# ABSTRACT

DUTTA, PRERONA. Metric Entropy and Nonlinear Partial Differential Equations. (Under the direction of Tien Khai Nguyen.)

The metric entropy (or $\varepsilon$-entropy) has been studied extensively in a variety of literature and disciplines. This notion was introduced by Kolmogorov and Tikhomirov in 1959 as the minimum number of bits needed to represent a point in a given subset $K$ of a metric space $(E,\rho)$, up to an accuracy $\varepsilon$ with respect to the metric $\rho$. Recently, the $\varepsilon$-entropy has also been used to measure the set of solutions of various nonlinear partial differential equations (PDEs). In this context, it provides a measure of the order of "resolution" and the "complexity" of a numerical scheme.

This dissertation is motivated by the above points of view and demonstrates techniques emerging from the field of nonlinear analysis, that are used to estimate the metric entropy for different classes of bounded variation (BV) functions. Since Helly's theorem states that a set of uniformly bounded total variation functions is compact in $\mathbf{L}^1$-space, a natural question arises on how to quantify the degree of compactness and we answer this using the concept of $\varepsilon$-entropy. Subsequently, we apply the obtained results to measure solution sets of conservation laws and Hamilton-Jacobi equations.

In the first half of this thesis, we elucidate the foundational principles involved in our work and show that the minimal number of functions needed to represent a bounded total variation function in $\mathbf{L}^1([0,L]^d,\mathbb{R})$ up to an error $\varepsilon$ with respect to $\mathbf{L}^1$-distance, is of the order $\varepsilon^{-d}$. We use this outcome to examine the metric entropy for sets of viscosity solutions to the Hamilton-Jacobi equation. Earlier works on this topic considered a uniformly convex Hamiltonian. We extend the analysis to the case when the Hamiltonian $H \in \mathscr{C}^1(\mathbb{R}^d)$ is strictly convex, coercive and a uniformly directionally convex function. Under these assumptions, we establish sharp estimates on the $\varepsilon$-entropy for sets of viscosity solutions to the Hamilton-Jacobi equation with respect to $\mathbf{W^{1,1}}$-distance in multi-dimensional cases.

The second half of the thesis focuses on finding sharp estimates for the metric entropy of a class of bounded total generalized variation functions taking values in a general totally bounded metric space $(E,\rho)$ upto an accuracy of $\varepsilon$ with respect to the $\mathbf{L}^1$–distance. We rely on the ideas of covering and packing in $(E,\rho)$ to derive such bounds and utilize them to study a scalar conservation law in one-dimensional space, which yields an upper bound on the $\varepsilon$-entropy of a set of entropy admissible weak solutions to it, in case of weakly genuinely nonlinear fluxes, i.e., for fluxes with no affine parts. In particular, for fluxes admitting finitely many inflection points with a polynomial degeneracy, this estimate is sharp.

Metric Entropy and Nonlinear Partial Differential Equations

by
Prerona Dutta

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Mathematics

Raleigh, North Carolina

2021

APPROVED BY:

| | |
|---|---|
| Patrick Combettes | Kazufumi Ito |
| Ryan Murray | Tien Khai Nguyen<br>Chair of Advisory Committee |

# DEDICATION

To my parents and all my teachers in life,
for their valuable lessons and eternal inspiration.

# BIOGRAPHY

Prerona Dutta was born in Kolkata, India and spent her growing years in the City of Joy. She completed her schooling from The Future Foundation School, Kolkata in 2011 and then pursued undergraduate education at St. Xavier's College (Autonomous), Kolkata from where she received the degree of B.Sc. Honors in Mathematics in 2014. Subsequently, she moved to the peaceful, coastal city of Puducherry in India where she studied at Pondicherry University and was awarded the degree of M.Sc. in Mathematics in 2016. Thereafter Prerona travelled halfway around the globe in Fall 2017 to join North Carolina State University in the charming, green city of Raleigh, USA and commenced Doctoral research in Mathematics with Dr. Tien Khai Nguyen as her advisor. She will continue her mathematical career through postdoctoral research at The Ohio State University where she will join as Ross Assistant Professor from Fall 2021. Apart from her academic pursuits, Prerona is an avid reader who seeks solace in the lap of Nature and finds pleasure in singing, playing badminton, photography and writing poetry.

## ACKNOWLEDGEMENTS

The path towards a doctoral degree has been a positively life-altering experience and I would like to thank all those who made it possible for me to undertake this journey.

I wish to begin by expressing my deepest gratitude to my advisor Prof. Tien Khai Nguyen for the guidance and care he has bestowed throughout. His detailed feedback and utmost generosity in all possible aspects made this learning curve most wholesome and fulfilling. He has always inspired me to delve deeper into a problem till there is a clear picture of it in my mind. This has helped me to grow more insightful than I was a few years ago. He has also taught me how to approach various problems, helped me to improve upon academic writing and encouraged me to listen to other mathematicians as a motivation for finding new problems and directions of research. With his helpful supervision, gradually I was enlightened further on research methodologies and gained in confidence towards becoming an independent researcher. I will be eternally grateful to my advisor for being most supportive and reassuring all along the way.

I am greatly indebted to all the members of my Graduate Advisory Committee, namely, Prof. Kazufumi Ito, Prof. Patrick Combettes, Prof. Ryan Murray and Prof. Steven Hunter, as well as my references who have provided me with recommendation letters, namely, Prof. Antonio Marigonda, Ms. Elizabeth Dempster, Prof. I. Subramania Pillai, Prof. Rajeswari Seshadri and Prof. S. Francis Raj. Additionally, I would like to thank Ms. Dempster for her feedback on my teaching, which is truly beneficial and much appreciated.

My co-authors have played significant roles in my mathematical development. I would especially like to thank Dr. Rossana Capuani as I have been helped greatly through our enriching discussions, both mathematical and light-hearted, when working together on bounded total generalized variation functions. I am also grateful to Prof. Stefano Bianchini for his valuable collaboration on the problem of finding metric entropy estimates for the Hamilton-Jacobi equation with a uniformly directionally convex Hamiltonian.

I express my sincere gratitude to the Department of Mathematics at the North Carolina State University for providing me with the opportunity to conduct my doctoral studies here and ensuring a smooth sailing throughout. I am also thankful to all the faculty and students of the Nonlinear Analysis Thematic Group at the University for creating and preserving an intellectually stimulating atmosphere that is conducive for research.

Every quest is driven by a dream that has been well-cherished from its inception and gradually built upon with the hope of attainment. In my case, the aspiration is mine but it has been fostered by the efforts of many, namely, my parents and teachers. I am forever indebted to my parents Mr. Bhanu Dutta and Dr. Krishna Dutta for their love, teachings and constant support in all my endeavours. Their unquestioning faith in my abilities and understanding of my desires have always empowered me with the freedom of choice and happen to be the most effective morale booster.

I would sincerely like to thank all my teachers from The Future Foundation School, St. Xavier's College (Autonomous) Kolkata and Pondicherry University, for having helped to cultivate the traits that have kept me going since the time in the seventh standard when I decided that I want to pursue higher studies in Mathematics, towards being able to realize that goal in actuality. In addition, I would like to thank the Mathematics Training and Talent Search (MTTS) Programme in India, for helping me develop a better understanding of basic undergraduate mathematical concepts and reinstating the confidence I had begun to lose in the time prior to attending their camp.

I have been most fortunate when it comes to the friends I have had in the course of my ventures and I thank each and every one of them. I am immensely grateful to Jithin Joseph for being my pillar of strength throughout this journey. I would also like to thank Anushka Khasnobish, Pratik Misra, Purnima Kovuri and Suparna Chakraborty for cheering me on unfailingly, from different parts of the world.

In conclusion, I would like to pay special regards to the city of Raleigh which has housed me for the past few years with all its lakes and greenery, making me feel at home, miles away from what used to be home. Finally, thank you to everyone who has loved, taught, inspired and supported me in my entire lifetime either in person or in spirit, I am grateful to you all.

# TABLE OF CONTENTS

# LIST OF FIGURES

CHAPTER

# 1

# INTRODUCTION

The metric entropy (or $\varepsilon$-entropy) has been studied extensively in a variety of literature and disciplines. It plays a central role in various areas of information theory and statistics, including nonparametric function estimation, density information, empirical processes and machine learning ([14, 29, 45]). This concept was first introduced by Kolmogorov and Tikhomirov in [33] and defined as follows.

**Definition.** *Let $(E,\rho)$ be a metric space and $K$ be a totally bounded subset of $E$. For $\varepsilon > 0$, let $\mathcal{N}_\varepsilon(K\big|E)$ be the minimal number of sets in an $\varepsilon$-covering of $K$, i.e., a covering of $K$ by balls in $E$ with radius no greater than $\varepsilon$. Then the $\varepsilon$-entropy of $K$ is defined as*

$$\mathcal{H}_\varepsilon(K|E) = \log_2 \mathcal{N}_\varepsilon(K|E).$$

A classical topic in the field of probability is to investigate the metric covering numbers for general classes of real-valued functions $\mathscr{F}$ defined on a space $E$ under the family of $\mathbf{L}^1(dP)$ where $P$ is a probability distribution on $E$. Upper bounds in terms of the Vapnik-Chervonenkis dimension and pseudo-dimension of the function class were established in [25] and then improved in [29, 30, 45]. Several results on lower bounds were also studied in [35]. Eventually, upper and lower estimates of the $\varepsilon$-entropy of $\mathscr{F}$ in $\mathbf{L}^1(dP)$ in terms of a scale-sensitive dimension of the function class were provided in [35, 40] and applied to machine learning.

According to Helly's theorem, a set of uniformly bounded variation functions is compact in $\mathbf{L}^1$-space. This raises the logical question of how to quantify the degree of compactness of such sets and attempts were made to do so by using the $\varepsilon$-entropy. In [35], the authors showed that the $\varepsilon$-entropy of any set of uniformly bounded total variation real-valued functions in $\mathbf{L}^1$ is of the order $\frac{1}{\varepsilon}$ in the scalar case. Related works have been done in the context of density estimation where attention has been given to the problem of finding covering numbers for the classes of densities that are unimodal or non-decreasing in [14, 27]. In multi-dimensional cases, the covering numbers of convex and uniformly bounded functions were studied in [28]. It was shown that the $\varepsilon$-entropy of a class of convex functions with uniform bound in $\mathbf{L}^1$ is of the order $\frac{1}{\varepsilon^{\frac{d}{2}}}$ where $d$ is the dimension of the state variable. The result was previously studied for scalar state variables in [24] and for convex functions that are uniformly bounded and uniformly Lipschitz with a known Lipschitz constant in [17]. These results have direct implications in the study of rates of convergence of empirical minimization procedures (see in [15, 47]) as well as optimal convergence rates in convexity constrained function estimation problems (see in [13, 39, 48]).

Recently, the $\varepsilon$-entropy has been used to measure the set of solutions of certain nonlinear partial differential equations. In this setting, it could provide a measure of the order of "resolution" and "complexity" of a numerical scheme, as suggested in [37, 38]. Roughly speaking, the order of magnitude of the $\varepsilon$-entropy indicates the minimum number of operations that one should perform in order to obtain an approximate solution with a precision of order $\varepsilon$ with respect to the considered topology. A starting point for the research on this topic is a result which was obtained in [23] for a scalar conservation law in one dimensional space

$$u_t(t,x)+f(u(t,x))_x = 0, \tag{1.0.1}$$

with uniformly convex flux $f$. It was shown that the upper bound of the minimum number of functions needed to represent an entropy solution $u$ of (1.0.1) at any time $t > 0$ up to an accuracy of $\varepsilon$ with respect to $\mathbf{L}^1$-distance is of the order $\frac{1}{\varepsilon}$. In [5] a lower bound on such an $\varepsilon$-entropy was established, which is of the same order as the upper bound in [23]. More generally, the authors in [5] also obtained the same estimate for a system of hyperbolic conservation laws in [6, 7]. In the scalar case, it is well-known that the integral form of an entropy solution of (1.0.1) is a viscosity solution of the related Hamilton-Jacobi equation. Therefore, it is natural to study the $\varepsilon$-entropy for the set of viscosity solutions to the Hamilton-Jacobi equation

$$u_t(t,x)+H\big(\nabla_x u(t,x)\big) = 0, \tag{1.0.2}$$

with respect to $\mathbf{W^{1,1}}$-distance in multi-dimensional cases. Most recently, it has been proved in [3] that the minimal number of functions needed to represent a viscosity solution of (1.0.2) up to an accuracy of $\varepsilon$ with respect to the $\mathbf{W^{1,1}}$-distance is of the order $\frac{1}{\varepsilon^d}$, provided that the Hamiltonian $H$ is uniformly convex. Here, $d$ is the dimension of the state variable. The same result for when $H$ depends on the state variable $x$ has also been obtained by the same authors in [4]. Interestingly, the authors in [3] also established an upper bound on the $\varepsilon$-entropy for the class of monotone functions in $\mathbf{L^1}$-space. As a consequence of Poincaré-type inequalities, it was possible to obtain the $\varepsilon$-entropy for a class of semi-convex/concave functions in Sobolev $\mathbf{W^{1,1}}$-space. This result extended the ones in [17, 24, 28] to a stronger norm, namely the $\mathbf{W^{1,1}}$-norm, instead of $\mathbf{L}^1$-norm.

Motivated by the results in [3, 17, 24, 28, 35] and considering a possible application to Hamilton-Jacobi equations with non-strictly convex Hamiltonian, in Section 3.1 we provide upper and lower estimates of the $\varepsilon$-entropy for a class of uniformly bounded total variation functions in $\mathbf{L^1}$-space in multi-dimensional cases. This is a joint work with Khai T. Nguyen and has been published in *Journal of Mathematical Analysis and Applications* ([26]).

Looking back at [4], we observe that the main idea here involved providing controllability results for Hamilton-Jacobi equations and a compactness result for a class of semiconcave functions. However, such a gain of BV regularity does not hold for (1.0.2) with general strictly convex Hamiltonian functions and the previous approach to finding the $\varepsilon$-entropy of the solution set as in [3, 4] cannot be applied. In this case, a study of the fine regularity properties of viscosity solutions is still lacking and quantitative estimates on the $\varepsilon$-entropy of viscosity solution sets are not available yet.

Therefore, in Section 3.2, we further the research in this direction by using our results from Section 3.1 to extend the analysis of the metric entropy for sets of viscosity solutions to (1.0.2) when the Hamiltonian $H \in \mathscr{C}^1\big(\mathbb{R}^d\big)$ is strictly convex, coercive and in the form of a uniformly directionally convex function. This work has been done in collaboration with Stefano Bianchini and Khai T. Nguyen ([12]), where we establish a BV bound on the slope of backward characteristics $DH(u(t,\cdot))$ starting at a positive time $t>0$ and relying on this BV bound, we quantify the metric entropy in $\mathbf{W}^{1,1}_{\text{loc}}(\mathbb{R}^d)$ for the map $S_t$ that associates to every given initial data $u_0 \in \mathbf{Lip}\big(\mathbb{R}^d\big)$, the corresponding solution $S_t u_0$.

From our review of the existing literature, we note that previous works on metric entropy for nonlinear partial differential equations which lead to our results in Chapter 3 strongly rely on the BV regularity properties of solutions. Thereafter, the results in [5, 23] have been extended to scalar conservation laws with a smooth flux function $f$ that is either strictly (but not necessarily uniformly) convex or has a single inflection point with a polynomial degeneracy [8] where entropy admissible weak solutions could be having unbounded total variation. In this case, the sharp estimate on the $\varepsilon$-entropy for sets of entropy admissible weak solutions have been provided by exploiting the BV bound of the characteristic speed $f'(u)$ at any positive time [20]. On the other hand, it was shown in [11, Example 7.2] that, in general, for fluxes having one inflection point where all derivatives vanish, the composition of the derivative of the flux with the solution of (1.0.1) fails in general to belong to the BV space and the analysis in [8] is not applicable here. However, for *weakly genuinely nonlinear* fluxes, that is, for fluxes with no affine parts, equibounded sets of entropy solutions to (1.0.1) at positive time are still relatively compact in $\mathbf{L}^1$ ([46, Theorem 26]). Therefore, for fluxes of such classes that do not fulfill the assumptions in [8], it remains an open problem to provide a sharp estimate on the $\varepsilon$-entropy for the solution set of (1.0.1) and a different approach must be pursued when studying (1.0.1) with weakly genuinely nonlinear fluxes, perhaps exploiting the uniform bound on total generalized variation of entropy admissible weak solutions studied in [41, Theorem 1].

