained in the same way:

$$\begin{aligned} D_p H(p, x) &= q(p, x) + p \cdot D_p q(p, x) - D_q L(q(p, x), x) \cdot D_p q(p, x) \\ &= q(p, x) + (p - D_q L(q(p, x), x)) \cdot D_p q(p, x)) \\ &\overset{(1.11)}{=} q(p, x) = \dot{x}(s) \end{aligned}$$

Finally, using this, we can easily prove that the mapping $s \mapsto H(p(s), x(s)$ is constant:

$$\begin{aligned} &\frac{d}{ds} H(p(s), x(s)) = D_p H(p(s), x(s)) \cdot \dot{p}(s) + D_x H(p(s), x(s)) \cdot \dot{x}(s) \\ &= D_p H(p(s), x(s)) \cdot (-D_x H(p(s), x(s))) + D_x H(p(s), x(s)) \cdot D_p H(p(s), x(s)) \\ &= 0 \end{aligned}$$

□

Unlike Lagrangian formulation, the Hamiltonian ODEs are coupled and the order of the equations is one, not two, hence mathematically it can be easier to solve them in some cases. But there's still a question, where the Hamiltonian formalism can be useful and why it was introduced in the first place? Let us list some of the reasons:

1. Hamiltonian is a physical quantity, namely a total energy of the system, as described in the example (1.1) above, and therefore all parts of Hamilton's equations have a physical interpretation.

2. Geometrically, Hamilton's equations say, that flowing in time is equivalent to flowing along a vector field in phase space. That gives a geometrical picture of time evolutions in dynamic Hamiltonian systems.

3. In Hamiltonian mechanics one can use canonical transformations which allow to change coordinates in phase space which can make it easier to solve problem. (This will be the topic of the next chapter).

4. And probably the most important feature of Hamilton's equations is that it gives the easy way to quantize classical systems, hence it's possible to use in quantum mechanics. Moreover, it can be used in thermodynamics, statistical physics and many other branches, where Lagrangian formalism fails or becomes incomprehensible. Later we'll show, how one can derive the Schrödinger equation directly from the Hamiltonian formalism.

1.3 Canonical Transformations

As we showed above, the Hamiltonian mechanics uses $2d$ parameters to describe the system of moving particles. Sometimes, however, it makes sense to perform a change of variables to use different $2d$ parameters. This change in its most general form is described mathematically in the following way:

$$(p(t), x(t)) \mapsto (P(p, x, t), X(p, x, t)). \tag{1.14}$$

There are several requirements we want to put on this transformation. First, it should be invertible, so that we can obtain the solution to the original Hamilton's equations $(p(t), x(t))$, once we know the solution $(P(t), X(t))$ to the transformed equation. Also, we want to have the equations for new parameters to be similar to the original Hamilton's equations.

Definition 1.4. *A map* $g : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d \times \mathbb{R}^d$ *is called a* canonical transformation*, if it preserves the canonical form of the Hamilton's equations, i.e. if* $g = (P, X)$*,* *where $P = P(p,x,t)$ and $X = X(p,x,t)$, and the Hamiltonian in new coordinates is denoted as $K(P,X) = H(p(P,X), x(P,X))$, then the Hamilton's equations in new coordinates have the following form:*

$$\begin{cases} \dot{P}(t) = -D_X K(P(t), X(t)) \\ \dot{X}(t) = D_P K(P(t), X(t)) \end{cases} \tag{1.15}$$

Now let us remind ourselves about Hamilton's principle. The path $x(t)$ minimizes the action functional, thus solving the variational problem (1.4). In terms of calculus of variation, it means, that $x(t)$ is a *stationary point* of the action functional, meaning that

$$\delta I[x(\cdot)] = \delta \int_0^t L(\dot{x}(s), x(s))\, ds = 0,$$

where δ denotes the functional derivative. In physics this version is often referred to as *the Hamilton's principle in configuration space*. We want to rewrite it via the Hamiltonian. Let $p(s) = D_q L(\dot{x}(s), x(s))$. Then from the definition 1.3 it follows, that $H(p(s), x(s)) = p(s)\cdot\dot{x}(s) - L(\dot{x}(s), x(s))$. We can rewrite the variational relation above now:

$$\delta I[x(\cdot)] = \delta \int_0^t [p(s)\cdot\dot{x}(s) - H(p(s), x(s))]\, ds = 0. \tag{1.16}$$

This version is called *the Hamilton's principle in phase space*. Next we want to apply the canonical transformation and expect the Hamilton's principle to hold in the new coordinates, i.e. there exists some path $X(s)$ in new coordinates, such that

$$\delta I[X(\cdot)] = \delta \int_0^t [P(s)\cdot\dot{X}(s) - K(P(s), X(s))]\, ds = 0, \tag{1.17}$$

where $K(P,X)$ is the Hamiltonian in new coordinates. The expressions under the integral inside (1.16) and (1.17) need not be equal, however, due to the fact, that we have fixed endpoint values for $x(t), x(0)$ and hence $p(t), p(0)$, so there's a zero variation at endpoints, we must have the following relation between the integrands in old and new coordinates:

$$\alpha(p(s)\cdot\dot{x}(s) - H(p(s), x(s))) = P(s)\cdot\dot{X}(s) - K(P(s), X(s)) + \frac{du}{dt}|_{t=s}, \tag{1.18}$$

where u is some good enough function with zero variation on endpoints (this will ensure the simultaneous satisfaction of both (1.16) and (1.17)) and α is a scaling constant. From now on we consider $\alpha = 1$, as it's not relevation for our discussion. This function F is called *a generating function*, depending on the variables, on which it depends. Goldstein in [14] introduces four generic types of generating functions:

1. $u_1 := u_1(x, X, t)$.

2. $u_2 := u_2(P, x, t) - X \cdot P$.

3. $u_3 := u_3(p, X, t) + x \cdot p$.

4. $u_4 := u_4(p, P, t) + x \cdot p - X \cdot P$.

We are interested in transformation laws that appear in case of the generating function type-2. Let $u_2 = u(P, x, t) - X \cdot P$. We remark, that the term $-X \cdot P$ is added to get rid of $\dot{X}$ in (1.17). Let us see what happens (we'll omit the parameter s for the sake of clearer calculations):

$$\begin{aligned} p \cdot \dot{x} - H(p, x) &= P \cdot \dot{X} - K(P, X) + \left.\frac{d(u(P, x, t) - X \cdot P)}{dt}\right|_{t=s} \\ &= P \cdot \dot{X} - K(P, X) + u_t + D_P u \cdot \dot{P} + d_x u \cdot \dot{x} - P \cdot \dot{X} - \dot{P} \cdot X \\ &= -K(P, X) + u_t + D_P u \cdot \dot{P} + D_x u \cdot \dot{x} - X \cdot \dot{P}. \end{aligned}$$

Or, rewriting it:

$$(p - D_x u) \cdot \dot{x} + (X - D_P u) \cdot \dot{P} = H(p, x) - K(P, X) + u_t. \tag{1.19}$$

Separately the new and old coordinates are independent of each other, then the whole equation can be identically true, if the corresponding coefficients for $\dot{x}$ and $\dot{P}$ vanish, i.e. if the following is true:

$$\begin{aligned} p(s) &= D_x u(P(s), x(s), s) \\ X(s) &= D_P u(P(s), x(s), s). \end{aligned}$$

And that means, that the rest of the equation should also be equal to zero, thus giving us the formula for the Hamiltonian in the new coordinates and thus proving the following theorem.

Theorem 1.3. *If $(p, x, t) \mapsto (P(p, x, t), X(p, x, t), t)$ is a canonical transformation, generated by a function u, then the Hamiltonian in the new coordinates is:*

$$K(P, X, t) = H(p, x) + \partial_t u(P, x, t). \tag{1.20}$$

Generally the process of applying the transformation is defined by several steps:

1. We start with the function $u : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, $u := u(P, x, t)$.

2. We define the coordinates $p = D_x u$, $X = D_P u$.

3. Then $(p(t), x(t)) \mapsto (P(t), X(t))$ is a canonical transformation.

1.4 Hamilton-Jacobi Equation and Its Properties

Now we are searching for a specific canonical transformation, that will make the resulting Hamiltonian K to be 0. Then the Hamilton's equations in the new coordinates will be trivial:

$$\begin{cases} \dot{P}(t) = 0 \\ \dot{X}(t) = 0 \end{cases} \tag{1.21}$$

So we are looking for a generating function $u(P, x, t)$, that will produce such a transformation. By theorem (1.3), $K(P, X, t) = H(p, x) + \partial_t u(P, x, t)$, and by the definition of a type-2 generating function, $p = D_x u$. So, in order for the new Hamiltonian K to be zero, we must have:

$$\partial_t u(P, x, t) + H(D_x u(P, x, t), x) = 0.$$

However, P doesn't play any mathematical role in this equation, so for further investigation, we assume that u has no explicit dependence on P and the equation becomes the Hamilton-Jacobi equation:

$$\partial_t u(x, t) + H(D_x u(x, t), x) = 0. \tag{1.22}$$

The function u, as a solution to the Hamilton-Jacobi equation, is also called the *Hamilton principle function.* We'll show, that it's actually closely related to the action integral. Since u is a generating function, then $D_x u = p$. Also $\dot{x} = q$. Using these facts and the Hamilton-Jacobi equation, we can calculate the time derivative of u:

$$\frac{du}{dt} = D_x u \cdot \frac{dx}{dt} + \partial_t u = p \cdot \dot{x} + \partial_t u = p \cdot \dot{x} - H(p, x) = p \cdot q - H(p, x) = L(p, q) \tag{1.23}$$

So actually, u is a classical action plus some constant.

1.4.1 Method of Characteristics

There's another important connection of Hamilton-Jacobi equation and Hamilton's ODEs: apparently Hamilton's ODEs are characteristic equations of Hamilton-Jacobi PDE. We'll prove this fact later in this chapter. Let us first recall the method of characteristics.

Assume, we have a nonlinear first-order PDE with boundary conditions in the set $\Omega \subset \mathbb{R}^d$:

$$\begin{cases} F(Du, u, x) = 0 \text{ in } \Omega, \\ u = g \text{ on } \Gamma \subseteq \partial\Omega. \end{cases} \tag{1.24}$$

We assume, that $F = F(p, z, x)$ and g are given smooth functions. The idea of the method of characteristics if the following: we assume, that u solves the equation (1.24) and we fix $x \in \Omega$ and $x^0 \in \Gamma$. We know the value of u in x^0 because of the boundary condition and then want to find the value $u(x)$ by finding a curve inside Ω, that connects x and x^0 and along which we can calculate the value of u.
The curve mathematically is described as $x(s) = (x_0(s), ..., x_d(s))$. We define $z(s) = u(x(s))$ and $p(s) = Du(x(s))$. Let's first differentiate:

$$\dot{p}_i(s) = \sum_{j=1}^{n} u_{x_i x_j} \dot{x}_j(s). \tag{1.25}$$

Second, we differentiate the original equation (1.24) with respect to x_i:

$$\sum_{j=1}^{n} \frac{\partial F}{\partial p_j} u_{x_j x_i} + \frac{\partial F}{\partial z} u_{x_i} + \frac{\partial F}{\partial x_i} = 0 \tag{1.26}$$

Now we calculate this expression at $x = x(s)$:

$$\begin{aligned} &\sum_{j=1}^{n} \frac{\partial F}{\partial p_j}(p(s), z(s), x(s)) u_{x_j x_i}(x(s)) \\ &\quad = -\frac{\partial F}{\partial z}(p(s), z(s), x(s)) u_{x_i}(x(s) - \frac{\partial F}{\partial x_i}(p(s), z(s), x(s)). \end{aligned} \tag{1.27}$$

We want to get rid of the second derivatives in (1.25), so we require:

$$\dot{x}_j(s) = \frac{\partial F}{\partial p_j}(p(s), z(s), x(s)). \tag{1.28}$$

Now we can pass these results back to (1.25):

$$\dot{p}_i(s) = -\frac{\partial F}{\partial z}(p(s), z(s), x(s)) p_i(s) - \frac{\partial F}{\partial x_i}(p(s), z(s), x(s)) \tag{1.29}$$

Finally, we calculate the derivative of $z(s)$:

$$\begin{aligned} \dot{z}(s) = \sum_{j=1}^{n} \frac{\partial u}{\partial x_j}(x(s)) \dot{x}_j(s) &= \sum_{j=1}^{n} \dot{p}_j(s) \frac{\partial F}{\partial p_j}(p(s), z(s), x(s)) \\ &= D_p F(p(s), z(s), x(s)) \cdot p(s). \end{aligned} \tag{1.30}$$

Equations (1.28), (1.29) and (1.30) are the characteristic lines we were looking for. Let's rewrite them in the vector notation:

$$\begin{cases} \dot{p}(s) = -D_z F(p(s), z(s), x(s))p(s) - D_x F(p(s), z(s), x(s)) \\ \dot{z}(s) = D_p F(p(s), z(s), x(s)) \cdot p(s) \\ \dot{x}(s) = D_p F(p(s), z(s), x(s)). \end{cases} \tag{1.31}$$

It should be noted, that for these equations to be used, appropriate initial conditions should be found. More about it can be found in Chapter 3.2 in [11]. We, however, return to the Hamilton-Jacobi PDE.

Theorem 1.4. *Hamilton's equations are characteristic equations of Hamilton-Jacobi PDE.*

Proof. In the notation introduced above, we have the following situation:

$$F((D_x u, u_t), u, (x, t)) = u_t + H(D_x u, x) = 0.$$

Here we have $F(P, z, X) = q + H(p, x)$, where $P = (p, q)$ and $X = (x, t)$. Now we calculate:

$$D_P F = (D_p H, 1).$$

$$D_P F \cdot P = (D_p H, 1) \cdot (p, q) = D_p H \cdot p + q = D_p H \cdot p - H(p, x).$$

$$D_z F = 0.$$

$$D_X F = (D_x H, 0).$$

Now we can calculate these derivatives at $(y(s), z(s), y(s))$ and use (1.31) and we get the following system:

$$\begin{cases} \dot{P}(s) = -D_x H(p(s), x(s)) \\ \dot{z}(s) = D_p H(p(s), x(s)) - H(p(s), x(s)) \\ \dot{X}(s) = D_p H(p(s), x(s)). \end{cases}$$

The equations for $x(s)$ and $p(s)$ are exactly the Hamilton's ODEs, as we have derived them from the Lagrangian mechanics. Notice also, that the equation for $z(s)$ is trivial, once we find $x(s)$ and $p(s)$, because the function H doesn't directly depend on z, so we can immediately integrate the equation for $z(s)$. □

1.5 Applications

In this chapter we'll show some useful examples, that shows the relations of the Hamilton-Jacobi equation to such fields of physics as optics and quantum mechanics.

1.5.1 Eikonal Equation

If we consider the static process, then the Hamilton-Jacobi equation turns to the following form:

$$H(D_x u(x), x) = 0. \tag{1.32}$$

One example of such a static Hamilton-Jacobi equation arises in geometrical optics and is called the *eikonal equation*. Since there's no time dependency, the equation is usually considered in the bounded domain $\Omega \in \mathbb{R}^d$ and supplied with the boundary conditions:

$$\begin{cases} |D_x u(x)| = n(x) \text{ in } \Omega \\ u(x) = 0 \text{ on } \partial\Omega \end{cases} \tag{1.33}$$

This equation can be derived from the Maxwell's equations of electromagnetism and it provides a link between wave optics and ray optics. Or it can also be derived from the Fermat's principle of least time, which says that light travels between two points along the path that requires the least time, as compared to other nearby paths. Physically, the function u here describes the shortest time, needed for light to travel from the boundary $\partial\Omega$ to the point $x \in \Omega$, and $n(x)$ is the refractive index of a medium at point x.

1.5.2 Schrödiner Equation

We have shown previously, that u is a classical action. If we know u, then at any time t we can determine the isosurfaces[2] $u(x_0, t)$. The motion of this isosurface is determined by the motions of the particles, belonging to the system, that start at a point x_0. According to the Feynman path integral formulation of quantum mechanics, there're two postulates:

1. The probability that a particle has a path lying in a certain region of spacetime is the absolute square of a sum of contributions, one from each path in this region.

2. The paths contribute equally in magnitude, but the phase of the contributions is the classical action.

[2] An *isosurface* represents points of constant action within some volume of space.

Mathematically, this means, that the amplitude for a given path $x(s)$ is $Ae^{\frac{iu(x,t)}{\hbar}}$, where A is some real function and u is a classical action, performed by the system between times t_1 and t_2. And the total amplitude of a system is

$$K_{21} = \sum_n K_{21}^n = \sum_n A_{21}^n e^{\frac{iu_n}{\hbar}} \tag{1.34}$$

, where A_{21}^n and u_n correspond to the path $x_n(t)$. This function K is called a space-time propagator and it's related to the quantum-mechanical wavefunction.
Now, to establish the connection between Hamilton-Jacobi equation and Schrödinger equation, we'll consider the following simplified wavefunction:

$$\psi(x,t) := e^{\frac{iu(x,t)}{\hbar}} \tag{1.35}$$

$\hbar$ is a Planck constant. This function has all the properties of wave function of wave mechanics. Inverting this, we get:

$$u(x,t) = -i\hbar \ln \psi(x,t). \tag{1.36}$$

Now consider the Hamiltonian from the example (1.1) $H(p,x) = \frac{1}{2m}|p|^2 + V(x)$. We want write a Hamilton-Jacobi equation for this H and u in terms of the wave function ψ. Now, let's derive some relations for this:

$$\frac{\partial u}{\partial x_i} = -\frac{i\hbar}{\psi}\frac{\partial \psi}{\partial x_i} \Leftrightarrow \frac{\partial \psi}{\partial x_i} = \frac{i\psi}{\hbar}\frac{\partial u}{\partial x_i}$$

Now, as we assume, that u is a generating function type-2 for the canonical transformation, hence $D_x u = p$, and that p is a momentum, so $p = mq = mx_t$ and thus:

$$\frac{\partial^2 u}{\partial^2 x_i} = \frac{\partial p_i}{\partial x_i} = \frac{d}{dt}\frac{\partial x_i}{\partial x_i} = 0.$$

Now, we can calculate:

$$\frac{\partial^2 \psi}{\partial^2 x_i} = \frac{i}{\hbar}\frac{\partial \psi}{\partial x_i}\frac{\partial u}{\partial x_i} + \frac{i\psi}{\hbar}\frac{\partial^2 u}{\partial^2 x_i} = -\frac{\psi}{\hbar^2}\left(\frac{\partial u}{\partial x_i}\right)^2$$

Which is equivalent to:

$$H(D_x u, t) = \frac{1}{2m}|D_x|^2 + V(x) = -\frac{\hbar^2}{2m\psi}D_x^2\psi + V(x). \tag{1.37}$$

Finally, the time derivative is:

$$\frac{\partial u}{\partial t} = -\frac{i\hbar}{\psi}\frac{\partial \psi}{\partial t}$$

And now we can rewrite the Hamilton-Jacobi equation in terms of ψ:

$$-\frac{\hbar^2}{2m}D_x^2\psi + V(x)\psi - i\hbar\frac{\partial \psi}{\partial t} = 0 \tag{1.38}$$

This is the time-dependent non-relativistic Schrödiner equation, which is a fundamendal mathematical description of behavior of quantum systems. Although we gave only the brief physical motivation of our guess, that u can be views as an amplitude of the wave, the derivation above suggests a deep link between Hamiltonian formulation of classical mechanics and quantum mechanics. More about this connection can be found in [12].