The above points of view elicited our study of the $\varepsilon$-entropy for classes of uniformly bounded total generalized variation functions taking values in a general totally bounded metric space $(E,\rho)$. This work illustrated in Chapter 4 has been done jointly with Rossana Capuani and Khai T. Nguyen and published in *SIAM Journal on Mathematical Analysis* ([19]). For deriving sharp estimates explicitly, we use the notions of doubling and packing dimensions of $(E,\rho)$, denoted by $\mathbf{d}(E)$ and $\mathbf{p}(E)$ respectively, which were first introduced by Assouad in [9]. More precisely, we consider $\mathscr{F}^{\Psi}_{[L,V]}$ to be a set of functions $g:[0,L]\to E$ such that the $\Psi$-total variation of $g$ over the interval $[0,L]$ is bounded by $V$ where $\Psi$ is a convex function and $\Psi:[0,+\infty)\to[0,+\infty)$ satisfies $\Psi(0)=0$ and $\Psi(s)>0$ for all $s>0$. In Section 4.1, we prove that for every $\varepsilon>0$ sufficiently small, the sharp bounds on $\mathscr{H}_\varepsilon\Big(\mathscr{F}^{\Psi}_{[L,V]}\Big|\mathbf{L}^1([0,L],E)\Big)$ can be approximated in terms of $\mathbf{p}(E)$, $\mathbf{d}(E)$ and $\Psi$.

We conclude this study on the notion of metric entropy and its application to nonlinear partial differential equations, by applying the above result in Section 4.2 to provide an upper estimate on the $\varepsilon$-entropy of a set of entropy admissible weak solutions to scalar conservation laws (1.0.1) with general weakly genuinely nonlinear fluxes in Theorem 62, which partially extends the recent developments made in [8]. The estimate is sharp in the case of fluxes having finite inflection points with a polynomial degeneracy. Hereafter, a natural question arises about establishing similar sharp estimates for the $\varepsilon$-entropy of such solution sets to (1.0.1) with general weakly genuinely nonlinear fluxes, however this topic remains open to further research.

CHAPTER

# 2

# PRELIMINARIES

## 2.1 Notation

Let $d$ be a given positive integer and $\Omega \subseteq \mathbb{R}^d$ be a measurable set. Also, let $E$ be a metric space with distance $\rho$ and $I$ be an interval in $\mathbb{R}$. Throughout this thesis we adopt the following notation.

- $|\cdot|$, the Euclidean norm in $\mathbb{R}^d$ and for any $R > 0$

$$B_d(x,R) = \{y \in \mathbb{R}^d : |x-y| < R\}, \qquad \square_R = (-R,R)^d;$$

- $\langle \cdot, \cdot \rangle$, the Euclidean inner product in $\mathbb{R}^d$;

- $\mathrm{int}(\Omega)$, the interior of $\Omega$;

- $\partial\Omega$, the boundary of $\Omega$;

- $[x,y]$, the segment joining two points $x, y \in \mathbb{R}^d$;

- $B_\rho(z,r)$, the open ball of radius $r$ and center $z$, with respect to the metric $\rho$ on $E$, i.e.,

$$B_\rho(z,r) = \left\{ y \in E \mid \rho(z,y) < r \right\};$$

- $\mathrm{diam}(F) = \sup_{x,y\in F} \rho(x,y)$, the diameter of the set $F$ in $(E,\rho)$;
- $\mathrm{Vol}(\Omega)$, the Lebesgue measure of a measurable set $\Omega \subset \mathbb{R}^d$;
- $\omega_d := \mathrm{Vol}(B_d(0,1))$, the Lebesgue measure of the unit ball in $\mathbb{R}^d$;
- $\mathscr{C}^1(\Omega)$, the space of continuously differentiable functions on $\Omega$;
- $\mathscr{C}^1_c(\Omega,\mathbb{R}^d)$, with $\Omega \subset \mathbb{R}^d$ an open set, the set of all continuously differentiable functions from $\Omega$ to $\mathbb{R}^d$ with a compact support in $\Omega$;
- $\mathscr{C}^\infty(\mathbb{R},\mathbb{R})$, the space of smooth functions having derivatives of all orders;
- $\mathscr{B}(I,[0,+\infty))$, a set of bounded functions from $I$ to $[0,+\infty)$;
- $\mathbf{L}^1(\Omega)$, the Lebesgue space of all (equivalence classes of) summable real-valued functions on $\Omega$, equipped with the usual norm $\|\cdot\|_{\mathbf{L}^1(\Omega)}$ (the same symbol will be used for the case when $u$ is vector-valued);
- $\mathbf{L}^1(I,E)$, the Lebesgue metric space of all (equivalence classes of) summable functions $f : I \to E$, equipped with the usual $\mathbf{L}^1$-metric distance, i.e.,

  $$\rho_{\mathbf{L}^1}(f,g) := \int_I \rho(f(t),g(t))dt < +\infty$$

  for every $f,g \in \mathbf{L}^1(I,E)$;
- $B_{\mathbf{L}^1(I,E)}(\varphi,r)$, the open ball of radius $r$ and center $\varphi$ in $\mathbf{L}^1(I,E)$, with respect to the metric $\rho_{\mathbf{L}^1}$ on $\mathbf{L}^1(I,E)$, i.e.,

  $$B_{\mathbf{L}^1(I,E)}(\varphi,r) = \left\{g \in \mathbf{L}^1(I,E) \,\middle|\, \rho_{\mathbf{L}^1}(\varphi,g) < r\right\};$$
- $\mathbf{L}^1(\mathbb{R})$, the Lebesgue space of all (equivalence classes of) summable functions on $\mathbb{R}$, equipped with the usual norm $\|\cdot\|_{\mathbf{L}^1}$;
- $\mathbf{L}^\infty(\Omega)$, the space of all essentially bounded real-valued functions on $\Omega$ and $\|u\|_{\mathbf{L}^\infty(\Omega)}$ is the essential supremum of a function $u \in \mathbf{L}^\infty(\Omega)$ (the same symbol will be used for the case when $u$ is vector-valued);
- $\mathbf{L}^\infty(\mathbb{R})$, the space of all essentially bounded functions on $\mathbb{R}$, equipped with the usual norm $\|\cdot\|_{\mathbf{L}^\infty}$;
- $\mathrm{Supp}(u)$, the essential support of a function $u \in \mathbf{L}^\infty(\mathbb{R})$;

- $\mathbf{W}^{1,1}(\Omega)$, the Sobolev space of functions with summable first order distributional derivatives and $\|\cdot\|_{\mathbf{W}^{1,1}(\Omega)}$ is its norm;
- $BV(\Omega,\mathbb{R}^m)$, the space of all vector-valued functions $F:\Omega\to\mathbb{R}^m$ of bounded variation (i.e., all $F\in\mathbf{L}^1(\Omega,\mathbb{R}^m)$ such that the first partial derivatives of $F$ in the sense of distributions are measures with finite total variation in $\Omega$);
- $TV(g,I)$, total variation of $g$ over the interval $I$;
- $TV^{\Psi}(g,I)$, $\Psi$-total variation of $g$ over the interval $I$;
- $TV^{\frac{1}{\gamma}}(g,I)$, $\gamma$-total variation of $g$ over the interval $I$, i.e., $\Psi$-total variation of $g$ with $\Psi$ defined by $\Psi(s)=|s|^{\gamma}$;
- $\mathbf{Lip}(\Omega)$, the space of all Lipschitz functions $f:\Omega\to\mathbb{R}$ and $\mathrm{Lip}[f]$ is the Lipschitz seminorm of $f$;
- $\mathscr{H}^k(E)$, the $k$-dimensional Hausdorff measure of $E\subset\mathbb{R}^d$;
- $\mathscr{L}^d$, the Lebesgue outer measure on $\mathbb{R}^d$;
- For any function $f$, the function $f_{\llcorner\Omega}$ is the restriction of $f$ on $\Omega$;
- $\mathbf{I}_d$, the identity matrix of size $d$;
- $\lfloor a\rfloor := \max\{z\in\mathbb{Z}: z\le a\}$, the integer part $a$;
- $\chi_{\Omega}(x)=\begin{cases} 1 & \text{if} \quad x\in\Omega,\\ 0 & \text{if} \quad x\in\mathbb{R}^d\backslash\Omega\end{cases}$ the characteristic function of a subset $\Omega$ of $\mathbb{R}^d$.
- $\overline{1,N}$, the set of natural numbers from 1 to $N$;
- $\binom{n}{k}=\dfrac{n!}{k!(n-k)!}$, number of ways in which $k$ objects can be chosen from among $n$ objects.

## 2.2 Metric entropy

The notion of metric entropy (or $\varepsilon$-entropy) introduced in [33] is defined as follows.

**Definition 1.** Let $(E,\rho)$ be a metric space and $K$ be a totally bounded subset of $E$. For $\varepsilon>0$, let $\mathcal{N}_{\varepsilon}(K\,|\,E)$ be the minimal number of sets in an $\varepsilon$-covering of $K$, i.e., a covering of $K$ by balls in $E$ with radius no greater than $\varepsilon$.

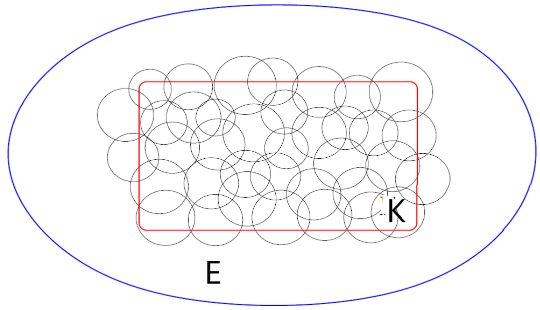


**Figure 2.1** A covering of $K$ in $E$

Then the $\varepsilon$-entropy of $K$ is defined as

$$\mathscr{H}_\varepsilon(K|E) = \log_2 \mathscr{N}_\varepsilon(K|E).$$

In other words, it is the minimum number of bits needed to represent a point in a given set $K$ in the space $E$ with an accuracy of $\varepsilon$ with respect to the metric $\rho$.

Some well-known examples of metric entropy estimates are as follows.

- $E = \mathbb{R}$, $K = [0, L]$

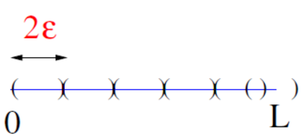


**Figure 2.2** A covering of $[0, L]$ in $\mathbb{R}$

$$\mathscr{N}_\varepsilon([0, L]|\mathbb{R}) \approx \frac{L}{2\varepsilon} \text{ and } \mathscr{H}_\varepsilon([0, L]|\mathbb{R}) \approx -\log_2(\varepsilon)$$

- $E = \mathbb{R}^2$, $K = [0, L] \times [0, L]$

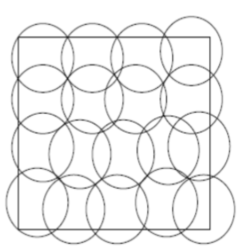

**Figure 2.3** A covering of $[0, L] \times [0, L]$ in $\mathbb{R}^2$

$$\mathscr{N}_\varepsilon([0, L]^2|\mathbb{R}^2) \approx \frac{2L^2}{\varepsilon^2} \text{ and } \mathscr{H}_\varepsilon([0, L]^2|\mathbb{R}^2) \approx -2\log_2(\varepsilon)$$

- $E = \mathbb{R}^d$, $\rho(x, y) = \|x - y\|$ and $K = B_d(0, r)$ for some $r > 0$

$$d \cdot \log_2\left(\frac{r}{\varepsilon}\right) \leq \mathscr{H}_\varepsilon\left(B_d(0, r)\middle|\mathbb{R}^d\right) \leq d \cdot \log_2\left(\frac{2r}{\varepsilon} + 1\right).$$

- **n-dimensional Lipschitz functions**: Let $F_n$ be the set of $L$-Lipschitz functions (with respect to $\|\cdot\|_\infty$) from $[0,1]^n$ to $[0,1]$. Then from [17],

$$\mathcal{H}_\varepsilon(F_n|\mathbf{L}^1([0,1]^n,[0,1])) \approx \left(\frac{L}{\varepsilon}\right)^n.$$

- **1-dimensional BV functions**: Given $L, V > 0$, let $\mathcal{F}_{[L,V]}$ be a set of functions from $[0,L]$ to $[-\frac{V}{2}, \frac{V}{2}]$ with $TV(f,[0,L]) \le V$. Then from [10],

$$\mathcal{H}_\varepsilon(\mathcal{F}_{[L,V]} \mid \mathbf{L}^1([0,L],\mathbb{R})) \le \frac{12LV}{\varepsilon},$$

  for all $\varepsilon \le LV/12$.

## 2.3 Covering and Packing

This section deals with the concepts of covering number and packing number in a totally bounded metric space $(E,\rho)$ and provides useful results to be utilized in Chapter 4. For any $K \subseteq E$ and $\varepsilon > 0$, we say that

- the set $\mathcal{A} = \{a_1, a_2, \ldots, a_n\} \subseteq E$ is an $\varepsilon$-covering of $K$ if $K \subseteq \bigcup_{i=1}^n B_\rho(a_i,\varepsilon)$, or equivalently, for every $x \in K$, there exists $i \in \overline{1,n}$ such that $\rho(x,a_i) < \varepsilon$; Card($\mathcal{A}$) is called the *size* of this $\varepsilon$-covering;
- the set $\mathcal{B} = \{b_1, b_2, \ldots, b_m\} \subseteq K$ is an $\varepsilon$-packing of $K$ if $\rho(b_i,b_j) > \varepsilon$ for all $i \neq j \in \overline{1,m}$, or equivalently, $\{B_\rho(b_i,\varepsilon/2)\}_{i=1}^m$ is a finite set of disjoint balls; Card($\mathcal{B}$) is called the *size* of this $\varepsilon$-packing.

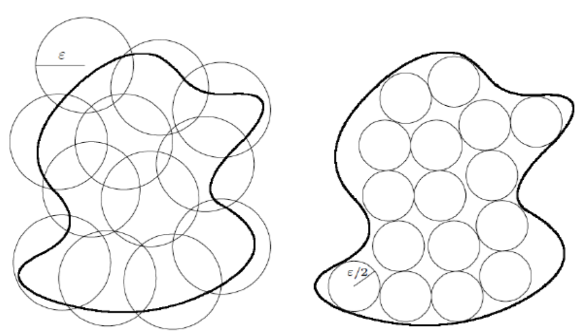


**Figure 2.4** An $\varepsilon$-covering and an $\varepsilon$-packing of a set

**Definition 2.** The $\varepsilon$-covering and $\varepsilon$-packing numbers of $K$ in $(E,\rho)$ are defined by

$$\mathcal{N}_\varepsilon(K|E) = \min\left\{n \in \mathbb{N} \mid \exists\ \varepsilon\text{—covering of } K \text{ having size } n\right\}$$

and

$$\mathcal{M}_\varepsilon(K|E) = \max\left\{m \in \mathbb{N} \mid \exists\ \varepsilon\text{—packing of } K \text{ having size } m\right\},$$

respectively.

Since $E$ is totally bounded, $\mathcal{N}_\varepsilon(K|E)$ is finite for every $\varepsilon > 0$. Moreover, the maps $\varepsilon \mapsto \mathcal{N}_\varepsilon(K|E)$ and $\varepsilon \mapsto \mathcal{M}_\varepsilon(K|E)$ are non-increasing. The relation between $\mathcal{N}_\varepsilon(K|E)$ and $\mathcal{M}_\varepsilon(K|E)$ is described by the following double inequality which was proved in [33].

**Lemma 3.** *For any $\varepsilon > 0$, it holds that*

$$\mathcal{M}_{2\varepsilon}(K|E) \le \mathcal{N}_\varepsilon(K|E) \le \mathcal{M}_\varepsilon(K|E).$$

Let us now introduce a commonly used notion of dimension for a metric space $(E,\rho)$ as proposed in [9, §4].

**Definition 4.** The doubling and packing dimensions of $(E,\rho)$ are respectively defined by

- $\mathbf{d}(E)$ is the minimum natural number $n$ such that for every $x \in E$ and $\varepsilon > 0$, the ball $B_\rho(x, 2\varepsilon)$ can be covered by $2^n$ balls of radius $\varepsilon$;
- $\mathbf{p}(E)$ is the maximum natural number $m$ such that for every $x \in E$ and $\varepsilon > 0$, the ball $B_\rho(x, 2\varepsilon)$ contains an $\varepsilon$-packing of size $\mathcal{M}_\varepsilon(B_\rho(x,2\varepsilon)|E)$ which satisfies the inequality

$$2^m \le \mathcal{M}_\varepsilon(B_\rho(x,2\varepsilon)|E) < 2^{m+1}.$$

We conclude this section with a result from [19], relating $\varepsilon$-covering and $\varepsilon$-packing.

**Lemma 5.** *Given $R \ge 2\varepsilon > 0$, let $k$ and $m$ be natural numbers such that*

$$2 \cdot 7^k \le \frac{R}{\varepsilon} \le 2^m.$$

*For all $z \in E$, the following hold:*

$$\mathcal{N}_\varepsilon\left(B_\rho(z,R) \,\middle|\, E\right) \le 2^{m\mathbf{d}(E)} \tag{2.3.1}$$

*and*

$$\mathcal{M}_\varepsilon\left(B_\rho(z,R) \,\middle|\, E\right) \ge 2^{(k+1)\mathbf{p}(E)}. \tag{2.3.2}$$

*Proof.* **1.** For every $n \ge 0$, we first show that

$$\mathcal{N}_\varepsilon\left(B_\rho(z,2^n\varepsilon) \,\middle|\, E\right) \le 2^{n\mathbf{d}(E)} \qquad \text{for all } z \in E. \tag{2.3.3}$$

Assuming that (2.3.3) holds for $n = i \geq 0$, for any given $z_0 \in E$, from Definition 4, we have

$$\mathcal{N}_{2^i\varepsilon}\left(B_\rho(z_0, 2^{i+1}\varepsilon)\,\middle|\, E\right) \leq 2^{\mathbf{d}(E)}.$$

Equivalently, there exist $x_1, x_2, \ldots, x_{2^{\mathbf{d}(E)}} \in E$ such that

$$B_\rho(z_0, 2^{i+1}\varepsilon) \subseteq \bigcup_{j=1}^{2^{\mathbf{d}(E)}} B_\rho(x_j, 2^i\varepsilon)$$

and

$$\mathcal{N}_\varepsilon\left(B_\rho(z_0, 2^{i+1}\varepsilon)\,\middle|\, E\right) \leq \sum_{j=1}^{2^{\mathbf{d}(E)}} \mathcal{N}_\varepsilon\left(B_\rho(x_j, 2^i\varepsilon)\,\middle|\, E\right) \leq 2^{\mathbf{d}(E)} \cdot 2^{i\mathbf{d}(E)} = 2^{(i+1)\mathbf{d}(E)}.$$

Thus, (2.3.3) holds for $n = i+1$ and the method of induction yields (2.3.3) for all $n \geq 0$. In particular, the non-decreasing property of the map $r \mapsto \mathcal{N}_\varepsilon\left(B_\rho(z,r)\,\middle|\, E\right)$ implies that

$$\mathcal{N}_\varepsilon\left(B_\rho(z,R)\,\middle|\, E\right) \leq \mathcal{N}_\varepsilon\left(B_\rho(z,2^m\varepsilon)\,\middle|\, E\right) \leq 2^{m\mathbf{d}(E)}.$$

**2.** To achieve the inequality in (2.3.2), we prove that

$$\mathcal{M}_\varepsilon\left(B_\rho(z, 2\cdot 7^n\varepsilon)\,\middle|\, E\right) \geq 2^{(n+1)\mathbf{p}(E)} \qquad \text{for all } z \in E. \tag{2.3.4}$$

It is clear from Definition 4 that (2.3.4) holds for $n = 0$. Assuming that (2.3.4) holds for $n = i \geq 1$, for any given $z_0 \in E$, from Definition 4, we have

$$\mathcal{M}_{6\cdot 7^i\varepsilon}\left(B_\rho(z_0, 12\cdot 7^i\varepsilon)\,\middle|\, E\right) \geq 2^{\mathbf{p}(E)}.$$

Equivalently, there exist $x_1, x_2, \ldots, x_{2^{\mathbf{p}(E)}} \in B_\rho(z_0, 12\cdot 7^i\varepsilon)$ such that

$$\rho(x_{j_1}, x_{j_2}) > 6\cdot 7^i\varepsilon \geq 4\cdot 7^i\varepsilon + 2\varepsilon \qquad \text{for all } j_1 \neq j_2 \in \{1, 2, \ldots, 2^{\mathbf{p}(E)}\}.$$

In particular, for every $j_1 \neq j_2 \in \{1, 2, \ldots, 2^{\mathbf{p}(E)}\}$, it holds that

$$\rho(z_1, z_2) > 2\varepsilon \qquad \text{for all } z_1 \in B_\rho(x_{j_1}, 2\cdot 7^i\varepsilon), z_2 \in B_\rho(x_{j_2}, 2\cdot 7^i\varepsilon).$$

Since $B_\rho(x_j, 2\cdot 7^i\varepsilon) \subseteq B_\rho(z_0, 2\cdot 7^{i+1}\varepsilon)$ for all $j \in \{1, 2, \ldots, 2^{\mathbf{p}(E)}\}$,

$$\mathcal{M}_\varepsilon\left(B_\rho(z_0, 2\cdot 7^{i+1}\varepsilon)\,\middle|\, E\right) \geq \sum_{j=1}^{2^{\mathbf{p}(E)}} \mathcal{M}_\varepsilon\left(B_\rho(x_j, 2\cdot 7^i\varepsilon)\,\middle|\, E\right) \geq 2^{\mathbf{p}(E)} \cdot 2^{(i+1)\mathbf{p}(E)} = 2^{(i+2)\mathbf{p}(E)}.$$

Thus, by the method of induction, (2.3.4) holds for all $n \geq 0$. In particular, the non-decreasing property of the map $r \mapsto \mathscr{M}_\varepsilon\Big(B_\rho(z,r)\Big| E\Big)$ implies that

$$\mathscr{M}_\varepsilon\Big(B_\rho(z,R)\Big| E\Big) \geq \mathscr{M}_\varepsilon\Big(B_\rho(z,2\cdot 7^k\varepsilon)\Big| E\Big) \geq 2^{(k+1)\mathbf{p}(E)}.$$

□

As a consequence of Lemma 3 and Lemma 5, we obtain

$$\left(\frac{R}{4\varepsilon}\right)^{\log_7(2)\cdot\mathbf{p}(E)} \leq \mathscr{N}_\varepsilon\Big(B_\rho(z,R)\Big| E\Big) \leq \left(\frac{2R}{\varepsilon}\right)^{\mathbf{d}(E)} \tag{2.3.5}$$

and

$$\left(\frac{R}{2\varepsilon}\right)^{\log_7(2)\cdot\mathbf{p}(E)} \leq \mathscr{M}_\varepsilon\Big(B_\rho(z,R)\Big| E\Big) \leq \left(\frac{4R}{\varepsilon}\right)^{\mathbf{d}(E)}. \tag{2.3.6}$$

## 2.4 Semiconcave functions

In this section we state some basic definitions and properties of semiconcave (semiconvex) functions in $\mathbb{R}^d$. We refer to [18] for a general introduction to the respective theories.

**Definition 6.** A continuous function $u : \Omega \to \mathbb{R}^d$ is semiconcave if there exists a nondecreasing continuous function $\omega : [0,\infty) \to [0,\infty)$ with $\lim_{s\to 0+} \omega(s) = 0$ such that

$$u(x+h) + u(x-h) - 2u(x) \leq \omega(|h|)\cdot|h|$$

for all $x, h \in \mathbb{R}^d$ such that $[x-h, x+h] \subset \Omega$. We say that

- $u$ is semiconcave in $\Omega$ with a semiconcavity constant $K$ if $\omega(s) = K s$ for all $s \in [0,+\infty)$;
- $u$ is semiconvex (with constant -$K$) if $-u$ is semiconcave (with constant K);
- $u$ is locally semiconcave (semiconvex) if $u$ is semiconcave (semiconvex) in every compact set $A \subset \Omega$.

For every $x \in \Omega$ with $\Omega \subseteq \mathbb{R}^d$ open, the sets

$$D^+u(x) := \left\{p \in \mathbb{R}^d : \limsup_{y\to x} \frac{u(y) - u(x) - \langle p, y-x\rangle}{|y-x|} \leq 0\right\}$$

and

$$D^-u(x) := \left\{p \in \mathbb{R}^d : \liminf_{y\to x} \frac{u(y) - u(x) - \langle p, y-x\rangle}{|y-x|} \geq 0\right\}$$

are called the superdifferential and the subdifferential of $u$ at $x$ respectively. It is clear that $D^{\pm}u(x)$ is convex and

$$D^{-}u(x) = -D^{+}(-u)(x) \qquad \text{for all } x \in \Omega.$$

The superdifferential of a semiconcave function possesses the following properties as proved in [18, Proposition 3.3.4, Proposition 3.3.10].

**Proposition 7.** *Let $u : \Omega \to \mathbb{R}$ be locally semiconcave with $\Omega \subseteq \mathbb{R}^d$ open, convex. Then:*

*(i) The superdifferential $D^{+}u(x)$ is a compact, convex, nonempty set for all $x \in \Omega$. Moreover, the set-valued map $x \mapsto D^{+}u(x)$ is upper semicontinuous;*

*(ii) $D^{+}u(x)$ is a singleton if and only if $u$ is differentiable at $x$;*

*(iii) If $D^{+}u(x)$ is a singleton for all $x \in \Omega$, then $u \in \mathscr{C}^1(\Omega)$;*

*(iv) For every $x, y \in \Omega$, it holds that*

$$\langle p_y - p_x, y - x\rangle \leq \omega(|x-y|)\cdot|y-x|$$

*for all $p_x \in D^{+}u(x)$ and $p_y \in D^{+}u(y)$.*

As a consequence of (ii)-(iii) if $u$ is both locally semiconcave and locally semiconvex then $u$ is in $\mathscr{C}^1(\Omega)$. This is crucial to prove further regularity results for viscosity solutions of Hamilton-Jacobi equations, to be dealt with in Section 2.7. From (iv), we get

**Corollary 8.** *If $u : \Omega \to \mathbb{R}$ is semiconvex with constant $-K$ then for every $x, y \in \Omega$, it holds that*

$$\langle p_y - p_x, y - x\rangle \geq -K\cdot|y-x|^2$$

*for all $p_x \in D^{-}u(x)$ and $p_y \in D^{-}u(y)$.*

For any given constants $r, K > 0$, we define

$$\mathscr{SC}_{[r,K]} := \left\{v \in \mathbf{Lip}(\mathbb{R}^d) : \mathrm{Lip}[v] \leq r \text{ and } v \text{ is semiconcave with constant } K\right\}. \tag{2.4.1}$$

From the proof of [3, Proposition 10], we obtain a lower bound on the $\varepsilon$-entropy for the set $\left\{Dv_{\llcorner\Box_R} : v \in \mathscr{SC}_{[r,K]}\right\}$ in $\mathbf{L}^1(\Box_R)$, which will be used to establish a lower estimate on the $\varepsilon$-entropy of a set of viscosity solutions in Section 3.2.

**Corollary 9.** *Given any $r, R, K > 0$, for every*

$$0 < \varepsilon \le \min\{r, K\} \cdot \frac{\omega_d \cdot R^d}{(d+1)2^{d+8}},$$

*there exists a subset $\mathscr{G}^R_{[r,K]}$ of $\mathscr{S}\mathscr{C}_{[r,K]}$ such that*

$$\mathrm{Card}\left(\mathscr{G}^R_{[r,K]}\right) \ge 2^{\beta_{[R,K]}\cdot\varepsilon^{-d}} \qquad \text{with} \qquad \beta_{[R,K]} = \frac{1}{3^d 2^{d^2+4d+3}\ln 2} \cdot \left(\frac{K\omega_d R^{d+1}}{(d+1)}\right)^d$$

*and*

$$\left\| D v_{\llcorner \square_R} - D w_{\llcorner \square_R} \right\|_{\mathbf{L}^1(\square_R)} \ge 2\varepsilon \qquad \textit{for all } v \neq w \in \mathscr{G}^R_{[r,K]}.$$

## 2.5 Bounded variation functions

This section introduces the concept of functions with uniformly bounded total variation and then we extend this notion to investigate classes of functions having bounded total generalized variation. We refer to [2] for a comprehensive analysis of this topic.

### 2.5.1 Functions of bounded total variation

**Definition 10.** The function $u \in \mathbf{L}^1(\Omega)$ taking values in $\mathbb{R}^m$ is a function of bounded variation on $\Omega \subseteq \mathbb{R}^d$, that is, $u \in BV(\Omega, \mathbb{R}^m)$, if the distributional derivative of $u$, denoted by $Du$, is an $m \times d$ matrix of finite measures $D_i u^\alpha$ in $\Omega$ satisfying

$$\sum_{\alpha=1}^m \int_\Omega u^\alpha \mathrm{div}\, \varphi^\alpha \, dx = -\sum_{\alpha=1}^m \sum_{i=1}^d \int_\Omega \varphi_i^\alpha \, dD_i u^\alpha \quad \text{for all } \varphi \in \left[\mathscr{C}_c^1(\Omega, \mathbb{R}^d)\right]^m, i \in \{1, \dots, d\}.$$

We denote by $|Du|$ the total variation of $Du$, i.e.,

$$|Du|(\Omega) = \sup\left\{ \sum_{\alpha=1}^m \int_\Omega u^\alpha \mathrm{div}^\alpha \varphi \;\middle|\; \varphi \in \left[\mathscr{C}_c^1(\Omega, \mathbb{R}^d)\right]^m, \|\varphi\|_{\mathbf{L}^\infty(\Omega)} \le 1 \right\}.$$

In particular, for real-valued functions, the above definition may be rewritten as follows.

**Definition 11.** The function $u \in \mathbf{L}^1(\Omega)$ is a function of bounded variation on $\Omega$, that is, $u \in BV(\Omega, \mathbb{R})$, if the distributional derivative of $u$ is representable by a finite Radon measure in $\Omega$, i.e., if

$$\int_\Omega u \cdot \frac{\partial \varphi}{\partial x_i} \, dx = -\int_\Omega \varphi \, dD_i u \qquad \text{for all } \varphi \in \mathscr{C}_c^1(\Omega, \mathbb{R}), i \in \{1, 2, \dots, n\}$$

for some Radon measure $Du = (D_1 u, D_2 u, ..., D_n u)$. We denote the total variation of the vector measure $Du$ by $|Du|$, i.e.,

$$|Du|(\Omega) = \sup\left\{ \int_\Omega u(x)\mathrm{div}\,(\varphi) \;\middle|\; \varphi \in \mathscr{C}_c^1(\Omega, \mathbb{R}^n), \|\varphi\|_{\mathbf{L}^\infty(\Omega)} \leq 1 \right\}.$$

Let us state Helly's theorem ([31]) on compactness of functions having bounded total variation, which motivates us to find metric entropy estimates for such function classes.

**Theorem 12.** *(Helly's selection principle) Every sequence of uniformly bounded variation functions has a convergent subsequence, i.e., $BV(\Omega, \mathbb{R}^m)$ is sequentially compact in $\mathbf{L}^1(\Omega)$.*

We recall a Poincaré-type inequality for bounded total variation functions on a convex domain. This result is based on [1, Theorem 3.2] and on [2, Proposition 3.2.1, Theorem 3.44] and is essential for our work in Section 3.1.

**Theorem 13.** *(Poincaré inequality) Let $\Omega \subset \mathbb{R}^d$ be an open, bounded, convex set with a Lipschitz boundary. For any $u \in BV(\Omega, \mathbb{R})$, it holds that*

$$\int_\Omega \left| u(x) - u_\Omega \right| dx \leq \frac{\mathrm{diam}(\Omega)}{2} \cdot |Du|(\Omega)$$

*where $u_\Omega = \frac{1}{\mathrm{Vol}(\Omega)} \cdot \int_\Omega u(x)\, dx$ is the mean value of $u$ over $\Omega$.*

Now we put forward some remarks on functions of bounded variation proved in [2], which are useful for our proofs in Section 3.2.

**Remark 14.** Let $\Omega \subset \mathbb{R}^2$ be an open, bounded set and let us assume that there exist pairwise disjoint open sets $\{\Omega_i\}_{1 \leq i \leq n}$ with piecewise $\mathscr{C}^1$ boundary such that

$$\bigcup_{i=1}^n \Omega_i \subset \Omega \subset \bigcup_{i=1}^n \overline{\Omega}_i .$$

If we have functions $u_i \in \mathscr{C}^1(\overline{\Omega}_i)$, we can construct a piecewise defined function $u : \Omega \to \mathbb{R}$ which is equal to $u_i$ on each $\Omega_i$ and defined arbitrarily on the remaining set of measure zero denoted by $\Sigma$. Applying Gauss-Green theorem to every $\Omega_i$, for $i = 1, \cdots, p$, it holds that

$$\int_{\Omega_i} u \mathrm{div}\, \varphi\, dx = -\int_{\Omega_i} \langle \nabla u, \varphi \rangle dx - \int_{\partial\Omega_i} u_i \langle \varphi, \nu_i \rangle d\mathscr{H}^1 \qquad \text{for all } \varphi \in [\mathscr{C}^1(\overline{\Omega}_i)]^2$$

where $\nu_i$ is the inner unit normal to $\Omega_i$. Summing over $i$ yields $u \in BV(\Omega, \mathbb{R})$ with $Du$ given by

$$\nabla u \mathscr{L}^2 + \sum_{i=1}^{p} u_i \nu_i \mathscr{H}^1_{\llcorner(\Omega \cap \partial\Omega_i)} .$$

**Definition 15.** Let $A$ be an $\mathscr{L}^d$-measurable subset of $\mathbb{R}^d$. For any open subset $\Omega \subset \mathbb{R}^d$, the perimeter of $A$ in $\Omega$, denoted by $P(A,\Omega)$, is the total variation of $\chi_A$ in $\Omega$, i.e.,

$$P(A,\Omega) := \sup\left\{ \int_A \operatorname{div} \varphi \, dx \;\middle|\; \varphi \in [\mathscr{C}^1_c(\Omega,\mathbb{R}^d]^d, \|\varphi\|_{\mathbf{L}^\infty(\Omega)} \leq 1 \right\}. \tag{2.5.1}$$

We say that $A$ is a set of finite perimeter in $\Omega$ if $P(A,\Omega) < \infty$ .