2 Traditional Approach

Now that we have established all the connections between the Hamilton-Jacobi PDE, Hamilton's ODEs and calculus of variation problems, we can move on to finding the solution to the problem (1.1). We'll demostrate how to solve the HJE by separation of variables and then show that in classical sense, the solution may be found only in some restricted cases, and this solution might not even be unique.

2.1 Separation Of Variables

In certain cases we can be lucky and the Hamilton-Jacobi equation can be easily solved using the separation of variables method in an additive way. Note, that in this case, solving this equation is actually easier than using the Hamilton's method directly. Assume, that the Hamiltonian has the following dependence on u and p:

$$H = H(f_1(x_1, p_1), ..., f_d(x_d, p_d))$$

So the HJE has the following form:

$$\partial_t u + H(f_1(x_1, \frac{\partial u}{\partial x_1}), ..., f_d(x_d, \frac{\partial u}{\partial x_d})) = 0 \tag{2.1}$$

Now we separate the first variable and look for the solution u in the following form:

$$u = u_1(x_1) + u'(x_2, ..., x_d, t) \tag{2.2}$$

Passing it back to (2.1), we get the following equation:

$$\partial_t u' + H(f_1(x_1, \frac{\partial u_1}{\partial x_1}), ..., f_d(x_d, \frac{\partial u'}{\partial x_d})) = 0 \tag{2.3}$$

Assume now, that we have already found a solution u in the form (2.2). Then (2.3) becomes an identity for any value of x_1. But when we change x_1, only the function f_1 is affected. Hence it must be constant. Thus, we obtain two equations:

$$f_1(x_1, \frac{\partial u_1}{\partial x_1}) = \alpha_1, \tag{2.4}$$

$$\partial_t u' + H(\alpha_1, f_2(x_2, \frac{\partial u'}{\partial x_2})..., f_d(x_d, \frac{\partial u'}{\partial x_d})) = 0 \tag{2.5}$$

α_1 is an arbitrary constant. Assume also, that f_1 is invertible in the second variable in a sense, that there exists $g(x_1, y_1)$, such that for each (x_1, p_1) and each $y_1 \in \mathbb{R}$, $f(x_1, g(x_1, y_1)) = y_1$. For our case this means, that $\frac{\partial u_1}{\partial x_1} = g(x_1, \alpha_1)$. Then after simple integration we find out u_1:

$$u_1(x_1) = \int g_1(x_1, \alpha_1) dq_1 + C_1, \tag{2.6}$$

where C_1 is the integration constant.

Now we make the similar separation steps consecutively for other variables, hence the solution has the following form:

$$u = u_1(x_1) + u_2(x_2) + ... + u_d(x_d) + u_0(t), \tag{2.7}$$

where each u_k is calculated via the inverse function of f_k, denoted as g_k:

$$u_k(x_k) = \int g_k(x_k, \alpha_1, ..., \alpha_k) dq_k + C_k, \tag{2.8}$$

Finally, the equation (2.1) takes the simple form:

$$\partial_t u_0 + H(\alpha_1, ...\alpha_d) = 0 \tag{2.9}$$

After integrating by t and combining all integration constants into one, we have the complete integral of the equation u:

$$u = -H(\alpha_1, ...\alpha_d)t + \sum_{k=1}^{d} \int g_k(x_k, \alpha_1, ..., \alpha_k) dq_k + C, \tag{2.10}$$

The separability of the equation depends not only on the Hamiltonian itself, but on the generalized coordinates too.

Example 2.1. *We'll demonstrate this method on the very trivial Hamilton-Jacobi PDE for the motion of a particles on a line (one dimension).*

$$u_t + (u_x)^2 = 0. \tag{2.11}$$

Then, following the steps of the method, we have that $H(x,p) = f(x,p) = p^2$*, hence* $g(x,\alpha) = \sqrt{\alpha}$*. Now, we are looking for a solution in a form* $u(x,t) = u_0(t) + u_1(x)$*. We can calculate* $u_1(x)$ *by (2.6):*

$$u_1(x) = \sqrt{\alpha}x + C,$$

and get the final solution by (2.10):

$$u(x,t) = -\alpha t + \sqrt{\alpha}x + C.$$

To specify α *and* C*, we need to supply the initial conditions. For example, if the initial condition is* $u(x,0) = 0$*, then obviously* $\alpha = 0$ *and hence also* $C = 0$*.*

Further, more complicated examples of the usages of this method can be found in §48 in [20].

2.2 Legendre Transform

Before we move on to studying other methods for solving the HJE, we introduce another important notion, connecting Lagrangian and Hamiltonian, namely the Legendre transform. First, we introduce the general notion of what it is, and then we proceed with its relations to the classical mechanics.

Definition 2.1. *Let* $f : \mathbb{R}^d \to \mathbb{R}$*,* $f = f(q)$ *be a convex function. Its* Legendre transform *is the following function:*

$$f^*(p) := \sup_{q \in \mathbb{R}^d} \{p \cdot q - f(q)\}. \tag{2.12}$$

The function f^* *is called* the convex conjugate *of* f*.*

Remark. *Historically, the Legendre transfrom is defined as above, for convex function only, however it can be generalized to non-convex functions, and this generalization is called the* Legendre-Fenchel transformation*. The convexity of the functions, to which the Legendre transform is applied, is important due to the property, called* convex duality*, which simply put, means, that*

$$f^{**} = f,$$

We'll prove this fact later in this chapter for the case of Lagrangian and Hamiltonian. Moreover, this equality above is in fact, a charaterization of a convex lower semicontinuous function. This fact is know as a Fenchel-Moreau theorem*.*
In case of f being non-convex, the biconjugate (the transformation applied twice) will not be the same as the original function, but it will be the largest lower semicontinuous convex function with $f^{**} = f$, which is called the closed convex hull*.*

Remark. *There is a good geometrical interpretation of what a Legendre transform is. Assume that $f = f(q)$ is a strictly convex smooth function. Consider a point q_0 and a supporting plane to the graph of the function f at this point. It will have the form $p \cdot q + b$ and has to satisfy two conditions:*

$$f(q_0) = p \cdot q_0 + b,$$

$$p = D_q f(q_0).$$

For strictly convex function, each component of the gradient is strictly monotone, so the second equation can be uniquely solved for $q_0 = q_0(p)$, which gives us a way to calculate b as a function of p:

$$b = f(q_0(p)) - p \cdot q_0(p) = -f^*(p).$$

Therefore, a family of supporting planes to the graph of f can be given by:

$$y = p \cdot q - f^*(p)$$

So the Legendre transform maps the graph of the function to the family of the tangent planes of this graph.

To simplify the discussion, we assume, that the Lagrangian L and its corresponding Hamiltonian H don't explicitly depend on x, so we write $L = L(q)$ and $H = H(p)$. From now on in this chapter we also assume that Lagrangian L is convex and satisfies the following condition:

$$\lim_{q \to \infty} \frac{L(q)}{|q|} = +\infty \tag{2.13}$$

This condition is called *superlinearity.* First we'll show that the Legendre transform of such a Lagrangian is actually a corresponding Hamiltonian.

Theorem 2.1. *Hamiltonian H, associated with the Lagrangian L is its Legendre transform, or:*

$$H(p) = L^*(p) \tag{2.14}$$

Proof. Remember the definition 1.3 of the Hamiltonian. In the view of condition (2.13), for each p there exists a point $q(p) \in \mathbb{R}^d$, such that the mapping $q \mapsto p{\cdot}q - L(q)$ has a maximum at $q(p)$ and:

$$L^*(p) = p \cdot q(p) - L(q(p)). \tag{2.15}$$

Now it follows that $D_q L(q(p)) = p$, which is a solvable for $q(p)$. It immediately follows, that (2.15) gives us the definition of the Hamiltonian. □

Now we turn to some properties of the Legendre transform, that make it so valuable to the discussion.

Theorem 2.2. *Assume, the convex Lagrangian L satisfies the condition (2.13). Define $H = L^*$. Then the following facts are true:*

1. *H is a convex map,*
2. *$\lim_{p\to\infty} \frac{H(p)}{|p|} = +\infty$,*
3. *$L = H^*$.*

These properties are called the *convex duality* of Hamiltonian and Lagrangian.

Proof. 1. Take $\tau \in [0,1)$ and $p_1, p_2 \in \mathbb{R}^d$. We have:

$$\begin{aligned} H(\tau p_1 + (1-\tau)p_2) = L^*(\tau p_1 + (1-\tau)p_2) &= \sup_q\{(\tau p_1 + (1-\tau)p_2) \cdot q - L(q)\} \\ &\leq \tau \sup_q\{p_1 \cdot q - L(q)\} + (1-\tau)\sup_q\{p_2 \cdot q - L(q)\} \\ &= \tau H(p_1) + (1-\tau)H(p_2) \end{aligned}$$

Hence H is a convex map by definition.

2. Take $p \neq 0$ and $\lambda > 0$. For estimate below we'll just use $q_0 = \lambda\frac{p}{|p|}$. We have:

$$\begin{aligned} H(p) &= \sup_q\{p \cdot q - L(q)\} \\ &\geq \lambda|p| - L(\lambda\frac{p}{|p|}) \\ &\geq \lambda|p| - \max_{q\in B(0,\lambda)} L(q). \end{aligned}$$

Now the latter term is fixed, so it's not affected by the change of p, so we can deduce that:

$$\liminf_{p\to\infty} \frac{H(p)}{|p|} \geq \lambda.$$

Since λ was an arbitrary constant, we proved the second property.

3. Since $H = L^*$, then from the definition of the Legendre transform it follows, that for all $p, q \in \mathbb{R}^d$:
$$H(p) + L(q) \geq p \cdot q.$$
In particular, the following is true:
$$L(q) \geq \sup_p \{p \cdot q - H(p)\} = H^*(q).$$
On the other hand:
$$\begin{aligned} H^*(q) = L^{**}(q) &= \sup_p \{p \cdot q - L^*(p)\} = \sup_p \{p \cdot q - \sup_r \{p \cdot r - L(r)\}\} \\ &= \sup_p \inf_r \{p \cdot (q - r) + L(r)\}\} \end{aligned} \tag{2.16}$$
Since L is a convex function, then by supporting hyperplane theorem (Theorem 1 in Appendix B.1 in [11]), there exists $s \in \mathbb{R}^d$, such that:
$$L(r) \geq L(q) + s \cdot (r - q)$$
So we can choose $p = s$ in (2.16), and then we get:
$$H^*(q) \geq \inf_r \{s \cdot (q - r) + L(r)\} \geq L(q) \tag{2.17}$$
□

2.3 Classical Hopf-Lax Formula

Now we try to solve the simplified initial value problem (1.1), where the Hamiltonian is convex and satisfies the condition $\lim_{p\to\infty} \frac{H(p)}{|p|} = +\infty$. The problem will have the following form:
$$\begin{cases} u_t + H(D_x u) = 0 \\ u(x, 0) = g. \end{cases} \tag{2.18}$$
We suppose that the function g is Lipschitz continuous.
Remember, that Hamilton-Jacobi equation has characteristic lines, which are exactly the Hamilton's ODEs. Furthermore, the Hamilton's ODEs are directly related to the variational problem (1.4). Next we'll try to build a connection between calculus of variation and the Hamilton-Jacobi equation bypassing the characteristic equations. Given $x \in \mathbb{R}^d$ and $t > 0$, we try to minimize the classical action:
$$\int_0^t L(\dot{w}(s))ds$$

over all $w : [0, t] \to \mathbb{R}^d$, such that, $w(t) = x$. We also need to ensure that the initial condition is take into account, so we'll add a term $g(w(0))$. So, we have the following variational problem:

$$u(x,t) := \inf \left\{ \int_0^t L(\dot{w}(s))ds + g(w(0)) | w(t) = x \right\} \tag{2.19}$$

We emphasize again, that this seemingly random guess comes from the relations of the Hamilton's equations to the Lagrange's variational problem and the fact that the Hamilton's equations are in fact the characteristic equations of the HJE. Next we simplify the expression (2.19).
Take $y \in \mathbb{R}^d$ and define $w(s) := y + \frac{s}{t}(x - y)$. Then by definition of u, we have

$$u(x,t) \le \int_0^t L(\dot{w}(s))ds + g(y) = tL\left(\frac{x-y}{t}\right) + g(y),$$

and hence also

$$u(x,t) \le \inf_{y \in \mathbb{R}^d} \{tL\left(\frac{x-y}{t}\right) + g(y)\}.$$

Next, if w is a C^1 function, the Jensen's inequality yields, that:

$$L\left(\frac{1}{t}\int_0^t \dot{w}(s)ds\right) \le \frac{1}{t}\int_0^t L(\dot{w}(s)ds.$$

If we again write $y = w(0)$, we obtain:

$$tL\left(\frac{x-y}{t}\right) + g(y) \le \int_0^t L(\dot{w}(s))ds + g(y),$$

and hence also:

$$\inf_{y \in \mathbb{R}^d} \left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\} \le u(x,t).$$

We have just shown, that:

$$u(x,t) = \inf_{y \in \mathbb{R}^d} \left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\}. \tag{2.20}$$

This looks already better, than the original form of $u(x,t)$. Notice now, that the function $\psi(y) = tL\left(\frac{x-y}{t}\right) + g(y)$ is continuous, due to the continuity of both components by assumption. We want to show, that the infimum in the above formula is in

fact a minimum.
Later we denote by Lip(g) the Lipschitz constant of g, which is

$$\mathrm{Lip}(g) = \sup_{x,y\in\mathbb{R}^d, x\neq y} \left\{ \frac{|g(x)-g(y)|}{|x-y|} \right\}. \tag{2.21}$$

In the case of Lipschitz functions Lip(g) $< \infty$. We make an estimate, using the Lipschitz continuity:

$$|g(y)-g(x)| - \mathrm{Lip}(g)|y-x| \leq 0.$$

Then we add this term to the definition of ψ to estimate it from below:

$$\begin{aligned}\psi(y) &\geq tL\left(\frac{x-y}{t}\right) + g(y) + |g(y)-g(x)| - \mathrm{Lip}(g)|y-x| \\ &\geq tL\left(\frac{x-y}{t}\right) + g(y) + |g(y)| - |g(x)| - \mathrm{Lip}(g)|y-x| \\ &\geq tL\left(\frac{x-y}{t}\right) - |g(x)| - \mathrm{Lip}(g)|y-x| \\ &= |x-y|\left(\frac{L\left(\frac{x-y}{t}\right)}{\frac{|x-y|}{t}} - \frac{|g(x)|}{|x-y|} - \mathrm{Lip}(g)\right)\end{aligned}$$

We claim, that the RHS of this inequality tends to $+\infty$, as $y \to \infty$. Indeed,

$$\frac{L\left(\frac{x-y}{t}\right)}{\frac{|x-y|}{t}} \to_{|y|\to\infty} +\infty,$$

by our assumption (2.13). Obviously, $\frac{|g(x)|}{|x-y|} \to 0$ as $|y| \to \infty$. That proves, that:

$$\lim_{|y|\to\infty} \psi(y) = +\infty, \tag{2.22}$$

which means by definition, that for any x, t there exists $R > 0$, such that if $y \notin \bar{B}(0,R)$ (by $\bar{B}(0,R)$ we denote a closed ball of radius R), then $\psi(y) > u(x,t) + \epsilon$ for some fixed $\epsilon > 0$. But by extreme value theorem inside this closed ball ψ attains its minimum at some point $y_0 \in \bar{B}(0,R)$.
By this construction, $\inf_{|y|\leq R}\psi(y) \geq \psi(y_0)$. Also by definition of the infimum, there exists a point $y' \in \bar{B}(0,R)$, such that $\psi(y') < u(x,t) + \frac{\epsilon}{2}$. From which it follows, that:

$$\inf_{|y|>R} \psi(y) > u(x,t) + \epsilon > u(x,t) + \frac{\epsilon}{2} > \psi(y') \geq \psi(y_0).$$

All of the above means, that $u(x,t) = \inf_{y\in\mathbb{R}^d} \psi(y) \geq \psi(y_0)$. It immediately follows from the definition of the infimum, that:

$$u(x,t) = \min_{y\in\mathbb{R}^d} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\}. \tag{2.23}$$

We have just proven the following theorem, which originally was constructed by Hopf in [17] and Lax in [21]:

Theorem 2.3 (Hopf-Lax)**.** *If $x \in \mathbb{R}^d$ and $t > 0$, then the solution $u = u(x,t)$ of the minimization problem (2.19) has the form (2.23), which is called* the Hopf-Lax formula*.*

Remark. *Actually, the Hopf-Lax formula can be motivated in a different way. Notice, that for any $y, z \in \mathbb{R}^d$ taken as a parameter, the function*

$$F(x,t) = (x-y)\cdot z - tH(z) + g(y)$$

solves the Hamilton-Jacobi equation from the initial value problem (2.18). Indeed, $D_xF(x,t) = z$, and hence $F_t(x,t) = -H(z) = -H(D_xF(x,t))$, which proves this claim.
Then the Hopf-Lax formula can be obtained from this $F(x,t)$ by the two steps envelope solution:

$$\begin{aligned} u(x,t) &= \inf_{y\in\mathbb{R}^d} \sup_{z\in\mathbb{R}^d} \{(x-y)\cdot z - tH(z) + g(y)\} \\ &= \inf_{y\in\mathbb{R}^d} \left\{ tH^*\left(\frac{x-y}{t}\right) + g(y) \right\}. \end{aligned}$$

Which is exactly the Hopf-Lax formula. The fact that inf *in this version is* min *can be proved in the same way as in the reasoning above.*
Moreover, Hopf in [17] constructed a second formula, by switching the inf *and* sup *in the above envelope formula, after which we can write a Legendre transform of g:*

$$\begin{aligned} u(x,t) &= \sup_{z\in\mathbb{R}^d} \inf_{y\in\mathbb{R}^d} \{(x-y)\cdot z - tH(z) + g(y)\} \\ &= \sup_{z\in\mathbb{R}^d} \{x\cdot z - g^*(z) - tH(z)\}. \end{aligned}$$

Before showing, how the Hopf-Lax formula can be used to define some reasonable solution to the Hamilton-Jacobi equation, let us first show some of the useful properties of the function u, defined by the Hopf-Lax formula.