**Remark 16.** The class of sets of finite perimeter in $\Omega$ includes all sets $A$ with $\mathscr{C}^1$ boundary inside $\Omega$ such that $\mathscr{H}^{d-1}(\Omega \cap \partial\Omega) < \infty$. Indeed, by Gauss-Green theorem, for every set $A$, it holds that

$$\int_A \operatorname{div} \varphi \, dx = -\int_{\Omega \cap \partial\Omega} \langle \nu_A, \varphi \rangle \, d\mathscr{H}^{d-1} \qquad \text{for all } \varphi \in [\mathscr{C}^1_c(\Omega,\mathbb{R}^d]^d \tag{2.5.2}$$

where $\nu_A$ is the inner unit normal to $E$. Using 2.5.2 in 2.5.1 yields $P(A,\Omega) = \mathscr{H}^{d-1}(\Omega \cap \partial\Omega)$ .

Finally, let us state an important theorem ([2, Theorem 3.9]) which characterizes bounded variation functions.

**Theorem 17.** *Let $u \in [\mathbf{L}^1(\Omega)]^m$. Then $u \in [BV(\Omega,\mathbb{R}^m)]^m$ if and only if there exists a sequence $(u_n)_{n\geq 1} \subset [\mathscr{C}^\infty(\Omega,\mathbb{R}^m)]^m$ which converges to $u$ in $[\mathbf{L}^1(\Omega)]^m$ and satisfies*

$$L := \lim_{n\to\infty} \int_\Omega |\nabla u_n| dx < \infty . \tag{2.5.3}$$

*Moreover, the least constant $L$ in 2.5.3 is $|Du|(\Omega)$.*

### 2.5.2 Functions of bounded total generalized variation

In this subsection, we explain the meaning of total generalized variation of a function $g : [a,b] \to E$ which was well-studied in [42] for the case $E = \mathbb{R}$.
Consider a convex function $\Psi : [0,+\infty) \to [0,+\infty)$ such that

$$\Psi(0) = 0 \qquad \text{and} \qquad \Psi(s) > 0 \qquad \text{for all } s > 0. \tag{2.5.4}$$

**Definition 18.** The $\Psi$-total variation of $g$ over $[a,b]$ is defined as

$$TV^\Psi\big(g,[a,b]\big) = \sup_{n\in\mathbb{N}, a=x_0<x_1<\ldots<x_n=b} \sum_{i=0}^{n-1} \Psi\big(\rho(g(x_i), g(x_{i+1}))\big). \tag{2.5.5}$$

If the supremum is finite then we say that $g$ has bounded $\Psi$-total variation and denote it by $g \in BV^{\Psi}([a,b],E)$. In the case of $\Psi(x) = |x|^{\gamma}$ for some $\gamma \geq 1$, we denote by

$$BV^{\frac{1}{\gamma}}([a,b],E) := BV^{\Psi}([a,b]), \qquad TV^{\frac{1}{\gamma}}\big(g,[a,b]\big) \;:=\; TV^{\Psi}\big(g,[a,b]\big)$$

the fractional BV space on $[a,b]$ and the $\gamma$-total variation of $g$, respectively.

For any function $g \in BV^{\Psi}([a,b],E)$, by a contradiction argument we infer that $g$ is a regulated function, i.e., the left and right hand side limits of $g$ at $x_0 \in [a,b]$ always exist and are denoted by

$$g(x_0-) \;:=\; \lim_{x \to x_0-} g(x) \qquad \text{and} \qquad g(x_0+) \;:=\; \lim_{x \to x_0+} g(x)\,.$$

Moreover, the set of discontinuities of $g$

$$\mathscr{D}_g \;:=\; \big\{x \in [a,b] \,\big|\, g(x+) = g(x) = g(x-) \;\text{ does not hold}\big\}$$

is at most countable. In particular, we have the following:

**Lemma 19.** *For any function $g \in BV^{\Psi}([a,b],E)$, the following function*

$$\tilde{g}(b) \;=\; g(b), \qquad \tilde{g}(x) \;:=\; g(x+) \qquad \textit{for all } x \in [a,b)$$

*is a continuous function from the right on the interval $[a,b)$ and belongs to $BV^{\Psi}([a,b],E)$ with*

$$\rho_{\mathbf{L}^1}(\tilde{g},g) \;=\; 0 \qquad \text{and} \qquad TV^{\Psi}\big(\tilde{g},[a,b]\big) \;\leq\; TV^{\Psi}\big(g,[a,b]\big). \tag{2.5.6}$$

*Proof.* Since $\mathscr{D}_g$ is at most countable, it holds that

$$\rho_{\mathbf{L}^1}(\tilde{g},g) \;=\; \int_{[a,b]\setminus\mathscr{D}_g} \rho(\tilde{g}(x),g(x))dx \;=\; 0\,.$$

On the other hand, for any partition $\{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a,b]$,

$$\begin{aligned}\sum_{i=0}^{n-1}\Psi(\rho(\tilde{g}(x_{i+1}),\tilde{g}(x_i))) \;&=\; \Psi(\rho(g(b),g(x_{n-1}+))) + \sum_{i=0}^{n-2}\Psi(\rho(g(x_{i+1}+),g(x_i+)))\\ &\leq\; TV^{\Psi}(g,[a,b])\end{aligned}$$

and this yields the second inequality in (2.5.6).

□

The following remark is used in the proof of the upper estimate in Section 4.1.

**Remark 20.** Under the assumption (2.5.4), the function $\Psi$ is strictly increasing on $[0,+\infty)$ and

$$\Psi(s) \leq \frac{s}{t}\cdot\Psi(t) \quad \text{for all } 0 \leq s < t\,. \tag{2.5.7}$$

Moreover, its inverse $\Psi^{-1}$ is also strictly increasing, concave and the map $s \longmapsto \dfrac{\Psi^{-1}(s)}{s}$ is strictly decreasing on $[0,+\infty)$.

*Proof.* By the convexity of $\Psi$ and (2.5.4),

$$\Psi(s) \leq \frac{t-s}{t}\cdot\Psi(0)+\frac{s}{t}\cdot\Psi(t) = \frac{s}{t}\cdot\Psi(t) < \Psi(t)$$

for all $0 \leq s < t$. Thus, $\Psi$ is strictly increasing and convex in $[0,+\infty)$ and this implies that its inverse $\Psi^{-1}$ exists, is strictly increasing and concave. In particular,

$$\frac{\Psi^{-1}(s)}{s} = \frac{\Psi^{-1}(s)-\Psi^{-1}(0)}{s} > \frac{\Psi^{-1}(r)}{r} \qquad \text{for all } 0 < s < r$$

and this yields the decreasing property of the map $s \longmapsto \dfrac{\Psi^{-1}(s)}{s}$.

□

We conclude this subsection by stating Helly's extracting theorem ([42, Theorem 1.3]), which extends Helly's selection principle to the case of bounded generalized variation functions.

**Theorem 21.** *(Helly's extracting theorem) Every sequence* $\left(f_n\right)_{n\geq1} \subset BV^{\Psi}([a,b],E)$ *has a subsequence which converges to a function* $f \in BV^{\Psi}([a,b],E)$ *pointwise on* $[a,b]$.

## 2.6 Conservation laws

A scalar conservation law in one-dimensional space is a first-order partial differential equation of the form

$$u_t(t,x)+f(u(t,x))_x = 0 \qquad (t,x)\in[0,\infty)\times\mathbb{R} \tag{2.6.1}$$

where $u:[0,\infty)\times\mathbb{R}\to\mathbb{R}$ is called the conserved quantity and $f:\mathbb{R}\to\mathbb{R}$ is the flux.

Here we introduce the fundamentals of this topic and refer to [16] for the related theories.

A basic feature of nonlinear systems of the above form is that, even for smooth initial data, the solution of the Cauchy problem may develop discontinuities in finite time. In order to prolong the solution to 2.6.1 after the formation of a discontinuity, we need to adopt the concept of a weak solution in the distributional sense, which will allow the presence of discontinuities in the solution or in its space derivatives.

**Definition 22.** A function $u \in \mathbf{L}^{\infty}((0,T)\times\mathbb{R})$ is said to be a weak solution of 2.6.1 if for every $\varphi \in \mathscr{C}_c^1((0,T)\times\mathbb{R},\mathbb{R})$, it holds that

$$\int\int_{(0,T)\times\mathbb{R}} [u(t,x)\varphi_t(t,x)+f(u(t,x))\varphi_x(t,x)]\,dt\,dx \;=\; 0\,.$$

**Remark 23.** A function $u \in \mathscr{C}^1((0,T)\times\mathbb{R})$ is a classical solution of 2.6.1 if and only if $u$ is a weak solution of 2.6.1.

**Lemma 24.** *(Closure of set of weak solutions in $\mathbf{L}^1_{\text{loc}}$) Let $(u_n)_{n\geq1}$ be a sequence of weak solutions of 2.6.1 such that*

$$u_n \longrightarrow u \text{ and } f(u_n) \longrightarrow f(u) \text{ in } \mathbf{L}^1_{\text{loc}}\,. \tag{2.6.2}$$

*Then $u$ is also a weak solution of 2.6.1.*

Let us now define a weak solution of a Cauchy problem

$$\begin{cases} u_t(t,x)+f(u(t,x))_x \;=\; 0 \\ \\ u(0,\cdot) \;=\; u_0(\cdot) \qquad\qquad \text{for a given } u_0 \in \mathbf{L}^1_{\text{loc}}(\mathbb{R})\;. \end{cases} \tag{2.6.3}$$

**Definition 25.** A function $u:[0,T]\times\mathbb{R}\longrightarrow\mathbb{R}$ is a weak solution of 2.6.3 if $u$ is a weak solution of 2.6.1 on the strip $(0,T)\times\mathbb{R}$ and the map $t\longmapsto u(t,\cdot)$ is continuous with values in $\mathbf{L}^1_{\text{loc}}$ for $t\in[0,T]$ with $u(0,\cdot) \;=\; u_0(\cdot)$.

In this case, the following lemma from [16] holds.

**Lemma 26.** *If $u:[0,T]\times\mathbb{R}\longrightarrow\mathbb{R}$ is a weak solution of 2.6.3, then $u$ is also a solution of 2.6.3 in the distributional sense, i.e., for every $\varphi\in\mathscr{C}_C^1((-1,T)\times\mathbb{R})$, it holds that*

$$\int\int_{(0,T)\times\mathbb{R}} [u(t,x)\varphi_t(t,x)+f(u(t,x))\varphi_x(t,x)]\,dt\,dx \;+\; \int_{-\infty}^{\infty} u_0(x)\varphi(0,x)dx \;=\; 0\,.$$

Now we state a necessary condition derived in [16] for the piecewise constant function $U$ as defined below, to be a weak solution of 2.6.1. The function $U$ is given by

$$U(t,x) = \begin{cases} u^+ & \text{if } x > \lambda \cdot t \\ u^- & \text{if } x < \lambda \cdot t \end{cases} \tag{2.6.4}$$

for some $u^\pm, \lambda \in \mathbb{R}$.

**Lemma 27.** *(Rankine-Hugoniot condition) The function $U$ in 2.6.4 is a weak solution of 2.6.1 if and only if it holds that*

$$\lambda = \frac{f(u^+) - f(u^-)}{u^+ - u^-}. \tag{2.6.5}$$

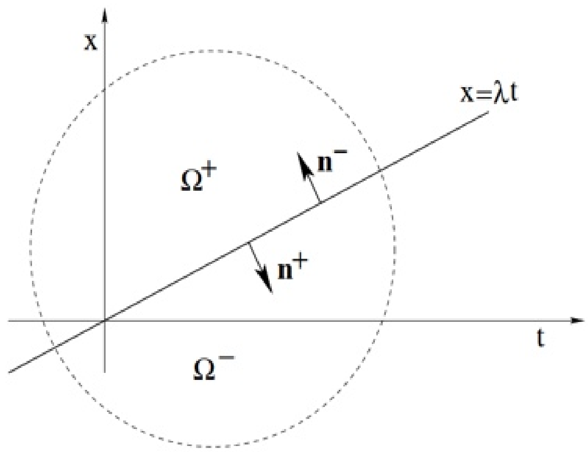


**Figure 2.5** Deriving the Rankine-Hugoniot condition

Towards deriving condition 2.6.5 for general weak solutions of 2.6.1, we define the following.

**Definition 28.** *(Approximate jump)* A real-valued function $u \in \mathbf{L}^1_{\text{loc}}$ has an approximate jump discontinuity at a point $(\tilde{t}, \tilde{x})$ if there exist $u^\pm, \lambda \in \mathbb{R}$ such that upon setting $U$ as defined in 2.6.4, it holds that

$$\lim_{r \to 0^+} \frac{1}{r^2} \int\int_{[-r,r]^2} \left| u(\tilde{t} + t, \tilde{x} + x - U(t,x) \right| dx\, dt = 0. \tag{2.6.6}$$

In this case, $u^-$ amd $u^+$ are the left and right approximate limits of $u$ at $(\tilde{t}, \tilde{x})$ and $\lambda$ is the jump speed.

The next proposition follows from this definition.

**Proposition 29.** *Let $u$ be a bounded weak solution of 2.6.1 having an approximate jump discontinuity at a point $(\tilde{t}, \tilde{x})$, i.e., 2.6.6 holds for some $u^\pm, \lambda \in \mathbb{R}$. Then the Rankine-Hugoniot condition 2.6.5 holds.*

The concept of weak solution is not sufficient to find a unique solution whenever a strong discontinuity appears in the solution. Therefore we additionally provide the following admissibility conditions that can be present in a weak solution, to achieve uniqueness of solutions and continuous dependence on the initial data.

- *Vanishing viscosity:* We say that a weak solution $u : \Omega \to \mathbb{R}$ of 2.6.1 is admissible in the vanishing viscosity sense if there exists a sequence of smooth solutions of the viscous parabolic approximation

$$u_t^\varepsilon + f'(u^\varepsilon) \cdot u_x^\varepsilon = \varepsilon u_{xx}^\varepsilon \tag{2.6.7}$$

  so that $u^\varepsilon$ converges to $u$ in $\mathbf{L}^1_{\text{loc}}$ as $\varepsilon \to 0^+$.

- *Entropy conditions:* Motivated by the second principle of thermodynamics, as kinetic energy is dissipated when a shock appears, we introduce the following concept of entropy which characterizes irreversible processes.

  **Definition 30.** *(Entropy-Entropy flux)* We say that a pair of $\mathscr{C}^1$ functions $(\eta, q) : \mathbb{R} \to \mathbb{R}$ is an entropy-entropy flux pair for 2.6.1 if

$$q'(u) = \eta'(u) \cdot f'(u) \tag{2.6.8}$$

  at every $u$ where $\eta, q$ and $f$ are differentiable.

  **Remark 31.** If $u$ is a classical solution of 2.6.1, then $u$ solves the equation

$$[\eta(u)]_t + [q(u)]_x = 0 .$$

  In this case, $\eta(u)$ is conserved. However it is not conserved in general, when $u$ is discontinuous, as shown in [16].

Let us now formulate the notion of an entropy admissible weak solution. Let $u^\varepsilon$ be the smooth solution of 2.6.7. We also see that it is a solution of the equation

$$[\eta(u^\varepsilon)]_t + [q(u^\varepsilon)]_x = \varepsilon \cdot \left( [\eta(u^\varepsilon)]_{xx} - \eta''(u^\varepsilon) \cdot [u_x^\varepsilon]^2 \right) .$$

In particular, if $\eta$ is convex and smooth, we have

$$[\eta(u^\varepsilon)]_t + [q(u^\varepsilon)]_x \le \varepsilon \cdot [\eta(u^\varepsilon)]_{xx} .$$

Thus for every non-negative test function $\varphi \in \mathscr{C}_C^1(\mathbb{R})$, it holds that

$$\int\int_\Omega [\eta(u^\varepsilon)\varphi_t + q(u^\varepsilon)\varphi_x] \, dt \, dx \ge -\varepsilon \int\int_\Omega \eta(u^\varepsilon)\varphi_{xx} \, dt \, dx .$$

If $u^\varepsilon$ converges to $u$ in $\mathbf{L}^1_{\text{loc}}$ then as $\varepsilon \to 0^+$, we get

$$\int\int_\Omega [\eta(u)\varphi_t + q(u)\varphi_x]\, dt\, dx \;\geq\; 0\,. \tag{2.6.9}$$

This yields the following entropy admissible condition.