Lemma 2.1. *The function u, defined by Hopf-Lax formula (2.23), where g is Lipschitz continuous, has the following properties:*

1. *u is Lipschitz continuous.*
2. *$u = g$ on $\mathbb{R}^d \times \{t = 0\}$.*
3. *For each $x \in \mathbb{R}^d$ and $0 \le s < t$ we have the following functional identity for $u(x,t)$, defined by (2.23):*

$$u(x,t) = \min_{y\in\mathbb{R}^d} \left\{ (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \right\} \tag{2.24}$$

Proof. 1. Fix $t > 0$, $x, \hat{x} \in \mathbb{R}^d$ and choose $y \in \mathbb{R}^d$, such that

$$u(x,t) = tL\left(\frac{x-y}{t}\right) + g(y).$$

Then

$$\begin{aligned} u(\hat{x},t) - u(x,t) &= \inf_{z\in\mathbb{R}^d} \left\{ tL\left(\frac{\hat{x}-z}{t}\right) + g(z) \right\} - tL\left(\frac{x-y}{t}\right) - g(y) \\ &= [\text{let } z = \hat{x} - x + y \text{ for the expression inside the inf}] \\ &\le tL\left(\frac{x-y}{t}\right) + g(\hat{x} - x + y) - tL\left(\frac{x-y}{t}\right) - g(y) \\ &\le g(\hat{x} - x + y) - g(y) \le \mathrm{Lip}(g)|\hat{x} - x|. \end{aligned}$$

The similar result can be archieved by interchanging roles of x and $\hat{x}$ in the above argument, from where it follows that:

$$|u(\hat{x},t) - u(x,t)| \le \mathrm{Lip}(g)|\hat{x} - x|. \tag{2.25}$$

This shows the Lipschitz continuity of $x \mapsto u(x,\cdot)$.

2. Let $0 < s < t$. The formula (2.24), that we need to prove, means is that once we know the value of $u(x,s)$, we can us it as a initial condition for calculating $u(x,t)$. Fix $y \in \mathbb{R}^d$ and $0 < s < t$ and choose $z \in \mathbb{R}^d$, such that

$$u(y,s) = sL\left(\frac{y-z}{s}\right) + g(z).$$

Notice, that:

$$\frac{x-z}{t} = \left(1 - \frac{s}{t}\right)\frac{x-y}{t-s} + \frac{s}{t}\frac{y-z}{s}.$$

It follows by convexity of L, that:

$$L\left(\frac{x-z}{t}\right) \leq \left(1 - \frac{s}{t}\right) L\left(\frac{x-y}{t-s}\right) + \frac{s}{t} L\left(\frac{y-z}{s}\right).$$

Therefore

$$\begin{aligned} u(x,t) \leq tL\left(\frac{x-z}{t}\right) + g(z) &\leq (t-s)L\left(\frac{x-y}{t-s}\right) + sL\left(\frac{y-z}{s}\right) + g(z) \\ &= (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s). \end{aligned}$$

This is true for all $y \in \mathbb{R}^d$. We have already shown the Lipschitz continuity, and hence the continuity of a map $y \mapsto u(y, \cdot)$. So it follows, that:

$$u(x,t) \leq \min_{y \in \mathbb{R}^d} \left\{ (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \right\}$$

Now we can choose w, such, that

$$u(x,t) = tL\left(\frac{x-w}{t}\right) + g(w) \tag{2.26}$$

and set $y = \frac{s}{t}x + (1 - \frac{s}{t})w$. Then

$$\frac{x-y}{t-s} = \frac{x-w}{t} = \frac{y-w}{s}.$$

And consequently, using the convexity of L again:

$$\begin{aligned} (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) &\leq (t-s)L\left(\frac{x-w}{t}\right) + sL\left(\frac{y-w}{s}\right) + g(w) \\ &= tL\left(\frac{x-w}{t}\right) + g(w) = u(x,t). \end{aligned}$$

This ends the proof of the identity (2.24).

3. We now need to show that there is also Lipschitz continuity with respect to the t variable. Let us again fix $t > 0$ and $x \in \mathbb{R}^d$ and choose $y = x$ in the Hopf-Lax formula (2.23). Then

$$u(x,t) \leq tL(0) + g(x). \tag{2.27}$$

Moreover, using the Lipschitz condition on g,

$$|g(x) - g(y)| \leq \mathrm{Lip}(g)|x - y| \Rightarrow g(y) \geq g(x) - \mathrm{Lip}(g)|x - y|.$$

we can estimate,

$$\begin{aligned}
u(x,t) &= \min_{y\in\mathbb{R}^d} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\} \\
&\geq g(x) + \min_{y\in\mathbb{R}^d} \left\{ tL\left(\frac{x-y}{t}\right) - \mathrm{Lip}(g)|x-y| \right\} \\
&[\text{now set } z = \frac{x-y}{t}] \\
&= g(x) + \min_{z\in\mathbb{R}^d} \{tL(z) - t\mathrm{Lip}(g)|z|\} \\
&[\text{next we use } min\{f(x)\} = -max\{-f(x)\} \\
&= g(x) - t \max_{z\in\mathbb{R}^d} \{\mathrm{Lip}(g)|z| - L(z)\} \\
&= g(x) - t \max_{w\in B(0,Lip(g))} \max_{z\in\mathbb{R}^d} \{w \cdot z - L(z)\} \\
&= g(x) - t \max_{w\in B(0,Lip(g))} L^*(w) = g(x) - t \max_{w\in B(0,Lip(g))} H(w)
\end{aligned}$$

From this result and (2.27) it follows, that

$$|u(x,t) - g(x)| \leq Ct, \tag{2.28}$$

where

$$C := \max\{|L(0)|, \max_{w\in B(0,Lip(g))} H(w)\}.$$

This inequality immediate yields the initial condition for u. If $t = 0$, then obviously $u(x, 0) = g(x)$.
Finally, by using the (2.24) and the fact, that $\mathrm{Lip}(u(\cdot, t) \leq Lip(g)$, proven above, we can repeat the above calculations.
Choose $0 < s < t$ and $x = y$ in (2.24). Then:

$$u(x,t) \leq (t - s)L(0) + u(x, s).$$

And, similarly to when s=0:

$$\begin{aligned} u(x,t) &\geq u(x,s) + \min_{y \in \mathbb{R}^d} \left\{ (t-s) L\left(\frac{x-y}{t-s} \right) - \mathrm{Lip}(u(\cdot,t))|x-y| \right\} \\ &= u(x,s) - C(t-s) \end{aligned}$$

This proves the Lipschitz continuity of $t \mapsto u(\cdot,t)$.

□

Theorem 2.4. *Let u be defined by Hopf-Lax formula (2.23). Then u is differentiable a.e. in $\mathbb{R}^d \times (0,\infty)$ and solves the initial value problem (2.18) in the points of differentiability.*

Proof. The differentiability a.e. follows immediately from the Lipschitz continuity by the Radamacher theorem.
Now let (x,t), where $x \in \mathbb{R}^d$, $t > 0$, be a point, where u is differentiable. Fix $q \in \mathbb{R}^d$, $h > 0$. By (2.24):

$$\begin{aligned} u(x+hq, t+h) &= \min_{y \in \mathbb{R}^d} \left\{ hL\left(\frac{x+hq-y}{h} \right) + u(y,t) \right\} \text{ [inside min let } y = x] \\ &\leq hL(q) + u(x,t). \end{aligned}$$

hence

$$\frac{u(x+hq,t+h) - u(x,t)}{h} \leq L(q). \tag{2.29}$$

Notice, that

$$\begin{aligned} \frac{u(x+hq,t+h) - u(x,t)}{h} &= \frac{u(x+hq,t+h) - u(x+hq,t)}{h} + \frac{u(x+hq,t) - u(x,t)}{h} \\ &\to q \cdot D_x u(x,t) + u_t(x,t) \text{ as } h \to 0^+ \end{aligned}$$

Combined with (2.29) we obtain:

$$q \cdot D_x u(x,t) + u_t(x,t) \leq L(q). \tag{2.30}$$

Since q is arbitrary, then, in particular, we have:

$$0 \geq u_t(x,t) + \max_{q \in \mathbb{R}^d} \{ q \cdot D_x u(x,t) - L(q) \} = u_t(x,t) + L^*(D_x u(x,t)). \tag{2.31}$$

To get the reversed inequality, we start by choosing z, such that:

$$u(x,t) = tL\left(\frac{x-z}{t} \right) + g(z)$$

and fix $h > 0$. We set $s = t - h$ and $y = \frac{s}{t}x + (1 - \frac{s}{t})z$. Then $\frac{x-z}{t} = \frac{y-z}{s}$. Therefore

$$\begin{aligned} u(x,t) - u(y,s) &\geq tL\left(\frac{x-z}{t}\right) + g(z) - \left(tL\left(\frac{y-z}{s}\right) + g(z)\right) \\ &= (t-s)L\left(\frac{x-z}{t}\right). \end{aligned}$$

From where it follows, that:

$$\frac{u(x,t) - u(x - h\frac{x-z}{t}, t-h)}{h} \geq L\left(\frac{x-z}{t}\right).$$

By letting $h \to 0^+$, we obtain:

$$\frac{x-z}{t} \cdot D_x u(x,t) + u_t(x,t) \geq L\left(\frac{x-z}{t}\right)$$

Finally

$$\begin{aligned} 0 &\leq u_t(x,t) + \frac{x-z}{t} \cdot D_x u(x,t) - L\left(\frac{x-z}{t}\right) \\ &\leq u_t(x,t) + \max_{q \in \mathbb{R}^d}\{q \cdot D_x u(x,t) - L(q)\} \\ &= u_t(x,t) + L^*(D_x u(x,t)). \end{aligned}$$

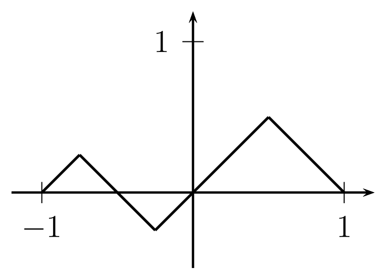
Taking into account the fact that $H = L^*$, this ends the proof. $\square$

One may conclude from this, that the Lipschitz continuous function, that solves the HJE a.e. and agrees with g, when $t = 0$, might be a good definition for the weak solution of (2.18), but unfortunately, such a solution is not unique, that can be illustrated by the following example:

Example 2.2. *Let's look at the eikonal equation in one dimension:*

$$\begin{cases} |u'(x)| = 1 \text{ on } [-1,1] \\ u(-1) = u(1) = 0. \end{cases} \tag{2.32}$$

We can notice, that the functions $u_1(x) := 1 - |x|$ and $u_2(x) := |x| - 1$ solve the equation almost everywhere (except for $x = 0$, as the derivative doesn't exist there). Moreover, since we only want a.e. solution, we can notice, that any function, that has the piecewise property $|u'(x)| = 1$ and agrees with the boundary conditions, will *solve the equation. For example, the function, that has the following graph, also solves the equation a.e., along with u_1 and u_2:*



also solves the initial value problem almost everywhere and it's Lipschitz continuous. Moreover, there're infinitely many function, that satisfy this problem almost everywhere. So as a conclusion, we need to put more restrictions on the functions g and H, and rethink our definition of the weak solution.

2.4 Weak Solutions

We'll start with looking into some more properties of the Hopf-Lax formula, establishing more connections between function u, defined by the formula and the given initial value problem. First, we give a definition of semiconcavity.

Definition 2.2. *A real-valued function g is called* semiconcave*, if there exists a constant C, such that for all $x, z \in \mathbb{R}^d$, the following inequality holds*

$$g(x+z) - 2g(x) + g(x-z) \leq C|z|^2 \tag{2.33}$$

Remark. *An extensive overview of the semiconcave functions and its applications is given in [5]. The term "semi"-concavity comes from some of the properties of the functions. Corollary 2.1.3 in [5] proves, that any semiconcave function can be represented as a sum of a smooth function and a concave function. Hence, to the surprise, a strictly convex function, as x^2 can be seen as an example of a semiconcave function, which is not concave.*

Lemma 2.2. *Let g be semiconcave. Then u, defined by (2.23), is also semiconcave w.r.t. x, meaning that for all $x, z \in \mathbb{R}^d$, and $t > 0$:*

$$u(x+z,t) - 2u(x,t) + u(x-z,t) \leq C|z|^2. \tag{2.34}$$

Proof. Take $y \in \mathbb{R}^d$ a minimizer in Hopf-Lax formula, i.e.:

$$u(x,t) = tL\left(\frac{x-y}{t}\right) + g(y).$$

Now, we can estimate $u(x+z,t)$ and $u(x-z,t)$ by the Hopf-Lax formula from above, by choosing respectively $y+z$ and $y-z$ as corresponding y-variable from the formula. It follows that:

$$\begin{aligned} u(x+z,t) - 2u(x,t) + u(x-z,t) &\leq \left(tL\left(\frac{x+z-y-z}{t}\right) + g(y+z) \right) \\ &- 2\left(tL\left(\frac{x-y}{t}\right) + g(y) \right) + \left(tL\left(\frac{x-z-y+z}{t}\right) + g(y-z) \right) \\ &= g(y+z) - 2g(y) + g(y-z) \leq C|z|^2. \end{aligned}$$

□

The second step will be to establish the connection between H and u. We drop the semiconcavity assumption on g and consider now the uniformly convex H. It happens, that this property is enough to ensure the semiconcavity of the solution u for any fixed $t > 0$. Again, we first give a definition.

Definition 2.3. *A C^2 function $H : \mathbb{R}^d \to \mathbb{R}$ is called* strongly convex *with constant $\theta > 0$, if the following inequality is true for all $p, \xi \in \mathbb{R}^d$:*

$$\xi^T \nabla^2 H(p) \xi \geq \theta |\xi|^2. \tag{2.35}$$

Remark. *This definition is actually a special case of a* uniformly convex *function, which means that there exists a function ϕ, which is increasing and vanishes only at zero, such that:*

$$\xi^T \cdot D_x^2 H(p) \cdot \xi \geq \phi(|\xi|) \tag{2.36}$$

Obviously, the case of $\phi(\alpha) = \theta\alpha^2$ gives the above definition of a strongly convex function.

Lemma 2.3. *Let H be a strongly convex Hamiltonian with constant θ, $L = H^*$ be the corresponding Lagrangian, and u is defined by the Hopf-Lax formula (2.23). Then u is semiconcave for any $t > 0$, in particular, for all $x, z \in \mathbb{R}^d$:*

$$u(x+z,t) - 2u(x,t) + u(x-z,t) \leq \frac{1}{\theta t}|z|^2.$$

Proof. Let $p_1, p_2 \in \mathbb{R}^d$. Then, by Taylor's formula, we have the following identities:

$$\begin{aligned} H(p_1) = H\left(\frac{p_1+p_2}{2}\right) &+ D_x H\left(\frac{p_1+p_2}{2}\right) \cdot \left(\frac{p_1-p_2}{2}\right) \\ &+ \frac{1}{2}\left(\frac{p_1-p_2}{2}\right)^T \cdot D_x^2 H(\hat{p}_1) \cdot \left(\frac{p_1-p_2}{2}\right)^T \end{aligned}$$

$$H(p_2) = H\left(\frac{p_1+p_2}{2}\right) + D_xH\left(\frac{p_1+p_2}{2}\right)\cdot\left(\frac{p_2-p_1}{2}\right)$$
$$+\frac{1}{2}\left(\frac{p_2-p_1}{2}\right)^T \cdot D_x^2H(\hat{p}_2)\cdot\left(\frac{p_2-p_1}{2}\right)^T$$

$\hat{p}_1$ and $\hat{p}_2$ are the mid-points, corresponding to the remainder term in Lagrange form. Now, summing it up and estimating each of the remainder terms by (2.35), we obtain:

$$H(p_1) + H(p_2) \geq 2H\left(\frac{p_1+p_2}{2}\right) + \frac{\theta}{4}|p_1-p_2|^2 \tag{2.37}$$

Next, let $q_1, q_2 \in \mathbb{R}^d$ be arbitrary. We choose p_1, p_2 be such, that $H(p_1) = p_1\cdot q_1 - L(q_1)$ and $H(p_2) = p_2\cdot q_2 - L(q_2)$. The existence and uniqueness is ensured by the relations between Hamiltonian and Lagrangian. Then we can estimate:

$$L(q_1) + L(q_2) = p_1\cdot q_1 + p_2\cdot q_2 - H(p_1) - H(p_2)$$
$$\leq p_1\cdot q_1 + p_2\cdot q_2 - 2H\left(\frac{p_1+p_2}{2}\right) - \frac{\theta}{4}|p_1-p_2|^2$$

Now, by definition of the Legendre transform, we have

$$H\left(\frac{p_1+p_2}{2}\right) \geq \frac{1}{4}(p_1+p_2)\cdot(q_1+q_2) - L\left(\frac{q_1+q_2}{2}\right)$$

Combining it with the previous inequality, we obtain:

$$L(q_1) + L(q_2) \leq p_1\cdot q_1 + p_2\cdot q_2 - \frac{1}{2}(p_1+p_2)\cdot(q_1+q_2) + 2L\left(\frac{q_1+q_2}{2}\right) - \frac{\theta}{4}|p_1-p_2|^2$$
$$\leq \frac{1}{2}(p_1-p_2)\cdot(q_1-q_2) + 2L\left(\frac{q_1+q_2}{2}\right) - \frac{\theta}{4}|p_1-p_2|^2 \tag{2.38}$$

Finally, we notice, that:

$$\frac{\theta}{4}|p_1-p_2|^2 - \frac{1}{2}(p_1-p_2)\cdot(q_1-q_2) + \frac{1}{4\theta}|q_1-q_2|^2$$
$$\geq \frac{\theta}{4}|p_1-p_2|^2 - 2\left(\frac{\sqrt{\theta}}{2}|p_1-p_2|\right)\cdot\left(\frac{1}{2\sqrt{\theta}}|q_1-q_2|\right) + \frac{1}{4\theta}|q_1-q_2|^2$$
$$= \left(\frac{\sqrt{\theta}}{2}|p_1-p_2| - \frac{1}{2\sqrt{\theta}}|q_1-q_2|\right)^2 \geq 0$$

This will help to get rid of the p_i from (2.38). We obtain:

$$L(q_1) + L(q_2) \leq 2L\left(\frac{q_1 + q_2}{2}\right) + \frac{1}{4\theta}|q_1 - q_2|^2 \tag{2.39}$$

Finally, we choose y to be a minimizer in the Hopf-Lax formula for $u(x,t)$, and then using the same value for estimating $u(x+z,t)$ and $u(x-z,t)$, we obtain:

$$\begin{aligned} u(x+z,t) - 2u(x,t) + u(x-z,t) \leq & \left(tL\left(\frac{x+z-y}{t}\right) + g(y)\right) \\ & - 2\left(tL\left(\frac{x-y}{t}\right) + g(y)\right) + \left(tL\left(\frac{x-z-y}{t}\right) + g(y)\right) \\ & \leq t\left(L\left(\frac{x+z-y}{t}\right) - 2L\left(\frac{x-y}{t}\right) + L\left(\frac{x-z-y}{t}\right)\right) \\ & \leq t\frac{1}{4\theta}\left|\frac{2z}{t}\right|^2 = \frac{1}{\theta t}|z|^2. \end{aligned}$$

Here we applied the inequality (2.39) here with $q_1 = \frac{x+z-y}{t}$ and $q_2 = \frac{x-z-y}{t}$. □

Now we can combine the results of both the lemmas to create a suitable definition of the weak solution of the HJE initial value problem (2.18).

Definition 2.4. *A Lipschitz continuous function $u : \mathbb{R}^d \times [0,\infty) \to \mathbb{R}$ is called the weak solution of the initial value problem (2.18), provided:*

1. *$u(x,0) = g(x)$, when $x \in \mathbb{R}^d$,*
2. *$u_t(x,t) + H(D_x u(x,t)) = 0$ for a.e. $(x,t) \in \mathbb{R}^d \times (0,\infty)$.*
3. *$u(x+z,t) - 2u(x,t) + u(x-z,t) \leq C(1 + \frac{1}{t})|z|^2$ for some constant $C \geq 0$ and all $x, z \in \mathbb{R}^d, t > 0$.*

The previous lemmas (2.2) and (2.3) directly lead us to the existence theorem of the weak solution, under some restrictions on the function H or g.

Theorem 2.5 (Hopf-Lax formula as a weak solution)**.** *If H is a C^2 superlinear convex function, $L = H*$ be a corresponding Lagrangian and g is Lipschitz continuous, then u, defined by the Hopf-Lax formula (2.23) is a weak solution of the initial value problem for the HJE (2.18), if either g is semiconcave or H is strongly convex.*

The uniqueness, however, can be ensured without additional semiconcavity and strong convexity assumptions on g and H, however the semiconcavity is used here, as the key point to the proof is the condition 3. from the definition 2.4.

Theorem 2.6 (Uniqueness of the weak solution)**.** *If H is a C^2 superlinear convex function and g is a Lipschitz continuous function, then there exists at most one solution to the initial value problem (2.18).*

Proof. Suppose, that u_1 and u_2 are two weak solutions of the (2.18). Let $w = u_1 - u_2$ and let (y, s) be an arbitrary point, where u_1 and u_2 are differentiable, and hence solve the HJE. Then

$$\begin{aligned} w_t(y,s) = u_{1t}(y,s) - u_{2t}(y,s) &= -H(D_x u_1(y,s) + H(D_x u_2(y,s)) \\ &= -\int_0^1 \frac{d}{d\tau} H(\tau D_x u_1(y,s) + (1-\tau) D_x u_2(y,s)) d\tau \\ &= -\int_0^1 D_p H(\tau D_x u_1(y,s) + (1-\tau) D_x u_2(y,s)) d\tau \cdot (D_x u_1(y,s) - D_x u_2(y,s)) \end{aligned}$$

We later denote the first multiplier as

$$b(y,s) := \int_0^1 D_p H(\tau D_x u_1(y,s) + (1-\tau) D_x u_2(y,s)) d\tau. \tag{2.40}$$

Consequently,

$$w_t + b \cdot D_x w = 0 \text{ a.e.} \tag{2.41}$$

Let ϕ be some smooth function $\phi : \mathbb{R} \to [0, \infty)$, whose further properties we'll establish later in the proof. Denote $v(t) := \phi(w(t)) \geq 0$. Then from (2.41), multiplied by $\phi'(w)$, it follows

$$\phi'(w) w_t + b \cdot (\phi'(w) D_x w) = v_t + b \cdot D_x v = 0 \text{ a.e.} \tag{2.42}$$

Let $\epsilon > 0$ be arbitrary and $\eta_\epsilon : \mathbb{R}^d \times (0, \infty) \to \mathbb{R}$ be the standard mollifier. We denote the mollifications of u_i by $u_i^\epsilon := \eta_\epsilon * u_i$. Now we can establish some facts, using the properties of convolution (they can be found in §C.4 in [11]).
First, since $u_i^\epsilon \in C^\infty$, then $D_x u_i^\epsilon$ exists, and moreover, using Lipschitz continuity of

u_i:

$$\left|\frac{u_i^\epsilon(x+he,t)-u_i^\epsilon(x,t)}{h}\right|$$
$$=\left|\int_{\mathbb{R}^d\times(0,\infty)}\eta_\epsilon(y,r)\frac{(u_i(x+he-y,t-r)-u_i(x-y,t-r))}{h}dydr\right|$$
$$\leq\int_{\mathbb{R}^d\times(0,\infty)}\eta_\epsilon(y,r)\left|\frac{(u_i(x+he-y,t-r)-u_i(x-y,t-r))}{h}\right|dydr$$
$$\leq \mathrm{Lip}(u_i)|e|\int_{\mathbb{R}^d}\eta_\epsilon(y,r)dydr=\mathrm{Lip}(u_i).$$

e here is a unit vector. By taking a limit $h\to 0$ and an absolute value, we get

$$|D_xu_i^\epsilon|\leq \mathrm{Lip}(u_i). \tag{2.43}$$

The implicit formula

$$D_xu_i^\epsilon(x,t)=\int_{\mathbb{R}^d\times(0,\infty)}\eta_\epsilon(x,t)D_xu(x-y,t-r)dydr$$

immediately yields, that $D_xu_i^\epsilon$ is a mollification of D_xu_i and hence

$$D_xu_i^\epsilon\to D_xu_i,\ \text{a.e. as } \epsilon\to 0 \tag{2.44}$$

Finally, by using the property 3 of a definition of a weak solution (2.4), we have one more inequality for all $\epsilon>0$, $y\in\mathbb{R}^d$ and $s>2\epsilon$:

$$D_x^2u_i^\epsilon(y,s)\leq C\left(1+\frac{1}{s}\right)I \tag{2.45}$$

By I we denote an identity matrix. This inequality can be obtained by the limit argument for the second derivative, like it was done above with the Lipschitz continuity. Simply notice, that for $z>0$, $|z|\leq 1$

$$\frac{u_i(y+ze_j,s)-2u_i(y,s)+u_i(y-ze_j,s)}{h^2}\leq C\left(1+\frac{1}{s}\right)|z|^2\leq C\left(1+\frac{1}{s}\right).$$

By letting $h\to 0$, we obtain the claim.
We denote

$$b_\epsilon(y,s):=\int_0^1 D_pH(rD_xu_1^\epsilon(y,s)+(1-r)D_xu_2^\epsilon(y,s))dr. \tag{2.46}$$

Then:

$$v_t + b_\epsilon \cdot D_x v = (b_\epsilon - b) \cdot D_x v \text{ a.e}$$

and therefore

$$v_t + \operatorname{div}(v b_\epsilon) = \operatorname{div}(b_\epsilon) v + (b_\epsilon - b) \cdot D_x v \text{ a.e.}$$

Clearly,

$$\begin{aligned} \operatorname{div}(b_\epsilon) &= \int_0^1 \sum_{k,l=1}^{d} H_{p_k p_l}(r D_x u_1^\epsilon + (1-r) D_x u_2^\epsilon)(r u^\epsilon_{1 x_l x_k} + (1-r) u^\epsilon_{1 x_l x_k}) dr \\ &\leq C\left(1 + \frac{1}{s}\right) \end{aligned} \tag{2.47}$$

Fix $x_0 \in \mathbb{R}^d$, $t_0 > 0$ and let

$$R := \max\{|D_p H(p)|, \text{ s.t. } |p| \leq \max\{\operatorname{Lip}(u_1), \operatorname{Lip}(u_2)\}\}$$

and

$$C := \{(x,t), \text{ s.t. } 0 \leq t \leq t_0, |x - x_0| \leq R(t_0 - t)\}.$$

Denote

$$e(t) = \int_{B(x_0, R(t_0 - t))} v(x,t) dx.$$

And now we compute for $t > 0$:

$$\begin{aligned} \dot{e}(t) &= \int_{B(x_0,R(t_0-t))} v_t dx - R \int_{\partial B(x_0,R(t_0-t))} v(x,t) dS \\ &= \int_{B(x_0,R(t_0-t))} -\operatorname{div}(v b_\epsilon) + (\operatorname{div}(b_\epsilon)) + (b_\epsilon - b) \cdot D_x v dx \\ &\quad - R \int_{\partial B(x_0,R(t_0-t))} v(x,t) dS \\ &= -\int_{\partial B(x_0,R(t_0-t))} v(b_\epsilon \cdot \nu + R) dS \\ &\quad + \int_{B(x_0,R(t_0-t))} \operatorname{div}(b_\epsilon) v + (b_\epsilon - b) \cdot D_x v dx \\ &\leq \int_{B(x_0,R(t_0-t))} \operatorname{div}(b_\epsilon) v + (b_\epsilon - b) \cdot D_x v dx \\ &\leq C\left(1 + \frac{1}{t}\right) e(t) + \int_{B(x_0,R(t_0-t))} (b_\epsilon - b) \cdot D_x v dx \end{aligned}$$

Letting $\epsilon \to 0$, we have by the Dominated Convergence Theorem:

$$\dot{e}(t) \leq C\left(1 + \frac{1}{t}\right)e(t), \text{ a.e. } 0 < t < t_0. \quad (2.48)$$

Now fix $0 < \epsilon < r < t$ and let the function $\phi(z) \geq 0$ be zero, if

$$|z| \leq \epsilon(\mathrm{Lip}(u_1) + \mathrm{Lip}(u_2)).$$

Now, by the initial condition $v = \phi(w) = \phi(u_1 - u_2) = 0$ at $\{t = \epsilon\}$. Hence $e(\epsilon) = 0$. So by Grönwall's inequality we have:

$$e(r) \leq e(\epsilon)e^{\int_\epsilon^r C\left(1+\frac{1}{s}\right)ds} = 0 \quad (2.49)$$

Therefore,

$$|u_1 - u_2| \leq \epsilon(\mathrm{Lip}(u_1) + \mathrm{Lip}(u_2)) \text{ on } B(x_0, R(t_0 - r)). \quad (2.50)$$

Since it's true for all $\epsilon > 0$, then $u_1(x_0, t_0) = u_2(x_0, t_0)$. □

As a conclusion, we can say, that the existence and the uniqueness of the solution of the HJE are being well established only in a restricted case. There're however a lot of limitations on the input data, that cannot be omitted. That's the reason why recently there was developed another notion of the weak solution, called viscosity solution, which helps to build an existence and uniqueness theory further without harsh restrictions on the initial data and the Hamiltonian. The next chapter is dedicated to this approach.

3 Viscosity Solutions

In this chapter we'll introduce the notion of viscosity solution, along with its original motivation, originating in the vanishing viscosity method. Nowadays this concept is widely used in the theory of degenerate elliptic PDEs (we'll give the definition for that later), but originally it was introduced as a generalized solution particularly to the Hamilton-Jacobi equation.

3.1 Method Of Vanishing Viscosity

The method of vanishing viscosity originated in the theory of conservations laws. If we have a conservation law (the initial data is omitted, as it is not relevant at this part of the discussion):

$$u_t + (F(u))_x = 0, \quad (3.1)$$

then the method is to have this equation as a limit ($\varepsilon \to 0$) of the following equation:

$$u_t^\varepsilon + (F(u^\varepsilon))_x = \varepsilon \Delta u^\varepsilon, \tag{3.2}$$

where $\Delta u = \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2}$.
Such a model originally appeared in fluid dynamics, namely in different specific cases of Navier-Stokes equations, where the term $\mu \Delta u$ represents the viscosity of the fluid, that's why the term $\varepsilon \Delta u$ is called *the vanishing viscosity*. This method is motivated, first of all, by the fact, that this additional small viscosity term makes the model more physically real. And more importantly for studying the solvability of (3.1), the equation (3.2) may have a smooth solution, even in case of discontinuous initial data. More generally, the method can be seen as:

$$A(u) = f \ \to \ A_\varepsilon(u_\varepsilon) = f,$$

where A is a nonlinear differential operator, f is given, and A_ε are operators that are expected to converge to A in some sense.
Now we turn back to the Hamilton-Jacobi equation (1.1). We forget about the restrictions on the Hamiltonian from the previous chapter and consider the general case, where H depends on x and may not be convex in p. This application of the vanishing viscosity method is originally due to Evans in [10].
We start by adding a viscosity term:

$$\begin{cases} u_t^\varepsilon + H(D_x u^\varepsilon, x) = \varepsilon \Delta u^\varepsilon \\ u^\varepsilon(x, 0) = g. \end{cases} \tag{3.3}$$

This is a quasilinear parabolic PDE. In our case of smooth Hamiltonian, it turn out to have smooth solutions (the proof of this fact can found in Friedman [13]), so we can call (3.3) a regularization of (1.1). Our hope is that u^ε converges to some solution u of (1.1), as $\varepsilon \to 0$. Evans in [10] proved the following theorem, using the Friedman's result:

Theorem 3.1. *Let $H = H(p, x)$ be a Hamiltonian and there exists a constant M, s.t.*

$$|H| + |H_x| + |H_p| \leq M,$$

let g be a bounded Lipschitz function. Then there exists a sequence $\varepsilon_j \searrow 0$, such that there exists a uniform limit on compact sets of $\mathbb{R}^d \times [0, T]$:

$$\lim_{\varepsilon_j \searrow 0} u^{\varepsilon_j}(x, t) = u(x, t)$$

and this limit is Lipschitz and is a solution of (1.1) in a weak sense (it solves the equation a.e. and satisfies the initial condition).

This is a good way to ensure the existence of such a weak solution to the original problem, although, as we can see, there are still some assumptions put on the Hamiltonian. Later we are going to build the existence theory, that follows from the control theory.
Now we turn to some observations regarding the limit u^ε, as $\varepsilon \to 0$. First, let's notice, that the estimates u^ε strongly depend on the regularizing term $\varepsilon \Delta u$, hence we can lose control over these estimates and their derivatives, as $\varepsilon \to 0$. However, in practice, if we assume, that u^ε converge uniformly on some compact subsets of $\mathbb{R}^d \times [0, \infty)$, then we can work with a family u^ε, that is bounded and equicontinuous on these compact subsets. Then, by the Arzela-Ascoli theorem, there exists a subsequence $\{u^{\varepsilon_j}\}_{j=1}^{\infty}$ and some function $u \in C(\mathbb{R}^d \times [0, \infty)$, such that

$$u^{\varepsilon_j} \to u \text{ locally uniformly in } \mathbb{R}^d \times [0, \infty) \tag{3.4}$$

This at least gives us some information about the solution u, however, as it's only continuous, we still have no idea about whether the derivatives exist and what they might look like. So what we are going to do is to exploit the maximum principle, which will allow us to move to differentiating smooth test functions.
Let u be defined as a limit above. Fix $v \in C^\infty(\mathbb{R}^d \times (0, \infty))$ and assume, that

$$u - v \text{ has a } \textit{strict} \text{ local maximum as some point} (x_0, t_0). \tag{3.5}$$

This means, that

$$(u - v)(x_0, t_0) > (u - v)(x, t),$$

for all (x, t) sufficiently close to (x_0, t_0). In particular, for any $r > 0$ and B_r, a closed ball in $\mathbb{R}^{d+1}$, centered at (x_0, t_0) with a radius r, we have that $\max_{\partial B_r}(u - v) < (u - v)(x_0, t_0)$. Now, by definition of u via the limit (3.4), we have that $u^{\varepsilon_j} \to u$ uniformly on B_r and hence for small ε_j we have that :

$$\max_{\partial B_r}(u^{\varepsilon_j} - v) < (u^{\varepsilon_j} - v)(x_0, t_0).$$

This means that $u^{\varepsilon_j} - v$ attains a local maximum somewhere inside B_r. If we repeat these steps for a sequence $r_j \to 0$, we get that there exists $(x_{\varepsilon_j}, t_{\varepsilon_j}) \to (x_0, t_0)$, such that

$$u^{\varepsilon_j} - v \text{ has a local maximum at } (x_{\varepsilon_j}, t_{\varepsilon_j}). \tag{3.6}$$

And this, consequently, means, that

$$\begin{cases} D_x u^{\varepsilon_j}(x_{\varepsilon_j}, t_{\varepsilon_j}) = D_x v(x_{\varepsilon_j}, t_{\varepsilon_j}) \\ u_t^{\varepsilon_j}(x_{\varepsilon_j}, t_{\varepsilon_j}) = v_t(x_{\varepsilon_j}, t_{\varepsilon_j}) \end{cases} \tag{3.7}$$

and

$$-\Delta u^{\varepsilon_j}(x_{\varepsilon_j}, t_{\varepsilon_j}) \geq -\Delta v(x_{\varepsilon_j}, t_{\varepsilon_j}) \tag{3.8}$$

We can now calculate, using the results above and (3.3):

$$\begin{aligned} &v_t(x_{\varepsilon_j}, t_{\varepsilon_j}) + H(D_x v(x_{\varepsilon_j}, t_{\varepsilon_j}), x_{\varepsilon_j}) \\ &= u_t^{\varepsilon_j}(x_{\varepsilon_j}, t_{\varepsilon_j}) + H(D_x u^{\varepsilon_j}(x_{\varepsilon_j}, t_{\varepsilon_j}), x_{\varepsilon_j}) \\ &= \varepsilon_j \Delta u^{\varepsilon_j}(x_{\varepsilon_j}, t_{\varepsilon_j}) \\ &\leq \varepsilon_j \Delta v(x_{\varepsilon_j}, t_{\varepsilon_j}) \end{aligned} \tag{3.9}$$

Finally, we let $\varepsilon_j \to 0$ ad using the fact that v is a smooth test function and H is a smooth Hamiltonian (in fact we can require it to be only continuous), and we obtain

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \leq 0. \tag{3.10}$$

If we soften the condition (3.5) by assuming that the maximum is not necessarily strict, we can still deduce the same inequality with some additional steps.
Let's denote for $\delta > 0$, $v^\delta(x, t) := v(x, t) + \delta(|x - x_0|^2 + |t - t_0|^2)$. If $u - v$ has a local maximum at (x_0, t_0), then $u - v^\delta$ has a strict local maximum at this point, so by the above reasoning $v_t^\delta(x_0, t_0) + H(D_x v^\delta(x_0, t_0), x_0) \leq 0$ and (3.10) follows for v as well.
In a similar manner, by reversing some inequalities, we can prove, that if

$$u - v \text{ has a local minimum at } (x_0, t_0), \tag{3.11}$$

then

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \geq 0. \tag{3.12}$$

The connection between (3.11) and (3.12) (and between (3.10) and its respective local maximum condition) is valid for all smooth test functions, so we have achieved what we were hoping for: we put the derivatives on test functions with full control over them.