**Definition 32.** A weak solution $u$ of 2.6.1 is entropy admissible if it satisfies the inequality

$$[\eta(u)]_t + [q(u)]_x \;\leq\; 0$$

in the distributional sense for every pair of convex entropy-entropy flux $(\eta, q)$, i.e., 2.6.9 holds for every non-negative test function $\varphi \in \mathscr{C}^1_C(\mathbb{R})$.

In order to establish some stability conditions, we recall $U$ as in 2.6.4 and consider a slightly perturbed solution where the original shock joining two states $u^\pm$ is split into two separated smaller shocks that join $u^+$ and $u^-$ with an intermediate state

$$u^\alpha = \alpha u^+ + (1-\alpha)u^- \qquad \text{for some } \alpha \in (0,1).$$

To ensure that the $\mathbf{L}^1$-distance between the original solution and the perturbed one does not increase in time, we must have [speed of jump behind] $\geq$ [speed of jump ahead]. By 2.6.5 this implies,

$$\frac{f(u^\alpha) - f(u^-)}{u^\alpha - u^-} \;\geq\; \frac{f(u^+) - f(u^\alpha)}{u^+ - u^\alpha}\,. \tag{2.6.10}$$

This is equivalent to

$$\begin{cases} f(\alpha u^+ + (1-\alpha)u^-) \;\geq\; \alpha f(u^+) + (1-\alpha) f(u^-) & \text{if } u^- < u^+ \\ \\ f(\alpha u^+ + (1-\alpha)u^-) \;\leq\; \alpha f(u^+) + (1-\alpha) f(u^-) & \text{if } u^- > u^+ \end{cases}\,. \tag{2.6.11}$$

**Proposition 33.** *The function $U$ is an entropy admissible solution of 2.6.1 if and only if the condition 2.6.11 holds.*

Another type of admissibility condition may be defined as follows.

**Definition 34.** We say that a weak solution of 2.6.1 is admissible in the sense of Lax if at every point $(\tilde{t}, \tilde{x})$ of approximate jump discontinuity with the left and right states $u^-, u^+$

and speed $\lambda$, the Lax condition holds, i.e.,

$$f'(u) \geq \lambda = \frac{f(u^+)-f(u^-)}{u^+-u^-} \geq f'(u^+). \tag{2.6.12}$$

In the case of convex flux $f''(u) \geq 0$, the stability condition 2.6.11 and Lax condition 2.6.12 are equivalent. Moreover, if $f''(u) > 0$ then the Lax condition is equivalent to the condition

$$u^- > u^+.$$

In the case of general flux, the Lax condition does not imply the stability condition.

We conclude this section with the following theorem from [16].

**Theorem 35.** *Let $f : \mathbb{R} \to \mathbb{R}$ be locally Lipschitz continuous. Then there exists a continuous semigroup $S : [0,\infty) \times \mathbf{L}^1(\mathbb{R}) \to \mathbf{L}^1(\mathbb{R})$ such that for each $\overline{u} \in \mathbf{L}^1(\mathbb{R}) \bigcap \mathbf{L}^\infty(\mathbb{R})$, the trajectory $t \mapsto S_t\overline{u}$ yields a unique bounded, entropy admissible weak solution of 2.6.1 with $u(0,\cdot) = \overline{u}$. Moreover, the following properties hold.*

*(i)* *(Semigroup property)* $S_0\overline{u} = \overline{u}, \qquad S_sS_t\overline{u} = S_{s+t}\overline{u}$.

*(ii)* $\|S_t\overline{u} - S_t\overline{v}\|_{\mathbf{L}^1(\mathbb{R})} \leq \|\overline{u} - \overline{v}\|_{\mathbf{L}^1(\mathbb{R})}$.

*(iii)* *If $\overline{u} \leq \overline{v}$ for all $x \in \mathbb{R}$ then*

$$S_t\overline{u}(x) \leq S_t\overline{v}(x) \qquad \textit{for all } (t,x) \in [0,\infty) \times \mathbb{R}.$$

## 2.7 Hamilton-Jacobi equations

Let us consider a first-order Hamilton-Jacobi equation

$$u_t(t,x) + H\big(D_x u(t,x)\big) = 0 \qquad \text{for all } (t,x) \in (0,\infty) \times \mathbb{R}^d \tag{2.7.1}$$

where $u : [0,+\infty) \times \mathbb{R}^d \to \mathbb{R}^d$, $D_x u = (u_{x_1}, \dots, u_{x_d})$ and $H : \mathbb{R}^d \to \mathbb{R}$ is a Hamiltonian. Due to the nonlinear dependence of the characteristic speeds on the gradient of the solution, in general a classical solution $u$ will develop singularities and the gradient $Du$ will become discontinuous in finite time. To cope with this difficulty, the concept of viscosity solution was introduced by Crandall and Lions in [21] to guarantee global existence, uniqueness and stability of the Cauchy problem. Under standard assumptions of the convexity and the coercivity on Hamiltonian $H$, (2.7.1) generates a Hopf-Lax semigroup of viscosity solutions $(S_t)_{t\geq 0} : \mathbf{Lip}(\mathbb{R}^d) \to \mathbf{Lip}(\mathbb{R}^d)$. More precisely, for every Lipschitz initial data $u_0 \in \mathbf{Lip}(\mathbb{R}^d)$, the corresponding unique viscosity solution of equation (2.7.1) with $u(0,x) = u_0(x)$ is computed by the Hopf-Lax representation formula

$$u(t,x) = S_t(u_0)(x) \;=\; \min_{y\in\mathbb{R}^d}\left\{u_0(y) + t\cdot L\left(\frac{x-y}{t}\right)\right\} \tag{2.7.2}$$

where $L$ is the Legendre transform of $H$. In addition, if $H$ is strongly convex, i.e., there exists a constant $\lambda > 0$ such that

$$D^2H(p) \geqslant \lambda\cdot\mathbf{I}_d \qquad \text{for all } p\in\mathbb{R}^d,$$

then the map $x \to u(t,x) - \frac{1}{2\lambda t}\cdot\|x\|^2$ is concave for every $t > 0$. In particular, $u(t,\cdot)$ is twice differentiable almost everywhere and $D_x u(t,\cdot)$ has locally bounded total variation.

We consider (2.7.1) with a coercive and strictly convex Hamiltonian $H \in \mathscr{C}^1(\mathbb{R}^d)$, i.e., $\lim_{|p|\to\infty} \frac{H(p)}{|p|} = +\infty$ and $H(tp+(1-t)q) < t\cdot H(p) + (1-t)H(q)$ for all $t\in(0,1)$ and $p,q\in\mathbb{R}^d$ and satisfying the additional condition of uniformly directional convexity, i.e., for every constant $R > 0$ it holds that

$$\inf_{p\neq q\in\overline{B}(0,R)}\left\langle \frac{DH(p)-DH(q)}{|DH(p)-DH(q)|}, \frac{p-q}{|p-q|}\right\rangle \;:=\; \lambda_R \;>\; 0. \tag{2.7.3}$$

Moreover, without loss of generality, we assume that $H$ satisfies further conditions

$$H(0) \;=\; 0 \qquad \text{and} \qquad DH(0) \;=\; 0, \tag{2.7.4}$$

otherwise the transformations $x \mapsto x + tDH(0)$, $u(t,\cdot) \mapsto u(t,x) + t\cdot H(0)$ and $H(p) \mapsto H(p) - \langle DH(0), p\rangle$ reduce the general case to this one.

**Remark 36.** If $H \in \mathscr{C}^2(\mathbb{R}^d)$ satisfies

$$D^2H(p) \;=\; |D^2H(p)|\cdot A(p) \qquad \text{with} \qquad A(p) \;\geq\; \lambda\cdot\mathbf{I}_d \tag{2.7.5}$$

for some $\lambda > 0$ then $H$ satisfies (2.7.3).

*Proof.* For any $p \neq q \in \mathbb{R}^d$, by mean value theorem, it holds that

$$\begin{aligned} DH(p) - DH(q) &= \int_0^1 D^2H(tp+(1-t)q)\cdot(p-q)\,dt \\ &= \left[\int_0^1 A(tp+(1-t)q)\left|D^2H(tp+(1-t)q)\right|dt\right]\cdot(p-q) \end{aligned}$$

and

$$|DH(p)-DH(q)| \le |p-q| \cdot \int_0^1 |D^2H(tp+(1-t)q)|dt.$$

Thus, using (2.7.5), we estimate

$$\begin{aligned}\langle DH(p)-DH(q), p-q\rangle &= \int_0^1 \Big[(p-q)^T A(tp+(1-t)q)(p-q)\Big]\cdot \Big|D^2H(tp+(1-t)q)\Big|dt \\ &\ge \lambda\cdot|p-q|^2 \int_0^1 \Big|D^2H(tp+(1-t)q)\Big|dt \\ &\ge \lambda\cdot|DH(p)-DH(q)|\cdot|p-q|\end{aligned}$$

and this implies (2.7.3).

□

It is well-known that in general, classical smooth solutions of (2.7.1) break down and the Lipschitz continuous functions that satisfy (2.7.1) almost everywhere together with a given initial condition are not unique. To deal with this problem, the concept of a generalized solution was introduced in [21] to guarantee global existence and uniqueness results.

**Definition 37.** (Viscosity solution) We say that a continuous function $u : [0,T]\times\mathbb{R}^d$ is a viscosity solution of (2.7.1) if:

(1) u is a viscosity subsolution of (2.7.1), i.e., for every point $(t_0, x_0) \in (0,T)\times\mathbb{R}^d$ and test function $v \in \mathscr{C}^1\big((0,+\infty)\times\mathbb{R}^d\big)$ such that $u-v$ has a local maximum at $(t_0,x_0)$, it holds that

$$v_t(t_0,x_0) + H\big(D_x v(t_0,x_0)\big) \leqslant 0\,.$$

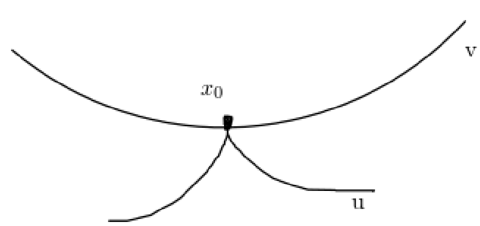


**Figure 2.6** Viscosity subsolution

(2) u is a viscosity supersolution of (2.7.1), i.e., for every point $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$ and test function $v \in \mathscr{C}^1\big((0,+\infty) \times \mathbb{R}^d\big)$ such that $u - v$ has a local minimum at $(t_0, x_0)$, it holds that

$$v_t(t_0, x_0) + H\big(D_x v(t_0, x_0)\big) \geqslant 0\,.$$

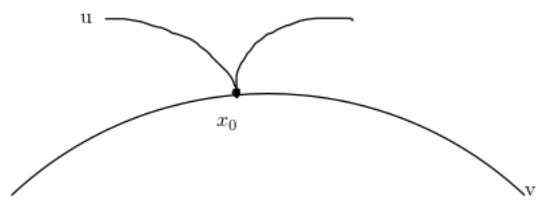


**Figure 2.7** Viscosity supersolution

By the alternative equivalent definition of viscosity solution expressed in terms of the subdifferential and superdifferential of the function (see [21]) and from Proposition 7 we observe that every $\mathscr{C}^1$ solution of (2.7.1) is also a viscosity solution of (2.7.1). On the other hand, if $u$ is a viscosity solution of (2.7.1) then $u$ satisfies the equation at every point of differentiability. Let us state a result on further regularity for viscosity solutions ([3, Proposition 3]) which says that smoothness in the pair $(t, x)$ follows from smoothness in the second variable.

**Proposition 38.** *Let $u$ be a viscosity solution of (2.7.1) in $[0, T] \times \mathbb{R}^d$. If $u(t,\cdot)$ is both locally semiconcave and semiconvex in $\mathbb{R}^d$ for all $t \in (0, T]$ then $u$ is a $\mathscr{C}^1$ solution of (2.7.1) in* $(0, T] \times \mathbb{R}$.

The viscosity solution of the Hamilton-Jacobi equation (2.7.1) with initial data $u(0,\cdot) = u_0 \in \mathrm{Lip}(\mathbb{R}^d)$ can be represented as the value function of a classical problem in calculus of variations, which admits the Hopf-Lax representation formula

$$u(t, x) = \min_{y \in \mathbb{R}^d} \left\{ t \cdot L\Big(\frac{x-y}{t}\Big) + u_0(y) \right\} \qquad t > 0, \;\; x \in \mathbb{R}^d\,, \tag{2.7.6}$$

where $L \in \mathscr{C}^1(\mathbb{R}^d)$ denotes the Legendre transform of $H$, defined by

$$L(q) := \max_{p \in \mathbb{R}^d} \big\{ p \cdot q - H(p) \big\} \qquad q \in \mathbb{R}^d. \tag{2.7.7}$$

The main properties of viscosity solutions defined by the Hopf-Lax formula, which will be used in Section 3.2, are listed below ([18, Section 1.1, Section 6.4]).

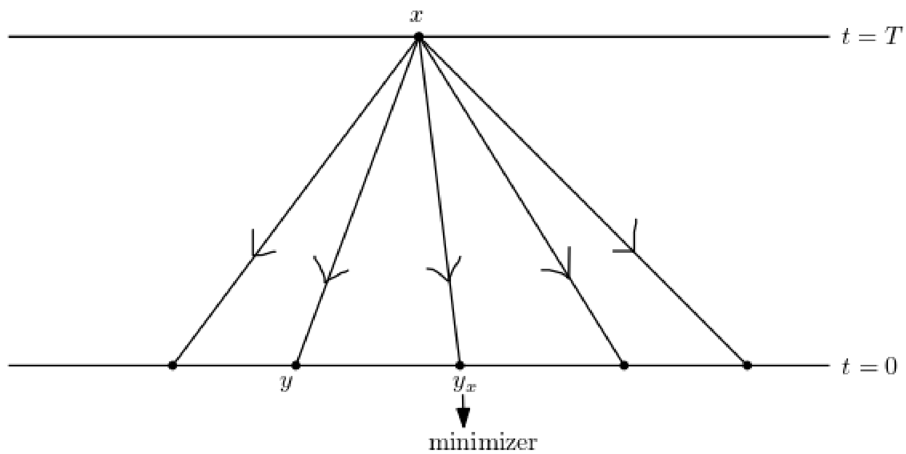


**Figure 2.8** Hopf-Lax representation formula

**Proposition 39.** *Let $u$ be the viscosity solution of* (2.7.1) *on* $[0,+\infty)\times\mathbb{R}^d$*, with continuous initial data* $u_0$*, defined by* (2.7.6)*. Then the following hold true.*

$(i)$ *Functional identity: For all* $x\in\mathbb{R}^d$ *and* $0\leqslant s<t$*, it holds that*

$$u(t,x)=\min_{y\in\mathbb{R}^d}\Big\{u(s,y)+(t-s)\cdot L\Big(\frac{x-y}{t-s}\Big)\Big\}.$$

$(ii)$ *Differentiability of $u$ and uniqueness:* (2.7.6) *admits a unique minimizer* $y_x$ *if and only if* $u(t,\cdot)$ *is differentiable at* $x$*. In this case we have*

$$y_x \;=\; x-t\cdot DH\big(D_x u(t,x)\big) \qquad \text{and} \qquad D_x u(t,x)\;\in\; D^- u_0(y_x).$$

$(iii)$ *Dynamic programming principle: Let* $t>s>0$, $x\in\mathbb{R}^d$*. Let us assume that $y$ is a minimizer for* (2.7.6) *and define* $z=\frac{s}{t}x+\Big(1-\frac{s}{t}\Big)y$*. Then $y$ is the unique minimizer over* $\mathbb{R}^d$ *of*

$$w\mapsto s\cdot L\Big(\frac{z-w}{s}\Big)+u_0(w) \quad \textit{for all } w\in\mathbb{R}^d.$$

As a consequence, the family of nonlinear operators

$$S_t:\mathbf{Lip}(\mathbb{R}^d)\to\mathbf{Lip}(\mathbb{R}^d), \qquad u_0\mapsto S_t u_0,\quad t\geqslant 0,$$

defined by

$$\begin{cases} S_t\, u_0(x) \;:=\; \min_{y\in\mathbb{R}^d}\left\{t\cdot L\Big(\dfrac{x-y}{t}\Big)+u_0(y)\right\} & t>0\,,\; x\in\mathbb{R}^d \\ \\ S_0\, u_0(x) \;:=\; u(x) & x\in\mathbb{R}^d \end{cases} \tag{2.7.8}$$

possesses the following properties as proved in [3]:

(i) For every $u_0\in\mathbf{Lip}(\mathbb{R}^d)$, $u(t,x):=S_t\,u_0(x)$ provides the unique viscosity solution of the Cauchy problem (2.7.1) with initial data $u(0,\cdot)=u_0$.

(ii) (Semigroup property)

$$S_{t+s}\,u_0=S_t\,S_s\,u_0\,,\quad \text{for all } t,s\geq 0\,,\text{ for all } u_0\in\mathbf{Lip}(\mathbb{R}^d).$$

(iii) (Translation) For every constant $c\in\mathbb{R}$ we have that

$$S_t(u_0+c)=S_t\,u_0+c\,,\qquad \text{for all } u_0\in\mathbf{Lip}(\mathbb{R}^d)\,,\text{ for all } t\geqslant 0\,. \tag{2.7.9}$$

(iv) The map $S_t$ is continuous on sets of functions with uniform Lipschitz constant with respect to $\mathbf{W}^{1,1}_{\mathrm{loc}}$-topology, i.e., for every $u_n\in\mathbf{Lip}(\mathbb{R}^d)$ with a Lipschitz constant $M$ such that

$$u_n\;\longrightarrow\;u\quad\text{in } \mathbf{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)\,,$$

we have that $S_t(u_n)$ also converges to $S_t(u)$ in $\mathbf{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ .

CHAPTER

# 3

# METRIC ENTROPY FOR HAMILTON-JACOBI EQUATIONS

This chapter is an in-depth study on how to find metric entropy estimates for the set of viscosity solutions to the Hamilton-Jacobi equation with a coercive and uniformly directionally convex Hamiltonian. Section 3.1 illustrates how to find upper and lower bounds on the minimal number of functions needed to represent a BV function with an accuracy of $\varepsilon$ with respect to $\mathbf{L}^1$–distance. These bounds are then utilized in Section 3.2 to analyze the BV-type regularity for viscosity solutions of the Hamilton-Jacobi equation and derive our desired estimates.

## 3.1 Metric entropy for BV functions

An approach towards quantifying the compactness of BV functions in $\mathbf{L}^1$-space was initialized in [35] which dealt this problem in the scalar case and proved that the $\varepsilon$-entropy of a class of real-valued BV functions in $\mathbf{L}^1$ is of the order $\frac{1}{\varepsilon}$. In this section we refine this conclusion further and provide estimates of the $\varepsilon$-entropy for a class of uniformly bounded total variation functions in $\mathbf{L}^1$-space in multi-dimensional cases. In particular, we show that the minimal number of functions needed to represent a BV function with an error up to $\varepsilon$ with respect to $\mathbf{L}^1$-distance is of the order $\frac{1}{\varepsilon^d}$, where $d$ is the dimension of the space under consideration.

We begin by stating a result on the $\varepsilon$-entropy for a class of bounded total variation functions in the scalar case and prove it using a method similar to that in [10]. Given $L, V, M > 0$, we denote

$$\mathscr{B}_{[L,M,V]} = \left\{ f \in \mathbf{L}^1([0,L],[0,M]) \,\middle|\, |Df|((0,L)) \leq V \right\}. \tag{3.1.1}$$

**Lemma 40.** *For all* $0 < \varepsilon < \frac{L(M+V)}{6}$, *it holds that*

$$\mathscr{H}_\varepsilon\left(\mathscr{B}_{[L,M,V]} \,\middle|\, \mathbf{L}^1([0,L])\right) \leq 8 \cdot \left[\frac{L(M+V)}{\varepsilon}\right]. \tag{3.1.2}$$

*Proof.* For any $f \in \mathscr{B}_{[L,M,V]}$, let $V_f(x)$ be the total variation of $f$ over $[0,x]$. We decompose f as

$$f(x) = f^+(x) - f^-(x) \qquad \text{for all } x \in [0,L]$$

where $f^- = \frac{V_f - f}{2} + \frac{M}{2}$ is a nondecreasing function from $[0,L]$ to $\left[0, \frac{L+M}{2}\right]$ and $f^+ = \frac{V_f + f}{2} + \frac{M}{2}$ is a nondecreasing function from $[0,L]$ to $\left[\frac{M}{2}, \frac{L+2M}{2}\right]$. Defining

$$\mathscr{I} := \left\{ g : [0,L] \to \left[0, \frac{V+M}{2}\right] \,\middle|\, g \text{ is nondecreasing} \right\},$$

we have

$$\mathscr{B}_{[L,M,V]} \subseteq \left(\mathscr{I} + \frac{M}{2}\right) - \mathscr{I} := \left\{ g - h \,\middle|\, g \in \mathscr{I} + \frac{M}{2} \quad \text{and} \quad h \in \mathscr{I} \right\}. \tag{3.1.3}$$

For any $\varepsilon > 0$, it holds that

$$\mathscr{N}_\varepsilon\left(\mathscr{B}_{[L,M,V]} \mid \mathbf{L}^1([0,L])\right) \leq \left[\mathscr{N}_{\frac{\varepsilon}{2}}(\mathscr{I} \mid \mathbf{L}^1([0,L]))\right]^2.$$

Indeed, from Definition 1, there exists a set $\mathscr{G}_{\frac{\varepsilon}{2}}$ of $\mathscr{N}_{\frac{\varepsilon}{2}}(\mathscr{I} \mid \mathbf{L}^1([0,L]))$ subsets of $\mathbf{L}^1([0,L])$ such that

$$\mathscr{I} \subseteq \bigcup_{\mathscr{E} \in \mathscr{G}_{\frac{\varepsilon}{2}}} \mathscr{E} \quad \text{and} \quad \operatorname{diam}(\mathscr{E}) = \sup_{h_1, h_2 \in \mathscr{E}} \|h_1 - h_2\|_{\mathbf{L^1([0,L])}} \leq \varepsilon.$$

Thus, (3.1.3) implies

$$\mathscr{B}_{[L,M,V]} \subseteq \bigcup_{(\mathscr{E}_1,\mathscr{E}_2) \in \mathscr{G}_{\frac{\varepsilon}{2}} \times \mathscr{G}_{\frac{\varepsilon}{2}}} \left[\left(\mathscr{E}_1 + \frac{M}{2}\right) - \mathscr{E}_2\right].$$

For any two functions

$$f_i \;=\; g_i - h_i \;\in \left(\mathscr{E}_1 + \frac{M}{2}\right) - \mathscr{E}_2 \qquad \text{for } i = 1,2\,,$$

we have

$$\begin{aligned} \|f_1 - f_2\|_{\mathbf{L}^1([0,L])} &\leq\; \|g_1 - g_2\|_{\mathbf{L}^1([0,L])} + \|h_1 - h_2\|_{\mathbf{L}^1([0,L])} \\ &\leq\; \mathrm{diam}\left(\mathscr{E}_1 + \frac{M}{2}\right) + \mathrm{diam}(\mathscr{E}_2) \;\leq\; \varepsilon + \varepsilon \;=\; 2\varepsilon \end{aligned}$$

and this implies that

$$\mathrm{diam}\left[\left(\mathscr{E}_1 + \frac{M}{2}\right) - \mathscr{E}_2\right] \;\leq\; 2\varepsilon\,.$$

By Definition 1, we have

$$\mathscr{N}_{\varepsilon}\left(\mathscr{B}_{[L,M,V]}\;\middle|\;\mathbf{L}^1([0,L])\right) \;\leq\; \left[\mathscr{N}_{\frac{\varepsilon}{2}}(\mathscr{I} \mid \mathbf{L}^1([0,L]))\right]^2$$

and thus

$$\mathscr{H}_{\varepsilon}\left(\mathscr{B}_{[L,M,V]}\;\middle|\;\mathbf{L}^1([0,L])\right) \;\leq\; 2\cdot\mathscr{H}_{\frac{\varepsilon}{2}}\left(\mathscr{I}\;\middle|\;\mathbf{L}^1([0,L])\right). \tag{3.1.4}$$

Finally applying [23, Lemma 3.1] in the case of $\mathscr{I}$, for $0 < \varepsilon < \frac{L(M+V)}{6}$, it holds that

$$\mathscr{H}_{\frac{\varepsilon}{2}}\left(\mathscr{I}\;\middle|\;\mathbf{L}^1([0,L])\right) \;\leq\; 4\cdot\left\lfloor\frac{L(M+V)}{\varepsilon}\right\rfloor$$

and then (3.1.4) yields (3.1.2). □

Now we proceed to establish upper and lower estimates of the $\varepsilon$-entropy for a class of uniformly bounded total variation functions,

$$\mathscr{F}_{[L,M,V]} \;=\; \left\{u \in \mathbf{L}^1([0,L]^d,\mathbb{R})\;\middle|\;\|u\|_{\mathbf{L}^\infty([0,L]^d)} \leq M, |Du|((0,L)^d) \leq V\right\}, \tag{3.1.5}$$

in the $\mathbf{L}^1([0,L]^d,\mathbb{R})$-space. Our main result in this section is stated as follows.

**Theorem 41.** *Given $L, M, V > 0$, for every $0 < \varepsilon < \frac{ML^d}{8}$, it holds that*

$$\frac{\log_2(e)}{8}\cdot\left\lfloor\frac{VL}{2^{d+2}\varepsilon}\right\rfloor^d \;\leq\; \mathscr{H}_{\varepsilon}\left(\mathscr{F}_{[L,M,V]}\;\middle|\;\mathbf{L}^1([0,L]^d)\right) \;\leq\; \Gamma_{[d,L,M,V]}\cdot\frac{1}{\varepsilon^d} \tag{3.1.6}$$

*where the constant $\Gamma_{[d,L,M,V]}$ is computed as*

$$\Gamma_{[d,L,M,V]} \;=\; \frac{8}{\sqrt{d}}\left(4\sqrt{n}LV\right)^d + \left(\frac{2^{d+7}V}{M} + 8\right)\cdot\left(\frac{ML^d}{8}\right)^d.$$

*Proof.* *(Upper estimate)* Let us first prove the upper-estimate of $\mathscr{H}_\varepsilon\left(\mathscr{F}_{[L,M,V]}\,\Big|\,\mathbf{L}^1([0,L]^d)\right)$. This proof is divided into several steps.

**1.** For any $N\in\mathbb{N}$, we divide the square $[0,L]^d$ into $N^d$ small squares $\square_\iota$ for $\iota=(\iota_1,\iota_2,\dots,\iota_d)\in\{0,1,\dots,N-1\}^d$ such that

$$\square_\iota = \frac{\iota L}{N}+\left(\left[0,\frac{L}{N}\right]\times\left[0,\frac{L}{N}\right]\times\dots\times\left[0,\frac{L}{N}\right]\right) \qquad \text{and} \qquad \bigcup_{\iota\in\{0,1,2,\dots,N-1\}^d}\square_\iota = [0,L]^d .$$

For any $u\in\mathscr{F}_{[L,M,V]}$, let $u_\iota$ defined as

$$-M \le u_\iota = \frac{1}{\mathrm{Vol}(\square_\iota)}\int_{\square_\iota} u(x)\,dx \le M$$

denote the average value of $u$ in $\square_\iota$ for every $\iota\in\{0,1,2,\dots,N-1\}^d$. Let $\tilde{u}$ be a piecewise constant function on $[0,L]^d$ such that

$$\tilde{u}(x) = \begin{cases} u_\iota & \text{for all } x\in\mathrm{int}(\square_\iota)\\ 0 & \text{for all } x\in\bigcup_{\iota\in\{1,2,\dots,N-1\}^d}\partial\square_\iota \end{cases}.$$

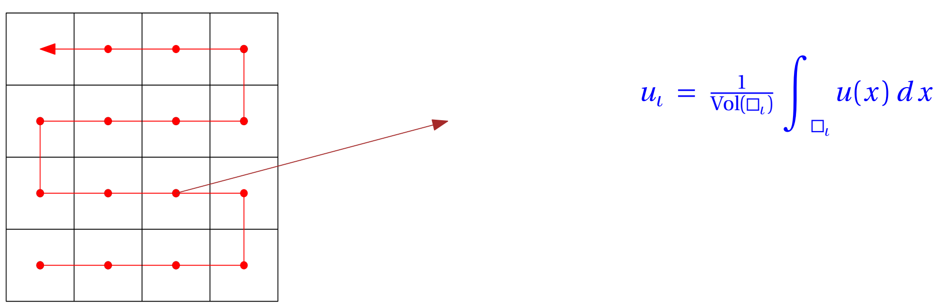


**Figure 3.1** Construction of $\tilde{u}$

By the Poincaré inequality, we have $\int_{\square_\iota}|u(x)-u_\iota|\,dx \le \frac{\mathrm{diam}(\square_\iota)}{2}\cdot|Du|(\mathrm{int}(\square_\iota))$ for all $\iota\in\{0,1,2,\dots,N-1\}^d$. Hence, the $\mathbf{L}^1$-distance between $u$ and $\tilde{u}$ can be estimated as

$$\begin{aligned}\|u-\tilde{u}\|_{\mathbf{L}_1([0,L]^d)} &= \int_{[0,L]^d}|u(x)-\tilde{u}(x)|\,dx = \sum_{\iota\in\{0,1,2,\dots,N-1\}^d}\int_{\square\iota}|u(x)-u_\iota|\,dx\\ &\le \sum_{\iota\in\{0,1,2,\dots,N-1\}^d}\left(\frac{\mathrm{diam}(\mathrm{int}(\square_\iota))}{2}\cdot|Du|(\mathrm{int}(\square_\iota))\right) \le \frac{L\sqrt{d}}{N}\sum_{\iota\in\{0,1,2,\dots,N-1\}^d}|Du|(\mathrm{int}(\square_\iota))\\ &= \frac{L\sqrt{d}}{N}|Du|((0,L)^d) \le \frac{L\sqrt{d}}{N}\cdot V. \qquad (3.1.7)\end{aligned}$$

**2.** Let $e_1, e_2, ..., e_d$ be the standard basis of $\mathbb{R}^d$ where $e_i$ denotes the vector with a 1 in the $i$-th coordinate and 0's elsewhere. For any $\iota \in \{0,1,2,...,N-1\}^d$ and $j \in \{1,2,...,d\}$, we estimate $\left|u_{\iota+e_j} - u_\iota\right|$ in the following way.

$$\begin{aligned}
|u_{\iota+e_j} - u_\iota| &= \left|\frac{1}{\mathrm{Vol}\left(\Box_{\iota+e_j}\right)}\int_{\Box_{\iota+e_j}} u(x)\,dx - \frac{1}{\mathrm{Vol}(\Box_\iota)}\int_{\Box_\iota} u(x)\,dx\right| \\
&= \frac{1}{\mathrm{Vol}(\Box_\iota)}\cdot\left|\int_{\Box_\iota} u\Big(x+\frac{L}{N}\cdot e_j\Big) - u(x)\,dx\right| = \frac{1}{\mathrm{Vol}(\Box_\iota)}\cdot\left|\int_{\Box_\iota}\int_0^{\frac{L}{N}} Du(x+se_j)(e_j)\,ds\,dx\right| \\
&\le \frac{1}{\mathrm{Vol}(\Box_\iota)}\cdot\int_0^{\frac{L}{N}}\left|\int_{\Box_\iota} Du(x+se_j)(e_j)\,dx\right| ds \le \left(\frac{N}{L}\right)^{d-1}\cdot|Du|(\mathrm{int}(\Box_\iota \cup \Box_{\iota+e_j})). \quad (3.1.8)
\end{aligned}$$

Let us rearrange the index set

$$\{0,1,2,\ldots,N-1\}^d = \left\{\kappa^1, \kappa^2, \ldots, \kappa^{N^d}\right\}$$

in a manner such that for all $j \in \{1,...,N^d-1\}$,

$$\kappa^{j+1} = \kappa^j + e_k \qquad \text{for some } k \in \{1,2,...,d\}.$$

From (3.1.8) and (3.1.5), we have

$$\begin{aligned}
\sum_{j=1}^{N^d-1}|u_{\kappa^{j+1}} - u_{\kappa^j}| &\le \left(\frac{N}{L}\right)^{d-1}\cdot\sum_{j=1}^{N^d-1}|Du|(\mathrm{int}(\Box_{\kappa^j}\cup\Box_{\kappa^{j+1}})) \\
&\le 2\left(\frac{N}{L}\right)^{d-1}\cdot|Du|((0,L)^d) \le 2V\left(\frac{N}{L}\right)^{d-1}. \quad (3.1.9)
\end{aligned}$$

To conclude this step, we define the function $f_{u,N}:[0, LN^{d-1}] \to [-M, M]$ associated with $u$ such that

$$f_{u,N}(x) = u_{\kappa^{i+1}} \qquad \text{for all } x \in \left[\frac{i\cdot L}{N}, \frac{(i+1)\cdot L}{N}\right), i \in \left\{0,1,...,N^d-1\right\}.$$

Then recalling (3.1.9), we have

$$|Df_{u,N}|((0, LN^{d-1})) \le 2V\left(\frac{N}{L}\right)^{d-1}. \qquad (3.1.10)$$

**3.** Let us denote

$$L_N := L\cdot N^{d-1}, \qquad \beta_N := 2V\left(\frac{N}{L}\right)^{d-1}. \qquad (3.1.11)$$

We introduce the set

$$\tilde{\mathscr{F}}_N = \Big\{ f:[0,L_N]\to[-M,M] \,\Big|\, |Df|((0,L_N)) \le \beta_N \text{ and}$$
$$f(x)=f\left(\frac{i\cdot L}{N}\right) \quad \text{for all } x\in\left[\frac{i\cdot L}{N},\frac{(i+1)\cdot L}{N}\right), i\in\{0,1,2,...,N^d-1\}\Big\}.$$