3.2 Definition and Basic Properties

The discussion in the previous chapter leads to the following final definition of the viscosity solution.

Definition 3.1. *Assume, that u is a bounded and uniformly continuous function on $\mathbb{R}^d \times [0, T]$ for any $T > 0$. We say that u is a* viscosity subsolution *(or respectively, a* viscosity supersolution*) of a problem (1.1), if:*

1. $u = g$ *on* $\mathbb{R}^d \times \{t = 0\}$,

2. *for each* $v \in C^\infty(\mathbb{R}^d) \times (0, \infty)$, *we have that, if* $u - v$ *has a local maximum (or, respectively, a local minumum) at a point* $(x_0, t_0) \in \mathbb{R}^d \times (0, \infty)$, *then,*
   $v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \leq 0$
   or, respectively,
   $v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \geq 0.$

u is called a viscosity solution, *if it's both viscosity subsolution and viscosity supersolution.*

Remark. *In this paper we'll use the definition above, but it's worth mentioning, that it can be defined in a much more general case. We'll briefly introduce it in this remark.*
We are looking at the equation in form $F(x, u, D_x u, D_x^2 u) = 0$, *where* $F = F(x, r, p, X)$ *is a function* $F : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{S}(\mathbb{R}^d)$, *where* $\mathcal{S}(\mathbb{R}^d)$ *is the set of symmetric matrices* $d \times d$ *equipped with the standard order. We have only one requirement to put on* F*: it has to satisfy the following proper degenerate ellipticity condition:*

$$F(x, r, p, X) \leq F(x, s, p, Y), \text{ when } r \leq s, Y \leq X \tag{3.13}$$

Let $\Omega \subset \mathbb{R}^d$ *and let* F *satisfy the above condition (3.13). A viscosity subsolution to* $F(x, u, D_x u, D_x^2 u) = 0$ *on* Ω *is then an upper semicontinuous function* $u : \Omega \to \mathbb{R}$, *such that for any* $v \in C^\infty(\mathbb{R}^d)$, *such that* $u - v$ *has a local maximum at* $x_0 \in \Omega$, *it follows that*

$$F(x_0, u(x_0), D_x v(x_0), D_x^2 v(x_0)) \leq 0. \tag{3.14}$$

And similarly, a supersolution is a lower semicontinuous function $u : \Omega \to \mathbb{R}$, *such that for any* $v \in C^\infty(\mathbb{R}^d)$, *such that* $u - v$ *has a local minimum at* $x_0 \in \Omega$, *it follows that*

$$F(x_0, u(x_0), D_x v(x_0), D_x^2 v(x_0)) \geq 0. \tag{3.15}$$

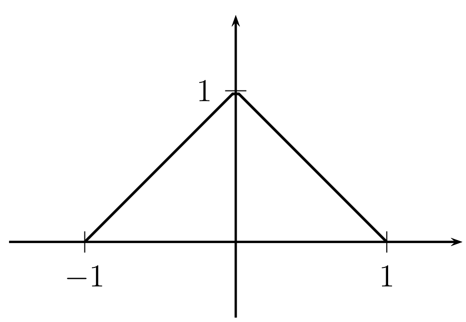
u is a viscosity solution, if it's both a subsolution and a supersolution. More about this definition and the corresponding properties can be found in [8].

Example 3.1. *Let's consider the following time-independent Hamilton-Jacobi equation in* $(-1, 1) \subset \mathbb{R}$*:*

$$|u'(x)| = 1 \tag{3.16}$$

with boundary conditions $u(-1) = u(1) = 0$. *In example 2.2 we have shown that this problems has infinitely many functions, that satisfy it in a weak sense and in this example we will illustrate, that only one of them is the viscosity solution. We'll show*

it by definition.
If u is a viscosity solution, then for any $v \in C^\infty(\mathbb{R})$, such that x_0 is a maximum (minimum) point of $u - v$, we need to have $|v'(x_0)| \leq 1$ (≥ 1). This means, that if we consider a smooth function $v(x) = c(x-1)$ for $c < -1$, then we should have $u \leq v$, for otherwise, there would exist a maximizing point x_0 for $u - v$ and $|v'(x)| = |c| > 1$, which contradicts our assumption, that u is a viscosity solution. Letting $c \to -1$, we get that $u \leq 1 - x$.
Similarly, by putting $v(x) = c(x+1)$ for $c > 1$ and letting $c \to 1$, we get that $u \leq 1 + x$, which means that $u \leq 1 - |x|$.
Finally, try to get v by smoothing the tip of $c(1 - |x|)$ where $0 < c < 1$ and see, that if there exists $\tilde{x} \in (-1, 1)$, such that $u(\tilde{x}) \leq v(\tilde{x})$, then there must be a minimizing point x_0, in which the following should be true $|v'(x_0)| = |c| \geq 1$, which contradicts our assumption on c. Thus we have shown that $u \geq v$ for any smooth v, arbitrarily close to $1 - |x|$. Along with the previous estimate, this proves that $u(x) = 1 - |x|$ is a viscosity solution of $|u'(x)| = 1$.
Notice, however, that u wouldn't be a viscosity solution of $1 - |u'(x)| = 0$, because then, if x_0 is a maximum (minimum) point of $u - v$, we would need to have $|v'(x_0)| \geq 1$ (≤ 1). In this case, the viscosity solution is $u(x) = |x| - 1$. This example illustrates, that the viscosity solution allows the orientation of "corners" or "bumps" in only one directions, so in general, if u solves $H(D_x u, u) = 0$ in a viscosity sense, it will not necessarily be a viscosity solution to $-H(D_x u, u) = 0$.
Notice also, that this equation is exactly the eikonal equation (1.33) in 1-dimension in the medium with uniform refraction index 1 and the viscosity solution is the one that has the physical meaning. The refraction index determines the speed of light at each point and and in out case the speed of light is the same everywhere, thus the time for travelling from the center of the domain ($x = 0$) to each of the boundaries will take the same time, which the graph of the function illustrates.



Remark. *It's worth mentioning, that the example above suggests another interpretation of viscosity solutions. They may be seen as a sort of uniquification of multiple* *solutions to the problem.*

The discussion in previous chapter showed, that the solution, obtained by the method of vanishing viscosity, thus, defined by (3.4), is a viscosity solution. Next we'll show some important properties of solutions, defined by Definition 3.1.
The first thing that comes to mind is whether a classical solution is a viscosity solution and vice versa. We'll prove the following result.

Theorem 3.2 (Consistency of viscosity solutions)**.** *The following facts are true:*

1. *If $u \in C^1(\mathbb{R}^d \times [0, \infty))$ is a classical solution of Hamilton-Jacobi PDE, i.e. it solves (1.1) and is bounded and uniformly continuous, then u is a viscosity solution.*

2. *If u is a viscosity solution and is differentiable at some point $(x_0, t_0) \in \mathbb{R}^d \times (0, \infty)$. Then*
$$u_t(x_0, t_0) + H(D_x(x_0, t_0), x_0) = 0. \tag{3.17}$$

Proof. The first fact is easy to prove. Let's take any test function $v \in C^\infty(\mathbb{R}^d \times (0, \infty))$ and assume, that $u - v$ attains a local extremum at (x_0, t_0), then

$$\begin{cases} D_x u(x_0, t_0) = D_x v(x_0, t_0) \\ u_t(x_0, t_0) = v_t(x_0, t_0), \end{cases}$$

from where it immediately follows:

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) = u_t(x_0, t_0) + H(D_x u(x_0, t_0), x_0) = 0.$$

Hence, by definition, u is a viscosity solution.
Before we proceed with proof of the second fact, we need to prove a lemma.

Lemma 3.1 (Touching by a C^1 function)**.** *Let $u : \mathbb{R}^d \to \mathbb{R}$ is continuous and differentiable at some point $x_0 \in \mathbb{R}^d$. Then there exists a function $v \in C^1(\mathbb{R}^d)$, such that*
$$\begin{cases} u(x_0) = v(x_0) \\ u - v \textit{ has a strict local maximum at } x_0. \end{cases} \tag{3.18}$$

Proof. We may assume, that $x_0 = 0$ and $u(0) = Du(0) = 0$, or otherwise consider $\tilde{u} = u(x + x_0) - u(x_0) - Du(x_0) \cdot x$ instead.

Under these assumptions, we have that $u(x) = |x|\rho_1(x)$, where $\rho_1 : \mathbb{R}^d \to \mathbb{R}$ is a continuous function, such that $\rho_1(0) = 0$. We next set for $r > 0$:

$$\rho_2(r) := \max_{x \in B_r} \{|\rho_1(x)|\} \tag{3.19}$$

Then $\rho_2(0) = 0$ and ρ_2 is non-decreasing. Now we define a function v:

$$v(x) := \int_{|x|}^{2|x|} \rho_2(r)dr + |x|^2. \tag{3.20}$$

Due to monotonicity of ρ_2, it follows, that

$$|v(x)| \leq |x|\rho_2(2|x|) + |x|^2,$$

hence $v(0) = Dv(0) = 0$. For other $x \neq 0$, we have

$$Dv(x) = \frac{2x}{|x|}\rho_2(2|x|) - \frac{x}{|x|}\rho_2(|x|) + 2x,$$

and therefore $v \in C^1(\mathbb{R}^d)$. Finally we have to show the strict local maximum property. If $x \neq 0$, then

$$\begin{aligned} u(x) - v(x) &= |x|\rho_1(x) - \int_{|x|}^{2|x|} \rho_2(r)dr - |x|^2 \\ &\leq |x|\rho_1(|x|) - \int_{|x|}^{2|x|} |\rho_1(|x|)|dr - |x|^2 \\ &\leq |x|\rho_1(|x|) - |x||\rho_1(|x|)| - |x|^2 \\ &\leq -|x|^2 < 0 = u(0) - v(0). \end{aligned}$$

□

Now we return to the proof of the second fact of the theorem. Applying the lemma above to u in $\mathbb{R}^{d+1}$, we deduce, that there exists a C^1 function v, such that $u - v$ has a strict maximum at (x_0, t_0). Next we take the standard mollifier η_ε and denote $v^\varepsilon := \eta_\varepsilon * v$. Then by the properties of mollification, we obtain:

$$\begin{cases} v^\varepsilon \to v \\ D_x v^\varepsilon \to D_x v \\ v_t^\varepsilon \to v_t, \end{cases} \tag{3.21}$$

uniformly near (x_0, t_0). Then it follows, that $u - v^\varepsilon$ has a maximum at some point $(x_\varepsilon, t_\varepsilon)$, such that $(x_\varepsilon, t_\varepsilon) \to (x_0, t_0)$ as $\varepsilon \to 0$. Since a mollification v^ε is a smooth function, we can apply the definition of a viscosity solution and deduce, that:

$$v_t^\varepsilon(x_0, t_0) + H(D_x v^\varepsilon(x_0, t_0), x_0) \leq 0. \tag{3.22}$$

Passing to limit $\varepsilon \to 0$, we get

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \leq 0. \tag{3.23}$$

But since u is differentiable at (x_0, t_0), at $u - v$ attains a maximum at (x_0, t_0), we see that

$$\begin{cases} D_x u(x_0, t_0) = D_x v(x_0, t_0) \\ u_t(x_0, t_0) = v_t(x_0, t_0). \end{cases} \tag{3.24}$$

so, also

$$u_t(x_0, t_0) + H(D_x u(x_0, t_0), x_0) \leq 0. \tag{3.25}$$

Finally, we can repeat the steps above for the function $-u$, then we'll find a C^1 function $\tilde{v}$, such that $u - v$ has a strict local minimum at (x_0, t_0). Then we can deduce in a similar manner, that

$$u_t(x_0, t_0) + H(D_x u(x_0, t_0), x_0) \geq 0, \tag{3.26}$$

thus completing the proof. □

Next we prove the stability result. In general, for non-linear first-order PDE, the set of solutions is not necessarily closed in the topology of uniform convergence. We may not conclude, that the uniform limit of solutions of some problem will be a solution itself without ensuring the uniform convergence of the corresponding gradients too. However, in the case of viscosity solutions, we can skip this requirement on the convergence of derivatives. The result is originally due to Crandall and Lions in [9].

Theorem 3.3 (Stability of Viscosity Solutions)**.** *Let H_n be a sequence of continuous Hamiltonians converging in $C(\mathbb{R}^d \times [0, \infty))$ to H. Let u_n be a subsolution (or, respectively, supersolution) of $u_t + H_n(D_x u(x, t), x) = 0$, and let $u_n \to u$ uniformly in $\mathbb{R}^d$. Then u is a viscosity solution of $u_t + H(D_x u(x, t), x) = 0$.*

Proof. Before we prove the theorem, let us first prove the following lemma.

Lemma 3.2. *Let $u : \mathbb{R}^d \to \mathbb{R}$ be a continuous function. Let v be a smooth function, such that $u - v$ has a strict local maximum (minumum) at x_0. If u_n is a sequence, converging to u uniformly, then there exists a sequence of points x_n, such that $x_n \to x_0$ with $u_n(x_n) \to u(x)$ and, moreover, $u_m - v$ has a local maximum (minumum) at x_m.*

Proof. We prove the result in case of $u - v$ having a local maximum. Then for every $\delta > 0$ there exists a small enough $\varepsilon_\delta > 0$, such that when $|x - x_0| = \delta$

$$u(x) - v(x) < u(x_0) - v(x_0) - \varepsilon_\delta. \tag{3.27}$$

Now, by the uniform convergence of u_n, there exists a sufficiently large N_δ, such that for all $m \geq N_\delta$, we have $|u_m(x) - u(x)| < \frac{\epsilon_\delta}{4}$. Then for $|x - x_0| = \delta$

$$\begin{aligned}
&u_m(x) - v(x) - u_m(x_0) + v(x_0) \\
&= (u_m(x) - u(x)) + (u(x) - v(x)) - u_m(x_0) + v(x_0) \\
&< \frac{\epsilon_\delta}{4} + u(x_0) - v(x_0) - \varepsilon_\delta - u_m(x_0) + v(x_0) \\
&< -\frac{\epsilon_\delta}{2}.
\end{aligned} \tag{3.28}$$

Or, rewriting it,

$$u_m(x) - v(x) < u_m(x_0) - v(x_0) - \varepsilon_\delta. \tag{3.29}$$

Which means that $u_m - v$ has a local maximum at some point x_m with $|x_m - x| < \delta$. We can construct the rest of x_m by letting $\delta, \varepsilon_\delta \to 0$.

□

Now, let's prove the theorem in case of subsolutions (the case of supersolutions is obtained in a similar manner). We take v, a test function, and let $u - v$ have a local maximum at (x_0, t_0). Then by the lemma there exists a sequence (x_m, t_m), such that $x_m \to x_0$, such that $u_m - v$ has a local maximum at (x_m, t_m) and $u_m(x_m, t_m) \to u(x_0, t_0)$. Then, since u_m is a subsolution, then

$$u_{mt}(x_m, t_m) + H_m(D_x u_m(x_m, t_m), x_m) \leq 0. \tag{3.30}$$

By letting $m \to \infty$, we finish the proof, using the continuity:

$$u_t(x_0, t_0)) + H(D_x u(x_0, t_0), x_0) \leq 0. \tag{3.31}$$

□

3.3 Hopf-Lax Formula as a Viscosity Solution

Finally, we'll show, that the Hopf-Lax formula for the problem we discussed in subsection 2.3, is consistent with the theory of viscosity solutions. We remind ourselves, that we are looking at a problem

$$\begin{cases} u_t + H(D_x u) = 0 \text{ in } \mathbb{R}^d \times (0, T] \\ u = g \text{ on } \mathbb{R}^d \times t = 0, \end{cases} \tag{3.32}$$

where H is a convex map, that satisfies the superlinearity condition:

$$\lim_{|p|\to\infty} \frac{H(p)}{|p|} = +\infty,$$

and g is Lipschitz continuous. In subsection 2.3 we showed, that Hopf-Lax formula provides the sort of weak solution to the given problem. Now we want to show, that it provides a correct viscosity solution.

Theorem 3.4 (Hopf-Lax formula as a viscosity solution)**.** *Assume, that H is as above, and $L = H^*$ its Legendre transform. The initial condition g is a bounded Lipschitz continuous function. Then u, defined by the Hopf-Lax formula for each $x \in \mathbb{R}^d$, $t > 0$:*

$$u(x,t) = \min_{y\in\mathbb{R}^d} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\},$$

gives the unique viscosity solution of the problem (3.32).