(3.1.10) implies

$$f_{u,N} \in \tilde{\mathscr{F}}_N \qquad \text{for all } u\in\mathscr{F}_{[L,M,V]}.$$

On the other hand, recalling that

$$\mathscr{B}_{[L_N,2M,\beta_N]} = \left\{ f\in \mathbf{L}^1([0,L_N],[0,2M]) \,\middle|\, |Df|((0,L_N))\le\beta_N \right\},$$

we have

$$\tilde{\mathscr{F}}_N \subset \mathscr{B}_{[L_N,2M,\beta_N]} - M.$$

From Lemma 40, for every $0<\varepsilon'<\frac{L_N\cdot(\beta_N+2M)}{6}$, it holds that

$$\mathscr{H}_{\varepsilon'}\left(\mathscr{B}_{[L_N,2M,\beta_N]} \,\middle|\, \mathbf{L}^1([0,L_N])\right) \le 8\cdot\left\lfloor\frac{L_N(\beta_N+2M)}{\varepsilon'}\right\rfloor,$$

which yields

$$\mathscr{H}_{\varepsilon'}\left(\tilde{\mathscr{F}}_N \,\middle|\, \mathbf{L}^1([0,L_N])\right) \le 8\cdot\left\lfloor\frac{L_N(\beta_N+2M)}{\varepsilon'}\right\rfloor.$$

By Definition 1, there exists a set of $\Gamma_{N,\varepsilon'}=2^{8\cdot\left\lfloor\frac{L_N(\beta_N+2M)}{\varepsilon'}\right\rfloor}$ functions in $\tilde{\mathscr{F}}_N$,

$$\mathscr{G}_{N,\varepsilon'} = \left\{g_1,g_2,\dots,g_{\Gamma_{N,\varepsilon'}}\right\} \subset \tilde{\mathscr{F}}_N,$$

such that

$$\tilde{\mathscr{F}}_N \subset \bigcup_{i=1}^{\Gamma_{N,\varepsilon'}} B(g_i,2\varepsilon').$$

So for every $u\in\mathscr{F}_{[L,M,V]}$, for its corresponding $f_{u,N}$, there exists $g_{j_u}\in\mathscr{G}_{N,\varepsilon'}$ such that

$$\|f_{u,N}-g_{j_u}\|_{\mathbf{L}^1([0,L_N])} \le 2\varepsilon'.$$

Let $\mathscr{U}_{N,\varepsilon'}$ be a set of $\Gamma_{N,\varepsilon'}$ functions $u_j^\dagger:[0,L]^N\to[-M,M]$ defined as follows.

$$u_j^\dagger = \begin{cases} 0 & \text{if} \quad x\in\bigcup_{\iota\in\{1,2,\dots,N\}^d}\partial\Box_\iota \\ g_j\left(\frac{(i-1)\cdot L}{N}\right) & \text{if} \quad x\in\operatorname{int}(\Box_{\kappa^i}), i\in\{1,2,\dots,N^d\} \end{cases}$$

for $j \in \{1,2,\ldots,d^N\}$. Then corresponding to every $u \in \mathscr{F}_{[L,M,V]}$, there exists $u^\dagger_{j_u} \in \mathscr{U}_{N,\varepsilon'}$ for some $j_u \in \{1,2,\ldots,\Gamma_{N,\varepsilon}\}$ such that

$$\begin{aligned}\left\|\tilde{u}-u^\dagger_{j_u}\right\|_{\mathbf{L}^1([0,L]^d)} &= \sum_{i=1}^{N^d}\left|u_{\kappa^i}-g_{j_u}\left(\frac{(i-1)\cdot L}{N}\right)\right|\cdot \mathrm{Vol}(\square_{\kappa^i})\\ &= \sum_{i=1}^{N^d}\left|f_{u,N}\left(\frac{(i-1)\cdot L}{N}\right)-g_{j_u}\left(\frac{(i-1)\cdot L}{N}\right)\right|\cdot\frac{L}{N}\cdot\frac{L^{d-1}}{N^{d-1}}\\ &= \frac{L^{d-1}}{N^{d-1}}\cdot\|f_{u,N}-g_{j_u}\|_{\mathbf{L}^1([0,L_N])} \le 2\varepsilon'\cdot\frac{L^{d-1}}{N^{d-1}}.\end{aligned}$$

Combining this with (3.1.7), we obtain

$$\left\|u-u^\dagger_{j_u}\right\|_{\mathbf{L}^1([0,L]^d)} \le \left\|u-\tilde{u}\right\|_{\mathbf{L}^1([0,L]^d)}+\left\|\tilde{u}-u_{g_{j_u}}\right\|_{\mathbf{L}^1([0,L]^d)} \le 2\varepsilon'\cdot\frac{L^{d-1}}{N^{d-1}}+\frac{L\sqrt{d}}{N}\cdot V. \quad (3.1.12)$$

**4.** For any $\varepsilon>0$, we choose

$$N = \left\lfloor\frac{2\sqrt{d}LV}{\varepsilon}\right\rfloor+1 \qquad \text{and} \qquad \varepsilon' = \frac{N^{d-1}\cdot\varepsilon}{4L^{d-1}} \quad (3.1.13)$$

such that

$$\left\|u-u^\dagger\right\|_{\mathbf{L}^1([0,L]^d)} \le 2\varepsilon'\cdot\frac{L^{d-1}}{N^{d-1}}+\frac{L\sqrt{d}}{N}\cdot V \le \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon$$

for all $u \in \mathscr{F}_{[L,M,V]}$ and for some $u^\dagger \in \mathscr{U}_{N,\varepsilon'}$. From the previous step, it holds that

$$\mathscr{F}_{[L,M,V]} \subseteq \bigcup_{u^\dagger\in\mathscr{U}_{N,\varepsilon'}} \overline{B}(u^\dagger,\varepsilon)$$

provided we have

$$\varepsilon' = \frac{N^{d-1}\varepsilon}{4L^{d-1}} \le \frac{L_N\cdot(\beta_N+2M)}{6} = \frac{N^{d-1}(VN^{d-1}+ML^{d-1})}{3L^{d-2}}. \quad (3.1.14)$$

This condition is equivalent to

$$\varepsilon \le \frac{4}{3}\cdot\left(LVN^{d-1}+ML^d\right).$$

(3.1.13) implies that the condition (3.1.14) holds if

$$\varepsilon \le \frac{4}{3}\cdot\left(\frac{2^{d-1}d^{\frac{d-1}{2}}L^dV^d}{\varepsilon^{d-1}}+ML^d\right). \quad (3.1.15)$$

Assuming $0<\varepsilon<\frac{2ML^d}{3}+d^{\frac{d-1}{2d}}LV$, we claim that (3.1.14) holds. Indeed, if $\frac{2ML^d}{3}>d^{\frac{d-1}{2d}}LV$ then

$$\varepsilon \ < \ \frac{2ML^d}{3}+d^{\frac{d-1}{2d}}LV \ \le \ \frac{4ML^d}{3},$$

which results in (3.1.15). Otherwise, we have that $\varepsilon<\frac{2ML^d}{3}+d^{\frac{d-1}{2d}}LV\le 2d^{\frac{d-1}{2d}}LV$. Thus

$$\begin{aligned}\frac{4}{3}\cdot\left(\frac{2^{d-1}d^{\frac{d-1}{2}}L^dV^d}{\varepsilon^{d-1}}+ML^d\right) \ &\ge \ \frac{4}{3}\cdot\frac{2^{d-1}d^{\frac{d-1}{2}}L^dV^d}{2^{d-1}d^{\frac{(d-1)^2}{2d}}L^{d-1}V^{d-1}}+\frac{4}{3}ML^d\\ &= \ \frac{4}{3}\cdot d^{\frac{d-1}{2d}}LV+\frac{4}{3}ML^d.\end{aligned}$$

and this implies (3.1.15).

To complete the proof, recalling (3.1.11) and (3.1.13), we estimate that

$$\begin{aligned}\mathrm{Card}(\mathscr{U}_{N,\varepsilon'}) \ &= \ \Gamma_{N,\varepsilon'} \ = \ 2^{8\cdot\left\lfloor\frac{L_N(\beta_N+2M)}{\varepsilon'}\right\rfloor} \ = \ 2^{8\cdot\left\lfloor\frac{8}{\varepsilon}\cdot\left(LVN^{d-1}+ML^d\right)\right\rfloor}\\ &\le \ 2^{\frac{64}{\varepsilon}\cdot\left(LV\left(\left\lfloor\frac{2\sqrt{d}LV}{\varepsilon}\right\rfloor+1\right)^{d-1}+ML^d\right)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathscr{H}_{\varepsilon}\left(\mathscr{F}_{[L,M,V]}\ \Big|\ \mathbf{L}^1([0,L]^d)\right) \ &\le \ \frac{64}{\varepsilon}\cdot\left(LV\left(\left\lfloor\frac{2\sqrt{d}LV}{\varepsilon}\right\rfloor+1\right)^{d-1}+ML^d\right)\\ &\le \ \frac{64}{\varepsilon}\cdot\left(LV\left(\frac{2^{2d-3}d^{\frac{d-1}{2}}L^{d-1}V^{d-1}}{\varepsilon^{d-1}}+2^{d-2}\right)+ML^d\right)\\ &= \ \frac{2^{2d+3}d^{\frac{d-1}{2}}L^dV^d}{\varepsilon^d}+\frac{2^{d+4}LV+ML^d}{\varepsilon}. \qquad (3.1.16)\end{aligned}$$

In particular, if $0<\varepsilon<\frac{ML^d}{8}$ then

$$\mathscr{H}_{\varepsilon}\left(\mathscr{F}_{[L,M,V]}\ \Big|\ \mathbf{L}^1([0,L]^d)\right) \ \le \ \left[2^{2d+3}d^{\frac{d-1}{2}}L^dV^d+\left(2^{d+4}LV+ML^d\right)\cdot\left(\frac{ML^d}{8}\right)^{d-1}\right]\cdot\frac{1}{\varepsilon^d}$$

and this yields the right hand side of (3.2.27).

*(Lower estimate)* We are now going to prove the lower estimate of $\mathscr{H}_{\varepsilon}\left(\mathscr{F}_{[L,M,V]}\ \Big|\ \mathbf{L}^1([0,L]^d)\right)$.

**1.** Again given any $N\in\mathbb{N}$, we divide the square $[0,L]^d$ into $N^d$ small squares $\square_\iota$, for $\iota=(\iota_1,\iota_2,\dots,\iota_d)\in\{0,1,\dots,N-1\}^d$ such that

$$\square_\iota = \frac{\iota L}{N} + \left(\left[0, \frac{L}{N}\right] \times \left[0, \frac{L}{N}\right] \times \ldots \times \left[0, \frac{L}{N}\right]\right) \quad \text{and} \quad \bigcup_{\iota \in \{0,1,2,\ldots,N-1\}^d} \square_\iota = [0, L]^d .$$

Consider the set of $N^d$-tuples

$$\Delta_N = \left\{ \delta = (\delta_\iota)_{\iota \in \{0,1,\ldots,N-1\}^d} \,\middle|\, \delta_\iota \in \{0,1\} \right\} .$$

Given any $h > 0$, for any $\delta \in \Delta_N$, we define the function $u_\delta : [0, L]^d \to \{0, h\}$ such that

$$u_\delta(x) = \sum_{\iota \in \{0,1,\ldots,N-1\}^d} h\delta_\iota \cdot \chi_{\mathrm{int}(\square_\iota)}(x) \qquad \text{for all } x \in [0, L]^d .$$

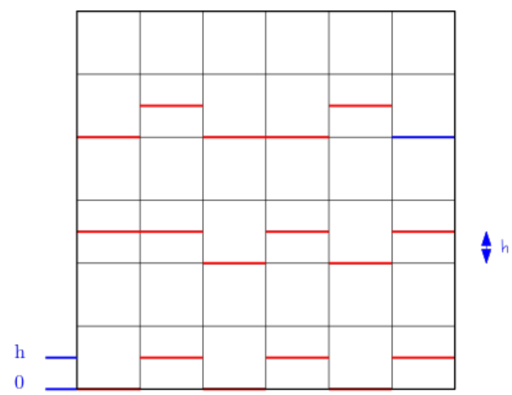


**Figure 3.2** Example of $u_\delta$

This directly implies $u_\delta \in BV((0, L)^d, \mathbb{R}^d)$ and

$$|Du_\delta|((0, L)^d) \leq \sum_{\iota \in \{0,1,\ldots,N-1\}^d} |Du_\delta|(\square_\iota) \leq 2^{d-1}\left(\frac{L}{N}\right)^{d-1} N^n h = (2L)^{d-1} N h .$$

Assuming that

$$0 < h \leq \min\left\{M , \frac{V}{2^{d-1} L^{d-1} N}\right\} , \tag{3.1.17}$$

we have

$$|Du_\delta|((0, L)^d) \leq (2L)^{d-1} N \cdot \frac{V}{2^{d-1} L^{d-1} N} = V \qquad \text{for all } \delta \in \Delta_N$$

and this implies

$$\mathcal{G}_{h,N} := \{u_\delta \mid \delta \in \Delta_N\} \subset \mathcal{F}_{[L,M,V]} \qquad \text{for all } N \in \mathbb{N} .$$

Hence,

$$\mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,M,V]} \,\middle|\, \mathbf{L}^1([0, L]^d)\right) \geq \mathcal{N}_\varepsilon\left(\mathcal{G}_{h,N} \,\middle|\, \mathbf{L}^1([0, L]^d)\right) \qquad \text{for all } \varepsilon > 0 . \tag{3.1.18}$$

Towards an estimate of the covering number $\mathcal{N}_\varepsilon\left(\mathscr{G}_{h,N} \,\middle|\, \mathbf{L}^1([0,L]^d)\right)$, for a fixed $\tilde{\delta}\in\Delta_N$, we can define

$$\mathscr{I}_{\tilde{\delta},N}(2\varepsilon) \;=\; \left\{\delta\in\Delta_N \,\middle|\, \|u_\delta - u_{\tilde{\delta}}\|_{\mathbf{L}^1([0,L]^d)} \;\le\; 2\varepsilon\right\} \qquad \text{and} \qquad C_N(2\varepsilon) \;=\; \mathrm{Card}(\mathscr{I}_{\tilde{\delta},N}(2\varepsilon)) \tag{3.1.19}$$

since the cardinality of the set $\mathscr{I}_{\tilde{\delta},N}(\varepsilon)$ is independent of the choice of $\tilde{\delta}\in\Delta_N$. We observe that an $\varepsilon$-cover of $\mathscr{G}_{h,N}$ in $\mathbf{L}^1$ contains at most $C_N(2\varepsilon)$ elements. Since $\mathrm{Card}(\mathscr{G}_{h,N}) = \mathrm{Card}(\Delta_N) = 2^{N^d}$, it holds that

$$\mathcal{N}_\varepsilon\left(\mathscr{G}_{h,N} \,\middle|\, \mathbf{L}^1([0,L]^d)\right) \;\ge\; \frac{2^{N^d}}{C_N(2\varepsilon)}. \tag{3.1.20}$$

**2.** Now we need to provide an upper bound on $C_N(2\varepsilon)$. For any given pair $\delta,\tilde{\delta}\in\Delta_N$,

$$\|u_\delta - u_{\tilde{\delta}}\|_{\mathbf{L}^1([0,L]^d)} \;=\; \sum_{\iota\in\{0,1,\dots,N\}^d} \|u_\delta - u_{\tilde{\delta}}\|_{\mathbf{L}^1(\square_\iota)} \;=\; \eta(\delta,\tilde{\delta})\cdot\frac{hL^n}{N^n}$$

where $\eta(\delta,\tilde{\delta}) := \mathrm{Card}\big(\{\iota\in\{0,1,\dots,N-1\}^d \mid \delta_\iota \ne \tilde{\delta}_\iota\}\big)$. From (3.1.19), we obtain

$$\mathscr{I}_{\tilde{\delta},N}(2\varepsilon) \;=\; \left\{\delta\in\Delta_N \,\middle|\, \eta(\delta,\tilde{\delta}) \;\le\; \frac{2\varepsilon N^d}{hL^d}\right\}$$

and it yields

$$C_N(2\varepsilon) \;=\; \mathrm{Card}\big(\mathscr{I}_{\tilde{\delta},N}(2\varepsilon)\big) \;\le\; \sum_{r=0}^{\left\lfloor \frac{2\varepsilon N^d}{hL^d}\right\rfloor} \binom{N^d}{r}.$$

To estimate the last term in the above inequality, let us consider $N^d$ independent random variables with uniform Bernoulli distribution $X_1, X_2, \dots, X_{N^d}$, i.e.,

$$\mathbb{P}(X_i = 1) \;=\; \mathbb{P}(X_i = 0) \;=\; \frac{1}{2} \qquad \text{for all } i\in\{1,2,\dots,N^d\}.$$