Proof. The accordance with the initial condition was already proved in subsection 2.3, when we were proving, that the Hopf-Lax formula provides a weak solution. So we prove that it's a viscosity solution.
Let $v \in C^\infty(\mathbb{R}^d \times (0,\infty))$ and assume, $u - v$ has a local maximum at (x_0, y_0), i.e.

$$u(x_0,t_0) - v(x_0,t_0) \geq u(x,t) - v(x,t). \tag{3.33}$$

By formula (2.24), we get the following formula for u

$$u(x_0,t_0) = \min_{x\in\mathbb{R}^d} \left\{ (t_0 - t)L\left(\frac{x_0 - x}{t_0 - t}\right) + u(x,t) \right\},$$

which in its turn means, that for $0 \leq t < t_0$, $x \in \mathbb{R}^d$

$$u(x_0,t_0) \leq (t_0 - t)L\left(\frac{x_0 - x}{t_0 - t}\right) + u(x,t).$$

Combining the estimates above, we obtain for $t < t_0$ and (x,t) close to (x_0,t_0).

$$v(x_0,t_0) - v(x,t) \leq (t_0 - t)L\left(\frac{x_0 - x}{t_0 - t}\right). \tag{3.34}$$

If we introduce now $h = t_0 - t > 0$ and let $x = x_0 - hq$, then the inequality becomes

$$v(x_0,t_0) - v(x_0 - hq, t_0 - h) \leq hL(q). \tag{3.35}$$

By dividing the inequality by h and letting $h \to 0$, we obtain for all $q \in \mathbb{R}^d$

$$v_t(x_0, t_0) + D_x v(x_0, t_0) \cdot q - L(q) \leq 0$$

Now, from convex duality and the definition of the Legendre transform, we have for each $p, q \in \mathbb{R}^d$

$$H(p) = \sup_{q \in \mathbb{R}^d} \{p \cdot q - L(q)\}.$$

Using this, we obtain

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0)) \leq 0, \tag{3.36}$$

which means that u is a viscosity subsolution of (3.32).
Next we prove, that it's also a viscosity supersolution. Suppose that $u - v$ has a local minimum at (x_0, t_0). Suppose, by contradiction, that the there is some $\delta > 0$, such that

$$v_t(x, t) + H(D_x v(x, t)) \leq -\delta < 0, \tag{3.37}$$

for all points (x, t), close to (x_0, t_0). Again, using the convex duality and the definition of Legendre transform, we get for all $q \in \mathbb{R}^d$.

$$v_t(x, t) + D_x v(x, t) \cdot q - L(q) \leq -\delta.$$

By formula (2.24), for small enough h there exists some $x_1 \in \mathbb{R}^d$, close to x_0, such that

$$u(x_0, t_0) = hL\left(\frac{x_0 - x_1}{h}\right) + u(x_1, t_0 - h).$$

Next we compute

$$\begin{aligned} v(x_0, t_0) - v(x_1, t_0 - h) &= \int_0^1 \frac{d}{ds} v(sx_0 + (1-s)x_1, t_0 + (s-1)h) ds \\ &= \int_0^1 D_x v(sx_0 + (1-s)x_1, t_0 + (s-1)h) \cdot (x_0 - x_1) \\ &+ v_t(sx_0 + (1-s)x_1, t_0 + (s-1)h) h ds \\ &= h\left(\int_0^1 D_x v \cdot q + v_t ds\right). \end{aligned}$$

Here $q = \frac{x_0 - x_1}{h}$. If $h > 0$ is small enough, we can use our contradictory assumption (3.37), and get

$$v(x_0, t_0) - v(x_1, t_0 - h) \leq hL\left(\frac{x_0 - x_1}{h}\right) - \delta h,$$

which in its turn, means by

$$v(x_0, t_0) - v(x_1, t_0 - h) \leq u(x_0, t_0) - u(x_1, t_0 - h) - \delta h,$$

which contradicts the local minimality of $u - v$ at (x_0, t_0). This contradiction proves that u is a viscosity supersolution and hence a viscosity solution. □

3.4 Uniqueness

In this and the following sections we'll be following the chapters 10.2 and 10.3 from [11]. The main focus of this chapter is the uniqueness result. We will show under which assumptions is the viscosity solution of the problem (1.1) is unique. We'll consider a problem for fixed time $T > 0$:

$$\begin{cases} u_t + H(D_x u, x) = 0 \text{ in } \mathbb{R}^d \times (0, T] \\ u(x, 0) = g \text{ on } \mathbb{R}^d \times \{t = 0\} \end{cases} \tag{3.38}$$

First we prove a lemma, showing that the inequalities from the definition of the viscosity solution, can be extended to the terminal time T.

Lemma 3.3. *Assume, u is a viscosity solution of (3.38) and $u - v$ has a local maximum (minimum) at a point $(x_0, t_0) \in \mathbb{R}^d \times (0, T]$. Then*

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \leq 0 (\geq 0) \tag{3.39}$$

Proof. We assume that $u - v$ has a strict local maximum at the point (x_0, T) (we can get rid of the strictness by the same technique we used in subsection 3.1. Define for $x \in \mathbb{R}^d$ and $0 < t < T$

$$\tilde{v}(x, t) := v(x, t) + \frac{\varepsilon}{T - t}.$$

Then for a small enough $\varepsilon > 0$, $u - \tilde{v}$ has a local maximum at $(x_\varepsilon, t_\varepsilon)$, where $0 < t_\varepsilon < T$ and $(x_\varepsilon, t_\varepsilon) \to (x_0, T)$. Therefore,

$$\tilde{v}_t(x, t) + H(D_x \tilde{v}(x_\varepsilon, t_\varepsilon), x_\varepsilon) \leq 0,$$

and hence

$$v_t(x, t) + \frac{\varepsilon}{(T - t)^2} + H(D_x v(x_\varepsilon, t_\varepsilon), x_\varepsilon) \leq 0.$$

By letting now $\varepsilon \to 0$, we get

$$v_t(x_0, T) + H(D_x v(x_0, T), x_0) < 0.$$

Which proves the theorem for the case of $u - v$ having a local maximum. The case of local minimum is proved by reversing the inequalities. □

Now we are going to prove one of most important facts in the theory of viscosity solutions for Hamilton-Jacobi equations.

Theorem 3.5 (Uniqueness of viscosity solutions)**.** *Assume, that Hamiltonian H satisfies the following conditions:*

$$\begin{cases} |H(p,x) - H(q,x)| \leq C|p-q| \\ |H(p,x) - H(p,y)| \leq C|x-y|(1+|p|) \end{cases} \tag{3.40}$$

for all $x, y, p, q \in \mathbb{R}^d$ and for some constant $C \geq 0$. Then there exists at most one viscosity solution of (1.1).

Proof. Assume, that we have u and $\tilde{u}$ to be different viscosity solutions, that both satisfy the initial condition, but

$$\sigma := \sup_{\mathbb{R}^d \times [0,T]} (u - \tilde{u}) > 0. \tag{3.41}$$

We choose $\varepsilon > 0$ and $\lambda < 1$ and define for $x, t \in \mathbb{R}^d$, $t, s > 0$

$$\Phi(x,y,t,s) := u(x,t) - \tilde{u}(y,s) - \lambda(t+s) - \frac{1}{\varepsilon^2}(|x-y|^2 + (t-s)^2 - \varepsilon(|x|^2 + |y|^2).$$

Then there exists a point $(x_0, y_0, t_0, s_0) \in \mathbb{R}^{2d} \times [0,T]^2$, such that

$$\Phi(x_0, y_0, t_0, s_0) = \max_{\mathbb{R}^{2d} \times [0,T]^2} \Phi(x,y,t,s). \tag{3.42}$$

Now fix ε and λ so small, that

$$\Phi(x_0, y_0, t_0, s_0) \geq \sup_{\mathbb{R}^d \times [0,T]} \Phi(x,x,t,t) \geq \frac{\sigma}{2}.$$

In addition, $\Phi(x_0, y_0, t_0, s_0) \geq \Phi(0,0,0,0)$, and hence

$$\begin{aligned} \lambda(t_0 + s_0) + \frac{1}{\varepsilon^2}(|x_0 - y_0|^2 + (t_0 - s_0)^2) + \varepsilon(|x_0|^2 + |y_0|^2) \\ \leq u(x_0, t_0) - \tilde{u}(y_0, s_0) - u(0,0) + \tilde{u}(0,0). \end{aligned} \tag{3.43}$$

Since u and $\tilde{u}$ are bounded, we get that

$$|x_0 - y_0|, |t_0 - s_0| = O(\varepsilon). \tag{3.44}$$

Moreover, $\varepsilon(|x_0|^2 + |y_0|^2) = O(1)$, and therefore

$$\begin{aligned}\varepsilon(|x_0| + |y_0|) &= \varepsilon^{\frac{1}{4}}\varepsilon^{\frac{3}{4}}(|x_0| + |y_0|) \\ &\leq \varepsilon^{\frac{1}{2}} + C\varepsilon^{\frac{3}{2}}(|x_0|^2 + |y_0|^2) \\ &\leq C\varepsilon^{\frac{1}{2}}\end{aligned} \tag{3.45}$$

Which means, that $\varepsilon(|x_0|+|y_0|) = O(\varepsilon^{\frac{1}{2}})$. Next, due to the fact, that $\Phi(x_0, y_0, t_0, s_0) \geq \Phi(x_0, x_0, t_0, t_0)$, we have

$$\begin{aligned}u(x_0, t_0) - \tilde{u}(y_0, s_0) - \lambda(t_0 + s_0) - \frac{1}{\varepsilon^2}(|x_0 - y_0|^2 + (t_0 - s_0)^2) \\ - \varepsilon(|x_0|^2 + |y_0|^2) \geq u(x_0, t_0) - \tilde{u}(x_0, t_0) - 2\lambda t_0 - 2\varepsilon|x_0|^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{\varepsilon^2}(|x_0 - y_0|^2 + (t_0 - s_0)^2) \leq \tilde{u}(x_0, t_0) - \tilde{u}(y_0, s_0) + \lambda(t_0 - s_0) \\ + \varepsilon(x_0 + y_0) \cdot (x_0 - y_0).\end{aligned}$$

By this and the uniform continuity of $\tilde{u}$, we have $|x_0 - y_0|, |t_0 - s_0| = o(\varepsilon)$.
Next, let ω be a modulus of continuity of u, meaning, that for all $x, y \in \mathbb{R}^d$, $0 \leq t, s \leq T$, we have that

$$|u(x, t) - u(y, s)| \leq \omega(|x - y| + |t - s|), \tag{3.46}$$

and $\omega(r) \to 0$ as $r \to 0$. Similarly we define $\tilde{\omega}$. Then

$$\begin{aligned}\frac{\sigma}{2} \leq u(x_0, t_0) - \tilde{u}(y_0, s_0) &= u(x_0, t_0) - u(x_0, 0) + u(x_0, 0) - \tilde{u}(x_0, 0) \\ &\quad + \tilde{u}(x_0, 0) - \tilde{u}(x_0, t_0) + \tilde{u}(x_0, t_0) - \tilde{u}(y_0, s_0) \\ &\leq \omega(t_0) + \tilde{\omega}(t_0) + \tilde{\omega}(o(\varepsilon))\end{aligned}$$

Now we can make ε even smaller, so that $\frac{\sigma}{4} \leq \omega(t_0) + \tilde{\omega}(t_0)$, which in turn implies, that $t_0, s_0 \geq \mu > 0$ for some constant μ.
Now, in the view of (3.42), the map $(x, t) \mapsto \Phi(x, y_0, t, s_0)$ has a maximum at (x_0, t_0), then by definition of Φ, $u - v$ has a maximum at (x_0, t_0), where v is defined by

$$v(x, t) := \tilde{u}(y_0, s_0) + \lambda(t + s_0) + \frac{1}{\varepsilon^2}(|x - y_0|^2 + (t - s_0)^2) + \varepsilon(|x| + |y_0|^2).$$

Since u is a viscosity solution, we have by definition

$$v_t(x_0, t_0) + H(D_x v(x_0, t_0), x_0) \leq 0. \tag{3.47}$$

Hence,

$$\lambda + \frac{2(t_0 - s_0)}{\varepsilon^2} + H\left(\frac{2}{\varepsilon^2}(x_0 - y_0) + 2\varepsilon x_0, x_0\right) \leq 0. \tag{3.48}$$

Similarly, the map $(y, s) \mapsto -\Phi(x_0, y, t_0, s)$ has a minimum at (y_0, s_0), then $\tilde{u} - \tilde{v}$ has a minimum at (y_0, s_0) for $\tilde{v}$ defined by

$$\tilde{v}(y, s) := u(x_0, t_0) - \lambda(t_0 + s) - \frac{1}{\varepsilon^2}(|x - y_0|^2 + (t_0 - s)^2) - \varepsilon(|x_0|^2 + |y|^2).$$

Again, since $\tilde{u}$ is a viscosity solution, then

$$\tilde{v}_s(y_0, s_0) + H(D_y\tilde{v}(y_0, s_0), y_0) \geq 0.$$

Hence

$$-\lambda + \frac{2(t_0 - s_0)}{\varepsilon^2} + H\left(\frac{2}{\varepsilon^2}(x_0 - y_0) - 2\varepsilon y_0, y_0\right) \geq 0. \tag{3.49}$$

Now, combining the two results, we get

$$2\lambda \leq H(\frac{2}{\varepsilon^2}(x_0 - y_0) - 2\varepsilon y_0, y_0) - H\left(\frac{2}{\varepsilon^2}(x_0 - y_0) + 2\varepsilon x_0, x_0\right). \tag{3.50}$$

In the view of the hypothesis on H, we obtain

$$\lambda \leq C\varepsilon(|x_0| + |y_0|) + C|x_0 - y_0|\left(1 + \frac{|x_0 - y_0|}{\varepsilon^2} + \varepsilon(|x_0| + |y_0|)\right) \tag{3.51}$$

Letting $\varepsilon \to 0$, we get that $0 < \lambda \leq 0$, which contradicts the assumption and hence ends the proof. ☐

3.5 Optimal Control Theory

In subsection 3.1 we mentioned one of the ways to establish the existence of viscosity solutions, obtained by the vanishing viscosity method. However, in this section we'll introduce another approach. It appears that the Hamilton-Jacobi equation is just a special case of the equation, central to the optimal control theory, namely, Hamilton-Jacobi-Bellman equation, and this connection allows us to use some techniques of the control theory to establish the existence of the viscosity solution. We'll start with an introduction into the optimal control theory.

3.5.1 Introduction

Optimal control theory is, in fact, an extension of the calculus of variation, dealing not just with an optimization problem, but with the control law for a system, such that the optimality conditions are achieved. Mathematically, the goal is to optimally control the solution to the following system of ordinary differential equations on a fixed time period $[t, T]$ ($t > 0$, $T > t$) by changing some parameters:

$$\begin{cases} \dot{x}(s) = f(x(s), \alpha(x)), \ (t < s < T) \\ x(t) = x. \end{cases} \tag{3.52}$$

Here, $x \in \mathbb{R}^d$ is an fixed initial point, f is a given bounded, Lipschitz continuous function $f : \mathbb{R}^d \times A \to \mathbb{R}^d$, where $A \in \mathbb{R}^k$ is a compact subset. The function α is called *a control*, i.e. a law adjusting parameters from the set A, thus changing the dynamics of the system. We denote the set of admissible controls

$$\mathcal{A} = \{\alpha : [0, T] \to A | \alpha \text{ is measurable}\}. \tag{3.53}$$

Due to the assumptions of f, the Picard-Lindelöf theorem yields that for each $\alpha \in \mathcal{A}$ there exists a unique Lipschitz continuous solution $x^\alpha(s)$ on $[t, T]$. Such a solution is called the *response* of the system to the control α. We have to find the optimal control α^*, but first we need to define the "optimality". We do it with the help of the *cost functional*, that is defined for a control α, given an initial condition $x \in \mathbb{R}^d$ to (3.52) and $0 \leq t \leq T$:

$$C_{x,t}[\alpha] := \int_t^T h(x^\alpha(s), \alpha(s))ds + g(x(T)). \tag{3.54}$$

Here $h : \mathbb{R}^d \times A \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ are given. These function have their name in the control theory: h is called the *running cost per unit time* and g the *the terminal cost*. We also assume, that h and g are bounded, h is Lipschitz continuous in the first variable, and g is Lipschitz continuous.

Now we can state the problem: given an initial point $x \in \mathbb{R}^d$ and a time $0 \leq t \leq T$, we need to find among all admissible controls an optimal control α^*, that minimizes the cost (3.54).

One of the techniques to deal with this problem is the method of *dynamic programming*. The idea is to look at the so-called *value function*:

$$u(x, t) := \inf_{\alpha \in \mathcal{A}} C_{x,t}[\alpha] \ (x \in \mathbb{R}^d, 0 < t < T). \tag{3.55}$$

By researching this function's behavior, we are embedding the original control problem (3.52) into the larger class of the similar problems, as we also let x and t vary. We named this function u for a reason, as it will appear, that u solves a PDE of Hamilton-Jacobi type. Our goal is to build this equation and to show that the solution of this PDE solves the optimal control problem for each x and t.

Theorem 3.6 (Optimality Conditions)**.** *For all $h > 0$, such that $t + h < T$, we get*

$$u(x,t) = \inf_{\alpha \in A} \left\{ \int_t^{t+h} h(x^\alpha(s), \alpha(s))ds + u(x^\alpha(t+h), t+h) \right\}. \tag{3.56}$$

Here x^α is as before - a solution of ODE (3.52) under the control α.

Proof. We choose any $\alpha_1 \in \mathcal{A}$ and let x_1 be a corresponding solution to the (3.52). Then we fix $\alpha_2 \in \mathcal{A}$ and let x_2 be its corresponding solution for initial time $t + h$, such that the following is true

$$u(x_1(t+h), t+h) + \varepsilon \geq \int_{t+h}^{T} h(x_2(s), \alpha_2(s))ds + g(x_2(T)), \tag{3.57}$$

Next we define a control

$$\alpha_3(s) := \begin{cases} \alpha_1(s), & \text{if } t \leq s < t+h \\ \alpha_2(s), & \text{if } t+h \leq s \leq T \end{cases} \tag{3.58}$$

The corresponding solution x_3 will have the following form (by the uniqueness of solutions to (3.52)) :

$$x_3(s) = \begin{cases} x_1(s), & \text{if } t \leq s < t+h \\ x_2(s), & \text{if } t+h \leq s \leq T. \end{cases} \tag{3.59}$$

Then by definition of $u(x,t)$ and the inequalities above we have

$$\begin{aligned} u(x,t) \leq C_{x,t}[\alpha_3] &= \int_t^T h(x_3(s), \alpha_3(s))ds + g(x_3(T)) \\ &= \int_t^{t+h} h(x_1(s), \alpha_1(s))ds + \int_{t+h}^{T} h(x_2(s), \alpha_2(s))ds + g(x_2(T)) \\ &\leq \int_t^{t+h} h(x_1(s), \alpha_1(s))ds + u(x_1(t+h), t+h) + \varepsilon. \end{aligned}$$

Since α_1 was arbitrary, we get

$$u(x,t) \le \inf_{\alpha\in A} \left\{ \int_t^{t+h} h(x^\alpha(s), \alpha(s))ds + u(x^\alpha(t+h), t+h) \right\} + \varepsilon. \tag{3.60}$$

To get the reversed inequality and complete the proof, choose a new $\varepsilon > 0$, then $\alpha_4 \in \mathcal{A}$ with the corresponding solution x_4 for the initial time t, such that

$$u(x,t) + \varepsilon \ge \int_t^T (h(x_4(s), \alpha_4(s))ds + g(x_4(T)), \tag{3.61}$$

Now, again from the definition of u, we observe:

$$u(x_4(t+h), t+h) \le \int_{t+h}^T h(x_4(s), \alpha_4(s))ds + g(x_4(T)). \tag{3.62}$$

Hence, summing the two above inequalities

$$\begin{aligned} u(x,t) + \varepsilon &\ge \int_t^T (h(x_4(s), \alpha_4(s))ds - \int_{t+h}^T h(x_4(s), \alpha_4(s))ds + u(x_4(t+h), t+h) \\ &= \int_t^{t+h} (h(x_4(s), \alpha_4(s))ds + u(x_4(t+h), t+h) \end{aligned}$$

Taking the infimum for $\alpha \in \mathcal{A}$, we finish the proof. □

3.5.2 Hamilton-Jacobi-Bellman Equation

Now we are going to rewrite the optimality conditions (3.56) as a PDE. First we'll prove some useful properties of the value function u.

Lemma 3.4. *The value function u, defined by (3.55) is bounded and Lipschitz continuous.*

Proof. By boundedness assumptions on g and h, it follows, that for any α, $x \in \mathbb{R}^d$, and $0 \le t \le T$

$$C_{x,t}[\alpha] \le C_h \int_t^T ds + C_g \le C_h T + C_g = C. \tag{3.63}$$

By taking an infimum by α, we get the boundedness of u.
Next we prove the Lipschitz continuity in x. Let's fix $x, \hat{x} \in \mathbb{R}^d$ and $0 \le t \le T$ and

take $\varepsilon > 0$. We choose $\hat{\alpha} \in \mathcal{A}$ and a corresponding solutions $\hat{x}(s)$ and $x(s)$ to (3.52) with initial value $\hat{x}$ and x, so that

$$u(\hat{x}, t) + \varepsilon \geq \int_t^T h(\hat{x}(s), \hat{\alpha}(s))ds + g(\hat{x}(T)). \tag{3.64}$$

Since f is Lipschitz continuous, we get:

$$|\dot{x}(s) - \dot{\hat{x}}(s)| \leq C|x(s) - \hat{x}(s)|,$$

Then by Grönwall's inequality

$$|x(s) - \hat{x}(s)| \leq \tilde{C}|x(t) - \hat{x}(t)| = \tilde{C}|x - \hat{x}|$$

Using this result, the Lipschitz continuity of h and g and the previous estimate (3.64), we deduce

$$\begin{aligned} u(x,t) - u(\hat{x},t) &\leq \int_t^T (h(x(s), \hat{\alpha}(s)) - h(\hat{x}(s), \hat{\alpha}(s))ds \\ &\quad + g(x(T)) - g(x(T)) + \varepsilon \\ &\leq C_h \int_t^T |\hat{x}(s) - x(s)|ds + C_g|\hat{x}(T) - x(T)| + \varepsilon \\ &\leq C_0|x - \hat{x}| + \varepsilon. \end{aligned}$$

If we repeat all the steps reversing the roles of x and $\hat{x}$, we get the Lipschitz continuity of u:

$$|u(x,t) - u(\hat{x},t)| \leq C_0|x - \hat{x}|. \tag{3.65}$$

Next take $x \in \mathbb{R}^d$ and $0 \leq t < \hat{t} \leq T$. Let $\varepsilon > 0$ and choose $\alpha \in \mathcal{A}$ with a corresponding solution $x(s)$, such that

$$u(x, t) + \varepsilon \geq \int_t^T h(x(s), \alpha(s))ds + g(x(T)). \tag{3.66}$$

Define for $\hat{t} \leq s \leq T$: $\hat{\alpha}(s) := \alpha(s + t - \hat{t}$. Let $\hat{x}$ be a corresponding solution with initial time $\hat{t}$, meaning

$$\begin{cases} \dot{\hat{x}}(s) = f(\hat{x}(s), \hat{\alpha}(x)), \ (\hat{t} < s < T) \\ \hat{x}(\hat{t}) = x. \end{cases} \tag{3.67}$$

Then $\hat{x}(s) = x(s + t - \hat{t})$. Therefore,

$$
\begin{aligned}
u(x,\hat{t}) - u(x,t) &\leq \int_{\hat{t}}^{T} h(\hat{x}(s),\hat{\alpha}(s))ds + g(\hat{x}(T) \\
&\quad - \int_{t}^{T} h(x(s),\alpha(s))ds - g(x(T)) + \varepsilon \\
&= -\int_{T+t-\hat{t}}^{T} h(x(s),\alpha(s))ds + g(x(T+t-\hat{t})) - g(x(T)) + \varepsilon \\
&\leq C|t-\hat{t}| + \varepsilon.
\end{aligned}
$$

For the second inequality, we choose $\hat{\alpha}$, so that

$$
u(x,\hat{t}) + \varepsilon \geq \int_{t}^{T} h(\hat{x}(s),\hat{\alpha}(s))ds + g(\hat{x}(T)), \tag{3.68}
$$

where

$$
\begin{cases} \dot{\hat{x}}(s) = f(\hat{x}(s),\hat{\alpha}(s)) \\ \hat{x}(\hat{t}) = x. \end{cases} \tag{3.69}
$$

Then we define

$$
\alpha(s) = \begin{cases} \hat{\alpha}(s+\hat{t}-t), \ \text{if } t \leq s \leq T+t-\hat{t} \\ \hat{\alpha}(T), \ \text{if } T+t-\hat{t} \leq s \leq T. \end{cases}
$$

And let x be a solution, corresponding to α. Then $x(s) = \hat{x}(s+\hat{t}-t)$, for $t \leq s \leq T+t-\hat{t}$. Therefore

$$
\begin{aligned}
u(x,t) - u(x,\hat{t}) &\leq \int_{t}^{T} h(x(s),\alpha(s))ds + g(x(T)) \\
&\quad - \int_{\hat{t}}^{T} h(\hat{x}(s),\hat{\alpha}(s))ds - g(\hat{x}(T) + \varepsilon \\
&= \int_{T+t-\hat{t}}^{T} h(x(s),\alpha(s))ds + g(x(T)) - g(x(T+t-\hat{t})) + \varepsilon \\
&\leq C|t-\hat{t}| + \varepsilon.
\end{aligned}
$$

This ends the proof. □

Definition 3.2. *The* Hamilton-Jacobi-Bellman *equation is the equation of the following form*

$$
u_t + H(D_x u, x) = 0, \ \textit{in } \mathbb{R}^d \times (0,T), \tag{3.70}
$$

where the Hamiltonian H is defined for $x, p \in \mathbb{R}^d$ as

$$H(p,x) := \min_{a \in A}\{f(x,a) \cdot p + h(x,a)\}. \tag{3.71}$$

Remark. *Before we proceed with the discussion, we have to note, that in case of a terminal-value problem for the Hamilton-Jacobi equation, the definition of the viscosity solution is a bit different. We say, that a uniformly continuous bounded function u, having the terminal value $u = g$ on $\mathbb{R}^d \times \{t = T\}$ is* a viscosity subsolution *of (3.70), provided, that for each $v \in C^\infty(\mathbb{R}^d \times (0,T))$*

$$\begin{cases} \textit{if } u - v \textit{ has a local maximum at } (x_0, t_0) \in \mathbb{R}^d \times (0,T) \\ \textit{then } v_t(x_0,t_0) + H(D_x v(x_0,t_0), x_0) \geq 0, \end{cases} \tag{3.72}$$

and is a viscosity supersolution*, if*

$$\begin{cases} \textit{if } u - v \textit{ has a local minumum at } (x_0, t_0) \in \mathbb{R}^d \times (0,T) \\ \textit{then } v_t(x_0,t_0) + H(D_x v(x_0,t_0), x_0) \leq 0. \end{cases} \tag{3.73}$$

As before, the function u is a viscosity solution*, if it's both a viscosity subsolution and a supersolution. Note, that the inequalities are reversed in comparison with the initial value problem.*

Now we'll prove one of the major results of the control theory, that connects the Hamilton-Jacobi equations and the optimal control problem.

Theorem 3.7. *The value function u is the unique viscosity solution of the terminal value problem for the Hamilton-Jacobi-Bellman equation:*

$$\begin{cases} u_t + \min_{a \in A}\{f(x,a) \cdot D_x u + h(x,a)\} = 0 \textit{ in } \mathbb{R}^d \times (0,T) \\ u = g \textit{ on } \mathbb{R}^d \times t = T. \end{cases} \tag{3.74}$$

Proof. First, by the lemma (3.4), u is bounded and Lipschitz continuous. Moreover, by definition of u, for each $x \in \mathbb{R}^d$

$$u(x,T) = \inf_{\alpha \in \mathcal{A}} C_{x,T}[\alpha] = g(x). \tag{3.75}$$

We prove first, that u is a viscosity subsolution to the given problem. We'll prove it by contradiction. Assume, $v \in C^\infty(\mathbb{R}^d \times (0,T))$, and assume $u - v$ has a local maximum at $(x_0, t_0) \in \mathbb{R}^d \times (0, T)$. Suppose, the inequality from the definition of the subsolution doesn't hold, so there exists some $\delta > 0$ and $a \in A$, that

$$v_t(x,t) + f(x,a) \cdot D_x v(x,t) + h(x,a) \leq -\delta < 0 \tag{3.76}$$

for all points close to (x_0, t_0), i.e.

$$|x - x_0| + |t - t_0| < \epsilon. \tag{3.77}$$

Since $u - v$ has a local maximum at (x_0, t_0), we can choose ϵ above such that for all (x, t), satisfying the inequality above, the following is true

$$(u - v)(x,t) \leq (u - v)(x_0, t_0). \tag{3.78}$$

Now let's take a constant control $\alpha(s) = a$ for $t_0 \leq s \leq T$ and the corresponding solution $x(s)$:

$$\begin{cases} \dot{x}(s) = f(x(s), a) \\ x(t_0) = x. \end{cases} \tag{3.79}$$

Now choose $0 < h < \epsilon$ to be so small, that $|x(s) - x_0| \leq \epsilon$, for $t_0 \leq s \leq t_0 + h$. Then according to the estimates above, for each s, $t_0 \leq s \leq t_0 + h$, we have

$$v_t(x(s), s) + f(x(s), a) \cdot D_x v(x(s), s) + h(x(s), a) \leq -\delta.$$

Now using the maximality at (x_0, t_0)

$$\begin{aligned} u(x(t_0+h), t_0+h) - u(x_0, t_0) &\leq v(x(t_0+h), t_0+h) - v(x_0, t_0) \\ \int_{t_0}^{t_0+h} \frac{d}{ds} v(x(s), s) ds &= \int_{t_0}^{t_0+h} v_t(x(s), s) + D_x v(x(s), s) \cdot \dot{x}(s) ds \\ &= \int_{t_0}^{t_0+h} v_t(x(s), s) + f(x(s), a) \cdot D_x v(x(s), s) ds. \end{aligned} \tag{3.80}$$

Finally, by optimality conditions from formula (3.56), we have that

$$u(x_0, t_0) \leq \int_{t_0}^{t_0+h} h(x(s), a) ds + u(x(t_0+h), t_0+h).$$

Combining the last two results, we obtain

$$0 \leq \int_{t_0}^{t_0+h} v_t(x(s), s) + f(x(s), a) \cdot D_x v(x(s), s) + h(x(s), a) ds \leq -\delta h. \tag{3.81}$$

This contradicts the assumption, and hence proves that u is indeed a viscosity subsolution.
Now we assume, that $u - v$ has a local minimum at $(x_0, t_0) \in \mathbb{R}^d \times (0, T)$. Again we try to contradict the definition of a viscosity supersolution and assume, that there is $\delta > 0$, such that for all $a \in A$

$$v_t(x, t) + f(x, a) \cdot D_x v(x, t) + h(x, a) \geq \delta \tag{3.82}$$

for (x, t), sufficiently close to (x_0, t_0)

$$|x - x_0| + |t - t_0| \leq \epsilon$$

, where ϵ is chosen so that by definition of a local minumum at (x_0, t_0), we have

$$(u - v)(x, t) > (u - v)(x_0, t_0).$$

Next choose $0 < h < \epsilon$, so that $|x(s) - x_0| < \epsilon$ for $t_0 \leq s \leq t_0 + h$, where $x(s)$ is a solution, corresponding to some $\alpha \in \mathcal{A}$:

$$\begin{cases} \dot{x}(s) = f(x(s), \alpha(s)) \\ x(t_0) = x_0. \end{cases}$$

Next, we estimate

$$\begin{aligned} u(x(t_0 + h), t_0 + h) - u(x_0, t_0) &\geq v(x(t_0 + h), t_0 + h) - v(x_0, t_0) \\ &= \int_{t_0}^{t_0+h} \frac{d}{ds} v(x(s), s) ds \\ &= \int_{t_0}^{t_0+h} v_t(x(s), s) + f(x(s), \alpha(s)) \cdot D_x v(x(s), s) ds, \end{aligned} \tag{3.83}$$

According to the optimality conditions from formula (3.56)

$$u(x_0, t_0) \geq \int_{t_0}^{t_0+h} h(x(s), \alpha(s)) ds + u(x(t_0 + h), t_0 + h) - \frac{\delta h}{2}. \tag{3.84}$$

Combining the above inequalities, we get the contradition

$$\begin{aligned} \frac{\delta h}{2} \geq \int_{t_0}^{t_0+h} & v_t(x(s), s) + f(x(s), \alpha(s)) \cdot D_x v(x(s), s) \\ & + h(x(s), \alpha(s)) ds \geq \delta h. \end{aligned} \tag{3.85}$$

This contradition proves that u is a viscosity supersolution, and hence also a solution of the given problem, which ends the proof. □

4 Numerical Approach

In the previous chapter we talked about the method of vanishing viscosity for approximating the solutions to the Hamilton-Jacobi equation, but it's not the only way to find the approximate solution. There're also a lot of numerical methods for solving the general version of the equation, as well as its particular versions, like eikonal equation. In this final chapter we'll see how some of the numerical methods agree with the theory presented in the previous chapters. As this chapter has an illustrative mission, most of the proofs will be skipped or referred to the sources.

4.1 First Order Monotone Schemes

For simplicity of the discussion and illustration, we'll consider the HJ equation in one dimension on a bounded domain $\Omega \in \mathbb{R}$, with convex Hamiltonian $H = H(p)$ and having a periodic boundary condition

$$u_t(x,t) + H(u_x(x,t)) = 0, \ x \in \Omega. \tag{4.1}$$

Next we cover the domain Ω by a uniform mesh $x_i = ih$, $i = 1..n$. We denote the approximations of $u(x,t)$ in x_i by $u_i := u_i(t) := u(x_i, t)$. The first problem, that we are facing right now is how to approximate the non-linear Hamiltonian. As we have build the mesh in the spacial variable, we can try approximating the derivative of u by a forward and a backward numerical differences:

$$u_i'^{+} = \frac{u_{i+1} - u_i}{h}, \ u_i'^{-} = \frac{u_i - u_{i-1}}{h}, \tag{4.2}$$

and then consider the *numerical Hamiltonian* $\hat{H} := \hat{H}(p^-, p^+)$, that in some way approximates the original H, and is consistent with it, i.e.

$$\hat{H}(p,p) = H(p).$$

So at the end we get the numerical difference scheme of the following form

$$\frac{d}{dt}u_i(t) + \hat{H}(u_i'^{-}, u_i'^{+}) = 0 \tag{4.3}$$

For now this scheme is semi-discrete, it's still continuous in time, but we'll deal with time later. Now we also require the numerical Hamiltonian to have the following monotonicity conditions: $\hat{H}(p^-, p^+)$ is *non-decreasing* in p^- and is *non-increasing* in p^+. If $\hat{H}$ has these properties, then it's called a *monotone numerical Hamiltonian* and the scheme (4.3) is then called a *monotone scheme.* Now it appears, that independently of the exact choice of the numerical Hamiltonian, the monotone schemes have the following properties (proofs can be found in [7]):

1. The monotone schemes are stable and convergent to the viscosity solution of the initial value problem.

2. The error between the numerical solution and the viscosity solution is at most $O(h^{\frac{1}{2}})$ in L^∞-norm.

In general, the monotone schemes are the simplest numerical methods for solving the Hamilton-Jacobi equations, and although they have the nice properties above, they still fail to achieve the desired order of accuracy for smooth solutions, as it can only be $O(h)$.

Remark. *There might be two generalizations of the scheme above.*

1. *If we consider the equation in many dimensions ($\mathbb{R}^d$), with the corresponding uniform mesh $x_j^{(i)} = jh^{(i)}$ for $i = 1..d$, $j = 1..n$, then the numerical Hamiltonian will be the a function of $2d$ variables:*

$$\hat{H} := \hat{H}(p^-, p^+) = \hat{H}(p_1^-, ..., p_d^-, p_1^+, ..., p_d^+),$$

   and the difference scheme would look like

$$\frac{d}{dt}u_\alpha(t) + \hat{H}(u_\alpha^{i_1,-}, ..., u_\alpha^{i_d,-}, u_\alpha^{i_1,+}, ..., u_\alpha^{i_d,+}) = 0,$$

   where $\alpha = (i_1, ..., i_d)$ is a multiindex, $u_\alpha = u(x_{i_1}, ...x_{i_d})$, $u_\alpha^{j,-}$ and $u_\alpha^{j,+}$ and backward and forward approximation of derivatives in $j-th$ variable, calculated at a point x_α.

2. *Another important generalization is looking at the mesh, which is not uniform. In this case we have a mesh $(x_0, ..., x_n)$, and denote $\Delta^+ x_i = x_{i+1} - x_i$ and $\Delta^- x_i = x_i - x_{i-1}$. Then the difference will be in calculation of the backward and forward numerical derivatives, which in this case will look like:*

$$u_i'^+ = \frac{u_{i+1} - u_i}{\Delta^+ x}, \; u_i'^- = \frac{u_i - u_{i-1}}{\Delta^- x},$$

Next, we'll show a couple of examples of numerical Hamiltonians. We start with the Lax-Friedrichs Hamiltonian, which was originally proposed by Lax in [22] in connection with non-linear parabolic equations.

Definition 4.1. *The* Lax-Friedrichs Hamiltonian *is defined by*

$$\hat{H}^{LF_1}(p^-, p^+) = H\left(\frac{p^- + p^+}{2}\right) - \alpha\frac{p^+ - p^-}{2}, \tag{4.4}$$

where α is the constant (artificial viscosity), that has to satisfy the following condition:

$$\alpha \geq \max |H'(p)|.$$

The other version can be given as :

$$\hat{H}^{LF_2}(p^-, p^+) = \frac{1}{2}\left(H(p^-) + H(p^+) - \alpha\frac{p^+ - p^-}{2}\right), \tag{4.5}$$

Remark. *One can see, that this numerical Hamiltonian is indeed consistent with the original H, as*

$$\hat{H}^{LF_1}(p, p) = H\left(\frac{p + p}{2}\right) - \alpha\frac{p - p}{2} = H(p),$$

and it's also monotone due to the convexity of H.

Definition 4.2. *The* Godunov Hamiltonian *(or Godunov monotone flux) is defined by*

$$\hat{H}^{G}(p^-, p^+) = \begin{cases} \min_{p^- \leq p \leq p^+} H(p), & \text{if } p^- \leq p^+ \\ \max_{p^+ \leq p \leq p^-} H(p), & \text{if } p^- > p^+ \end{cases} \tag{4.6}$$

Remark. *This scheme is more difficult to program due to the max and min involved, however, in case of convex Hamiltonian this scheme turns out to be quite efficient.*

Definition 4.3. *Denote $I(a, b) = [min(a, b), max(a, b)]$, then the* Roe Hamiltonian *is defined by*

$$\hat{H}^{R}(p^-, p^+) = \begin{cases} H(p^-), & \text{if } H'(p) \geq 0 \text{ for all } p \in I(p^-, p^+) \\ H(p^+), & \text{if } H'(p) \leq 0 \text{ for all } p \in I(p^-, p^+) \\ H\left(\frac{p^- + p^+}{2}\right) - \alpha\frac{p^+ - p^-}{2}, & \text{if } H'(p) \text{changes sign in } I(p^-, p^+) \end{cases} \tag{4.7}$$

where α is defined locally:

$$\alpha \geq \max_{p \in I(p^-, p^+)} H'(p)|.$$

We can say, that this Hamiltonian has a local Lax-Friedrichs fix.

Next we'll work out a simple example for the eikonal equation to illustrate, that the scheme converges to the viscosity solution.

Example 4.1. *Remember the problem for the eikonal equation*

$$\begin{cases} |u'(x)| = 1 \ on \ (-1,1) \\ u(-1) = u(1) = 0. \end{cases}$$

We know that the unique viscosity solution to this problem is $u(x) = 1 - |x|$. Now we want to apply the first order monotone scheme to see what results it will yield. We'll use the Lax-Friedrichs Hamiltonian. In our case $H(p) = |p| - 1$.The artificial viscosity α can immediately be taken as $\alpha = \max |H'(p)| = 1$, hence the numerical Hamiltonian in our case looks like:

$$\hat{H}(p^-, p^+) = \left| \frac{p^- + p+}{2} \right| - \frac{p^+ - p^-}{2}$$

The mesh we'll work with is going to have $h = 0.5$, so we have $x_0 = -1, x_1 = -0.5, x_2 = 0, x_3 = 0.5, x_4 = 1$. We have due to the boundary conditions: $u_0 = u_4 = 0$. Next we calculate backward differences for u:

$$\begin{aligned} u_1'^+ &= \frac{u_2 - u_1}{0.5} = 2(u_2 - u_1), & u_1'^- &= \frac{u_1 - u_0}{0.5} = 2u_1, \\ u_2'^+ &= \frac{u_3 - u_2}{0.5} = 2(u_3 - u_2), & u_2'^- &= \frac{u_2 - u_1}{0.5} = 2(u_2 - u_1) \\ u_3'^+ &= \frac{u_4 - u_3}{0.5} = -2u_3, & u_3'^- &= \frac{u_3 - u_2}{0.5} = 2(u_3 - u_2) \end{aligned}$$

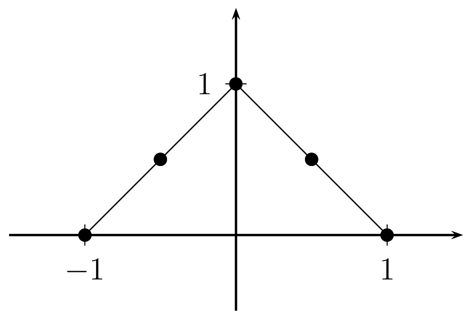
Next we calculate the Hamiltonian for different arguments and build the system of equations:

$$\begin{aligned} H(u_1'^+, u_1'^-) &= |u_2| + 2u_1 - u_2 = 1 \\ H(u_2'^+, u_2'^-) &= |u_3 - u_1| + 2u_2 - u_1 - u_3 = 1 \\ H(u_3'^+, u_3'^-) &= |u_2| + 2u_3 - u_2 = 1 \end{aligned}$$

From the first and the third equation, we can see that $u_1 = u_3$. Then we get:

$$\begin{aligned} |u_2| + 2u_1 - u_2 &= 1 \\ 2u_2 - 2u_1 &= 1 \end{aligned}$$

It follows, after summing up the equations, that $|u_2| + u_2 = 2$. So $u_2 = 1$, and hence $u_1 = u_3 = \frac{1}{2}$, which are exactly the values of the viscosity solution at the corresponding points of the mesh! So even at such a big mesh, the method gives the numerical solution, which is very close to the unique viscosity solution. The figure below illustrates this



4.2 Time Discretization

Remember, that the first order scheme (4.3) is still semi-discrete, the time variable is continuous. There're multiple ways to deal with it numerically. The simples way would be to use the first-order Euler method. We can rewrite the scheme (4.3) as:

$$\frac{du}{t} + \tilde{H}(u) = 0, \tag{4.8}$$

where u is a vector function $(u_1(t), ..., u_{n-1}(t))$, and

$$\tilde{H}(u) = (\hat{H}(u_1'^-(t), u_1'^+(t)), ..., \hat{H}(u_{n-1}'^-(t), u_{n-1}'^+(t)))$$

is the full discretization of the spatial operator H. We have the initial conditions, that can be rewritten in this form as

$$u(0) = ((u_1(0), ..., u_{n-1}(0)) = (g(x_1), ..., g(x_{n-1}) =: (g_1, ..., g_{n-1}).$$

We choose the time step Δt and denote $u^{(j)} = u(j\Delta t)$. The idea of the Euler's method is simply using the right time derivative approximation, meaning, that the scheme (4.8) can be approximated as:

$$\frac{u^{(j+1)} - u^{(j)}}{\Delta t} + \tilde{H}(u^{(j)}) = 0,$$

which is called *a forward Euler method*, or as

$$\frac{u^{(j+1)} - u^{(j)}}{\Delta t} + \tilde{H}(u^{(j+1)}) = 0,$$

which is called *a backward Euler method*.
The local error of such method is proportional to Δt^2 and although Euler method is very simple and easy to program, it's generally effective only for small time step. It also serves as a basis for further methods. We'll discuss some of these improvements. The simplest one is called *midpoint method*. First we rewrite the Euler scheme as

$$u^{(j+1)} = u^{(j)} - \Delta t \tilde{H}(u^{(j+1)}).$$

The idea of the midpoint method is to insert additional evaluation for the argument of $\tilde{H}$. Instead of $u^{(j+1)}$ we'll use the mid-point approximation, that comes from the forward Euler method:

$$u^{(j+\frac{1}{2})} = u^{(j)} - \frac{\Delta t}{2} \tilde{H}(u^{(j+1)}).$$

So the mid-point scheme will look like:

$$u^{(j+1)} = u^{(j)} - \Delta t \tilde{H}(u^{(j)} - \frac{\Delta t}{2} \tilde{H}(u^{(j+1)})).$$

We can continue such an argument further and that will lead to the family of explicit Runge-Kutta methods, which in general look like this for the semidiscrete scheme (4.8):

$$\begin{aligned}
k_1 &= -\tilde{H}(u^{(j)}) \\
k_2 &= -\tilde{H}(u^{(j)} + \Delta t a_{21} k_1) \\
&\dots \\
k_s &= -\tilde{H}(u^{(j)} + \Delta t(a_{s1} k_1 + a_{s2} k_2 + \dots + a_{s,s-1} k_{s-1})) \\
u^{(j+1)} &= u^{(j)} + \Delta t(b_1 k_1 + b_2 k_2 + \dots + b_s k_s).
\end{aligned}$$

The parameters a_{sj} and b_s are chosen so that the certain local truncation error requirement is met. In general, one should have $s > p$, if the desired error should be $O(\Delta t^{p+1})$. Further information about Runge-Kutta methods can be found in [4].

4.3 Higher Order Schemes and Further Reading

The monotone first order schemes, introduced in the first subsection, can be seen as the building blocks for the higher order schemes. Generally, high order schemes are expected to have high order accuracy only in smooth regions away from the singularities of the derivative of solutions.
Let us rewrite the previous first order scheme (4.3) in terms of interpolating polynomials. The steps would be the following:

1. Choose two mesh regions $S_2^- = \{x_{i-1}, x_i\}$, $S_2^+ = \{x_i, x_{i+1}\}$ that contain two grid points and will be used to approximate the left and the right derivatives.

2. Next we find two interpolating polynomials $p_-(x)$ and $p_+(x)$, such that $p_-(x_{i-1}) = u_{i-1}, p_-(x_i) = u_i$ and $p_+(x_i) = u_i, p_+(x_{i+i}) = u_{i+1}$. The polynomials would have the following form (we don't calculate constants b_1 and b_2, as they are not relevant for the discussion):

$$p_-(x) = \frac{u_i - u_{i-1}}{x_i - x_{i-1}} x + b_1 = \frac{u_i - u_{i-1}}{h} x + b_1$$
$$p_+(x) = \frac{u_{i+1} - u_i}{x_{i+1} - x_i} x + b_2 = \frac{u_{i+1} - u_i}{h} x + b_2.$$

3. Finally we take $u_i'^- = p_-'(x_i)$ and $u_i'^+ = p_+'(x_i)$ as approximations for the left and right derivatives of u and use them in the scheme difference scheme and we get exactly (4.3).

Now the idea of the higher order schemes is to use higher order approximating polynomials. Let's work out the idea in terms of using mesh regions with four points.

1. Choose two following mesh regions for approximating polynomial:

$$S_4^- = \{x_{i-2}, x_{i-1}, x_i, x_{i+1}\}$$
$$S_4^+ = \{x_{i-1}, x_i, x_{i+1}, x_{i+2}\}.$$

2. The conditions for calculating the coefficients of the two interpolating polynomials $p_1(x)$ and $p_2(x)$ are:

$$\begin{aligned} p_-(x_{i-2}) &= u_{i-2} & p_+(x_{i-1}) &= u_{i-1} \\ p_-(x_{i-1}) &= u_{i-1} & p_+(x_i) &= u_i \\ p_-(x_i) &= u_i & p_+(x_{i+1}) &= u_{i+1} \\ p_-(x_{i+1}) &= u_{i+1} & p_+(x_{i+2}) &= u_{i+2} \end{aligned}$$

The polynomials will be of degree three. We'll need $p_1'(x_i)$ and $p_2'(x_i)$ for calculating the derivatives. We calculate it for p_1 by the Lagrange method (description of the method can be found here [18]). The Lagrange polynomial in our case will look like

$$
\begin{aligned}
p_-(x) &= \frac{(x-x_{i-1})(x-x_i)(x-x_{i+1})}{(x_{i-2}-x_{i-1})(x_{i-2}-x_i)(x_{i-2}-x_{i+1})}u_{i-2} \\
&+ \frac{(x-x_{i-2})(x-x_i)(x-x_{i+1})}{(x_{i-1}-x_{i-2})(x_{i-1}-x_i)(x_{i-1}-x_{i+1})}u_{i-1} \\
&+ \frac{(x-x_{i-2})(x-x_{i-1})(x-x_{i+1})}{(x_i-x_{i-2})(x_i-x_{i-1})(x_i-x_{i+1})}u_i \\
&+ \frac{(x-x_{i-2})(x-x_{i-1})(x-x_i)}{(x_{i+1}-x_{i-2})(x_{i+1}-x_{i-1})(x_{i+1}-x_i)}u_{i+1} \\
&= -\frac{(x-x_{i-1})(x-x_i)(x-x_{i+1})}{6h^3}u_{i-2} \\
&+ \frac{(x-x_{i-2})(x-x_i)(x-x_{i+1})}{2h^3}u_{i-1} \\
&- \frac{(x-x_{i-2})(x-x_{i-1})(x-x_{i+1})}{2h^3}u_i \\
&+ \frac{(x-x_{i-2})(x-x_{i-1})(x-x_i)}{6h^3}u_{i+1}
\end{aligned}
$$

In a similar manner we can write down p_+.

3. Finally by calculating the derivative of this polynomial in x_i we obtain:

$$
\begin{aligned}
u_i'^- = p_-'(x_i) &= \frac{1}{h}\left(\frac{1}{6}u_{i-2} - u_{i-1} + \frac{1}{2}u_i + \frac{1}{3}u_{i+1}\right) \\
u_i'^+ = p_+'(x_i) &= \frac{1}{h}\left(\frac{1}{6}u_{i-1} - u_i + \frac{1}{2}u_{i+1} + \frac{1}{3}u_{i+2}\right),
\end{aligned}
$$

that we can use as the approximations for the scheme.

In this manner one can design finite difference schemes of any order. These schemes are extremely useful in case of smooth solutions.
Further information about the higher order schemes can be found in [23]. Possible improvements and adjustments to the Lax-Friedrichs method, applied to the Hamilton-Jacobi equation is in [19] and [3]. Extensive theoretical background on monotone schemes for Hamilton-Jacobi equation with the proof of its fine properties can be found in [7].

References

[1] Arnold V., *Mathematical Methods Of Classical Mechanics*, Springer, 1986.

[2] Bellman, R. E., *Dynamic Programming and a new formalism in the calculus of variations*, Proc. Natl. Acad. Sci. 40 (4), (1954), 231?235.

[3] Breuss, M., *The Correct Use of the Lax-Friedrichs Method*, ESAIM: Mathematical Modelling and Numerical Analysis, 38.3 (2010), 519-540.

[4] Butcher, John C., *Numerical Methods for Ordinary Differential Equations*, New York: John Wiley And Sons, 2008.

[5] Cannarsa, P., Sinestrari, C., *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, Birkhäuser, Boston, 2004.

[6] Crandall M.G., Evans L.C., Lions P.-L., *Some Properties of Viscosity Solutions of Hamilton-Jacobi Equations*, Trans. Amer. Math. Soc., 282 (1984) 487-502.

[7] Crandall M.G., Lions P.-L., *Two Approximations of Solutions of Hamilton-Jacobi Qquations*, Math. Comp., 43 (1984), n. 167, 1-19.

[8] Crandall M.G., Ishii H., Lions P.-L., *User's Guide to Viscosity Solutions of Second Order Partial Differential Equations*, Bull. Amer. Math. Soc., 27 (1992) 1-67.

[9] Crandall M.G., Lions P.-L., *Viscosity Solutions of Hamilton-Jacobi Equations*, Trans. Amer. Math. Soc., 277 (1983) 1-42.

[10] Evans L.C., *On Solving Certain Nonlinear Partial Differential Equations by Accretive Operator Methods*, Israel Journal of Mathematics, 36/3 (1980), 225-247.

[11] Evans L.C., *Partial Differential Equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 1998.

[12] Field J.H., *Derivation of the Schrödinger equation from the Hamilton-Jacobi equation in Feynman's path integral formulation of quantum mechanics*, Eur. J. Phys. 32 (2011) 63-86.

[13] Friedman A., *The Cauchy problem for first order partial differential equations*, Indiana Univ. Math. J. 23, 27-40 (1973).

[14] Goldstein H., *Classical Mechanics*, Addison Wesley, 2002.

[15] Hamilton, W. R., *On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function*, Philosophical Transactions of the Royal Society, p. 247-308, 1834.

[16] Hamilton, W. R., *Second essay on a general method in dynamics*, Philosophical Transactions of the Royal Society, p. 95-144, 1835.

[17] Hopf E., *Generalized solutions of nonlinear equations of first order*, J. Math. Mech. 14 (1965) 951-973.

[18] Jeffreys, H. and Jeffreys, B. S., *Lagrange's Interpolation Formula*, §9.011 in Methods of Mathematical Physics, 3rd ed. Cambridge, England: Cambridge University Press, p. 260, 1988.

[19] Kao, C., Osher, S., Qian, J., *Lax-Friedrichs Sweeping Scheme for Static Hamilton-Jacobi Equations*, UCLA CAM report (2003).

[20] Landau L. D., Lifshitz E. M., *Mechanics*, Pergamon Press, Oxford, 1969.

[21] Lax P.D., *Hyperbolic systems of conservation laws II*, Commun. Pure Appl. Math., 10 (1957) 537-566.

[22] Lax P.D., *Weak Solutions of Nonlinear Hyperbolic Equations and Their Numerical Approximation*, Comm. Pure Appl. Math., 7 (1954), 159-193.

[23] Osher, S., Shu, C.-W., *High-order Essentially Nonoscillatory Schemes for Hamilton-Jacobi Equations*, SIAM Journal on Numerical Analysis, 28 (1991), 907-922.

[24] Torby B., *Energy Methods*, Advanced Dynamics for Engineers, HRW Series in Mechanical Engineering, USA: CBS College Publishing, 1984.