Setting $S_{N^d} := X_1 + X_2 + \cdots + X_{N^d}$, we note that for any $k \le N^d$,

$$\sum_{r=1}^{k}\binom{N^d}{r} \;=\; 2^{N^d}\cdot\mathbb{P}(S_{N^d}\le k).$$

By Hoeffding's inequality from [32, Theorem 2], for all $\mu \le \frac{N^d}{2}$,

$$\mathbb{P}\big(S_{N^d} \le \mathbb{E}[S_{N^d}] - \mu\big) \;=\; \mathbb{P}\left(S_{N^d} \le \frac{N^d}{2} - \mu\right) \;\le\; e^{\left(-\frac{2\mu^2}{N^d}\right)}$$

where $\mathbb{E}[S_{N^d}]$ is the expectation of $S_{N^d}$. Hence, for every $0<\varepsilon\leq\frac{hL^d}{8}$, it holds that $\frac{2\varepsilon N^d}{hL^d}\leq\frac{N^d}{2}$ and $\frac{4\varepsilon}{hL^d}\leq\frac{1}{2}$. Thus we have

$$C_N(2\varepsilon)\;\leq\;\sum_{r=0}^{\left\lfloor\frac{2\varepsilon N^d}{hL^d}\right\rfloor}\binom{N^d}{r}\;=\;2^{N^d}\cdot\mathbb{P}\left(S_{N^d}\leq\left\lfloor\frac{2\varepsilon N^d}{hL^d}\right\rfloor\right)\;\leq\;2^{N^d}\cdot e^{\left(-\frac{2\left(\frac{N^d}{2}-\left\lfloor\frac{2\varepsilon N^d}{hL^d}\right\rfloor\right)^2}{N^d}\right)}$$

$$\leq\;2^{N^d}\cdot e^{\left(-\frac{\left(N^d-\frac{4\varepsilon N^d}{hL^d}\right)^2}{2N^d}\right)}\;=\;2^{N^d}\cdot e^{\left(-N^d\cdot\frac{\left(1-\frac{4\varepsilon}{hL^d}\right)^2}{2}\right)}\;\leq\;2^{N^d}\cdot e^{-N^d/8}.$$

From (3.1.20) and (3.1.17),

$$\mathcal{N}_\varepsilon\left(\mathcal{G}_{h,N}\;\middle|\;\mathbf{L}^1([0,L]^d)\right)\;\geq\;\frac{2^{N^d}}{C_N(2\varepsilon)}\;\geq\;e^{\frac{N^d}{8}}$$

provided

$$0\;<\;h\;\leq\;\min\left\{M\;,\;\frac{V}{2^{d-1}L^{d-1}N}\right\}\qquad\text{and}\qquad 0<\varepsilon\;\leq\;\frac{hL^d}{8}.\tag{3.1.21}$$

Therefore, for every $0<\varepsilon<\frac{ML^d}{8}$, by choosing

$$h\;=\;\min\left\{M\;,\;\frac{V}{2^{d-1}L^{d-1}N}\right\}\qquad\text{and}\qquad N\;:=\;\left\lfloor\frac{VL}{2^{d+2}\varepsilon}\right\rfloor$$

such that (3.1.21) holds, we obtain that

$$\mathcal{N}_\varepsilon\left(\mathcal{G}_{h,N}\;\middle|\;\mathbf{L}^1([0,L]^d)\right)\;\geq\;e^{\left(\frac{1}{8}\cdot\left\lfloor\frac{VL}{2^{d+2}\varepsilon}\right\rfloor^d\right)}.$$

Finally recalling (3.1.18), we have

$$\mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,M,V]}\;\middle|\;\mathbf{L}^1([0,L]^d)\right)\;\geq\;e^{\left(\frac{1}{8}\cdot\left\lfloor\frac{VL}{2^{d+2}\varepsilon}\right\rfloor^d\right)}$$

and this implies the first inequality in (3.2.27). □

**Remark 42.** The upper estimate of $\mathcal{H}_\varepsilon\left(\mathcal{F}_{[L,M,V]}\;\middle|\;\mathbf{L}^1([0,L]^d)\right)$ can be improved. Indeed, from (3.1.16), for every $0<\varepsilon<\frac{4}{3}\cdot\left(\frac{2^{d-1}d^{\frac{d-1}{2}}L^dV^d}{\varepsilon^{d-1}}+ML^d\right)$, it holds that

$$\mathcal{H}_\varepsilon\left(\mathcal{F}_{[L,M,V]}\;\middle|\;\mathbf{L}^1([0,L]^d)\right)\;\leq\;\frac{2^{2d+3}d^{\frac{d-1}{2}}L^dV^d}{\varepsilon^d}+\frac{2^{d+4}LV+ML^d}{\varepsilon}.$$

## 3.2 Application to Hamilton-Jacobi equation

Let us consider a first-order Hamilton-Jacobi equation

$$u_t(t,x)+H\big(D_x u(t,x)\big) = 0 \qquad \text{for all } (t,x)\in(0,\infty)\times\mathbb{R}^d \tag{3.2.1}$$

where $u:[0,+\infty)\times\mathbb{R}^d\to\mathbb{R}^d$, $D_x u=(u_{x_1},\dots,u_{x_d})$ and $H:\mathbb{R}^d\to\mathbb{R}$ is a Hamiltonian. Under standard assumptions of convexity and coercivity on the Hamiltonian $H$, for every Lipschitz initial data $u_0\in\mathbf{Lip}(\mathbb{R}^d)$, the corresponding unique viscosity solution of equation (3.2.1) with $u(0,x)=u_0(x)$ is computed by the Hopf-Lax representation formula

$$u(t,x)=S_t(u_0)(x) = \min_{y\in\mathbb{R}^d}\left\{u_0(y)+t\cdot L\left(\frac{x-y}{t}\right)\right\} \tag{3.2.2}$$

where $L$ is the Legendre transform of $H$. In addition, if $H$ is strongly convex, i.e., there exists a constant $\lambda>0$ such that

$$D^2H(p) \geqslant \lambda\cdot\mathbf{I}_d \qquad \text{for all } p\in\mathbb{R}^d,$$

then the map $x\to u(t,x)-\dfrac{1}{2\lambda t}\cdot\|x\|^2$ is concave for every $t>0$. In particular, $u(t,\cdot)$ is twice differentiable almost everywhere and $D_x u(t,\cdot)$ has locally bounded total variation.

The first results on the $\varepsilon$-entropy for sets of viscosity solutions of (3.2.1) were obtained in [3]. The authors proved that the minimal number of bits needed to represent a viscosity solution of (3.2.1) up to an accuracy $\varepsilon$ with respect to the $\mathbf{W^{1,1}}$-distance is of the order $\varepsilon^{-d}$ under the strongly convex condition on Hamiltonian $H$. A similar result was also proved in [4] by the same authors, for the case when $H$ depends on the state variable $x$. There the main idea was to provide controllability results for Hamilton-Jacobi equations and a compactness result for a class of semiconcave functions. However, such a gain of BV regularity does not hold for (3.2.1) with general strictly convex Hamiltonian functions and the previous approach in [3, 4] to finding the $\varepsilon$-entropy of the solution set cannot be applied.

Presently, we have extended the analysis of the metric entropy for sets of viscosity solutions to (3.2.1) when the Hamiltonian $H\in\mathscr{C}^1\big(\mathbb{R}^d\big)$ is strictly convex, coercive and in the form of a uniformly directionally convex function, i.e., for every constant $R>0$ it holds that

$$\inf_{p\neq q\in\overline{B}(0,R)}\left\langle\frac{DH(p)-DH(q)}{|DH(p)-DH(q)|},\frac{p-q}{|p-q|}\right\rangle := \lambda_R > 0. \tag{3.2.3}$$

By the Hopf-Lax representation formula (3.2.2), it is well-known from [18] that the set of slopes of backward optimal rays through $(t,x)$, denoted by

$$\mathbf{b}(t,x)=\left\{\frac{x-y}{t}: u(t,x)=u_0(y)+t\cdot L\left(\frac{x-y}{T}\right)\right\}, \tag{3.2.4}$$

reduces to a singleton $\mathbf{b}(t,x)=DH(D_x u(t,x))$ for almost every $(t,x)\in\mathbb{R}_+\times\mathbb{R}^d$. Moreover, if $M$ is a Lipschitz constant of $u_0$ then for all $t>0$, $\mathbf{b}(t,\cdot)$ can be viewed as an element in $\mathbf{L}^\infty(\mathbb{R}^d)$ with

$$||\mathbf{b}(t,\cdot)||_{\mathbf{L}^\infty(\mathbb{R}^d)} \le \Lambda_M := \max\{|q|: L(q)\le M|q|\}. \tag{3.2.5}$$

Towards the sharp estimate on $\varepsilon$-entropy of the semigroup $S_t$ for all $t>0$, we first establish a BV bound on $\mathbf{b}(t,\cdot)$.

**Theorem 43.** *Assume that $H\in\mathscr{C}^1\left(\mathbb{R}^d\right)$ is strictly convex, coercive and satisfies (3.2.3). For every $t>0$ and $u_0\in\mathbf{Lip}(\mathbb{R}^d)$ with a Lipschitz constant $M$, the function $\mathbf{b}(t,\cdot)$ has locally bounded variation and for every open and bounded set $\Omega\subset\mathbb{R}^d$ of finite perimeter*

$$|D\mathbf{b}(t,\cdot)|(\Omega) \le \frac{1}{\gamma_M}\cdot\left(\Lambda_M+\frac{\operatorname{diam}(\Omega)}{t}\right)\cdot\mathscr{H}^{d-1}(\partial\Omega)+\frac{\sqrt{d}}{t}\cdot|\Omega| \tag{3.2.6}$$

*with* $\gamma_M := \lambda_{\left(\max_{|q|\le\Lambda_M}|DL(q)|\right)}$.

### 3.2.1 BV bound on $\mathbf{b}(t,\cdot)$

In this subsection, we shall establish a BV bound on $\mathbf{b}(t,\cdot)$ as stated in Theorem 43 for a given $t>0$. Here, we recall (3.2.4) and note that $u_0$ is a Lipschitz function with a Lipschitz constant $M$. For any $n\ge 1$, setting

$$\mathscr{Z}_n = \frac{2^{-n+1}}{\sqrt{d}}\cdot\mathbb{Z}^d = \{y_1,y_2,\dots,y_k,\dots\},$$

we approximate the solution $u$ by a monotone decreasing sequence of continuous functions $u_n:(0,+\infty)\times\mathbb{R}^d\to\mathbb{R}$ defined by

$$u_n(t,x) := \min_{y\in\mathscr{Z}_n}\left\{(1-\varepsilon_n)\cdot u_0(y)+t\cdot L\left(\frac{x-y}{t}\right)\right\} \tag{3.2.7}$$

with

$$\varepsilon_n = \frac{1}{M\cdot\Lambda_M}\cdot\max_{|q|\le\frac{2-n}{t}}\left[M|q|+L(q)\right] \qquad \text{and} \qquad \Lambda_M = \max\{|q|: L(q)\le M|q|\}. \tag{3.2.8}$$

Let $C^n_{t,x}$ be the set of optimal points $y\in\mathscr{Z}_n$ such that (3.2.7) holds, i.e.,

$$C^n_{t,x} := \operatorname{argmin}_{y\in\mathscr{Z}_n}\left\{(1-\varepsilon_n)\cdot u_0(y)+t\cdot L\left(\frac{x-y}{t}\right)\right\}.$$

We define the multivalued-function $\mathbf{b}_n : (0,\infty)\times\mathbb{R}^d \to \mathbb{R}^d$ by

$$\mathbf{b}_n(t,x) = \left\{\frac{x-y}{t} : y\in C^n_{t,x}\right\} \qquad \text{for all } (t,x)\in(0,\infty)\times\mathbb{R}^d. \tag{3.2.9}$$

**1.** For any $x\in\mathbb{R}^d$ and $y^n_x\in C^n_{t,x}$, it holds that

$$\begin{aligned} u_n(t,x')-u_n(t,x) &\leq t\cdot\left[L\left(\frac{x'-y^n_x}{t}\right)-L\left(\frac{x-y^n_x}{t}\right)\right] \\ &\leq DL\left(\frac{x-y^n_x}{t}\right)\cdot(x'-x)+O(|x'-x|) \end{aligned}$$

for all $x'\in\mathbb{R}^d$. In particular, we have

$$\mathbf{b}_n(t,x) \subseteq DH(D^+u_n(t,x))$$

and the set of points where $\mathbf{b}_n(t,\cdot)$ is not singleton,

$$\Sigma^n_t = \{x \mid \mathrm{Card}(\{\mathbf{b}_n(t,x)\})\geq 2\},$$

is $\mathscr{H}^{n-1}$-rectifiable. Moreover, $\{\mathbf{b}_n(t,\cdot)\}_{n\geq 1}$ is a bounded sequence in $[\mathbf{L}^\infty(\mathbb{R}^d)]^d$.

Indeed, for any given $(x,y^n_x)\in\mathbb{R}^d\times C^n_{t,x}$, let $\bar{x}\in\mathscr{Z}_n$ be the closest point to $x$ such that $|x-\bar{x}|\leq 2^{-n}$. Using the Lipschitz estimate on $u_0$, we obtain

$$\begin{aligned} L\left(\frac{x-y^n_x}{t}\right) &\leq \frac{1-\varepsilon_n}{t}\cdot\big(u_0(\bar{x})-u_0(y^n_x)\big)+L\left(\frac{x-\bar{x}}{t}\right) \\ &\leq \frac{(1-\varepsilon_n)\cdot M}{t}\cdot\big(|x-y^n_x|+2^{-n}\big)+L\left(\frac{x-\bar{x}}{t}\right) \\ &\leq (1-\varepsilon_n)\cdot M\cdot\left|\frac{x-y^n_x}{t}\right|+\max_{|q|\leq\frac{2^{-n}}{t}}\big[M|q|+L(q)\big]. \end{aligned}$$

Thus, (3.2.8) implies that

$$L(\mathbf{b}_n(t,x))-M\cdot|\mathbf{b}_n(t,x)| \leq \varepsilon_n\cdot M\cdot(\Lambda_M-|\mathbf{b}_n(t,x)|)$$

and this yields

$$\|\mathbf{b}_n\|_{\mathbf{L}^\infty(\mathbb{R}^d)} \leq \Lambda_M \qquad \text{for all } n\geq 1. \tag{3.2.10}$$

On the other hand, let $z_x\in\mathscr{Z}_n$ be the closest point to $y_x\in C_{t,x}$ such that $|z_x-y_x|\leq 2^{-n}$.

Since $|x-y_x|/t \le \Lambda_M$, we estimate

$$\begin{aligned}
|u_n(t,x)-u(t,x)| &= \min_{y\in\mathscr{Z}_n}\left\{(1-\varepsilon_n)\cdot u_0(y)+t\cdot L\left(\frac{x-y}{t}\right)\right\}-u_0(y_x)-t\cdot L\left(\frac{x-y_x}{t}\right)\\
&\le (1-\varepsilon_n)\cdot u_0(z_x)-u_0(y_x)+t\cdot\left[L\left(\frac{x-z_x}{t}\right)-L\left(\frac{x-y_x}{t}\right)\right]\\
&\le \left((1-\varepsilon_n)\cdot M+\sup_{|q|\le\Lambda_M+\frac{2^{-n}}{t}}|DL(q)|\right)\cdot 2^{-n}+\varepsilon_n\cdot|u_0(y_x)|\\
&\le \left(M+\sup_{|q|\le\Lambda_M+\frac{2^{-n}}{t}}|DL(q)|\right)\cdot 2^{-n}+(|u_0(0)|+M|x|+M\Lambda_M t)\cdot\varepsilon_n.
\end{aligned}$$

In particular, $u_n(t,\cdot)$ converges uniformly to $u(t,\cdot)$ in any compact subset of $\mathbb{R}^d$.

**2.** Fixing $n\in\mathbb{Z}^+$, for any $i\neq j\ge 1$, the set

$$\mathscr{O}^t_{i,j} = \left\{x\in\mathbb{R}^d : (1-\varepsilon_n)u_0(y_i)+t\cdot L\left(\frac{x-y_i}{t}\right) < (1-\varepsilon_n)u_0(y_j)+t\cdot L\left(\frac{x-y_j}{t}\right)\right\}$$

is an open subset of $\mathbb{R}^d$ with $\mathscr{C}^1$-boundary

$$\Gamma^t_{i,j} : \left\{x\in\mathbb{R}^d : (1-\varepsilon_n)u_0(y_i)+t\cdot L\left(\frac{x-y_i}{t}\right) = (1-\varepsilon_n)u_0(y_j)+t\cdot L\left(\frac{x-y_j}{t}\right)\right\}.$$

Set $\mathscr{V}^t_i := \bigcup_{j\neq i}\mathscr{O}^t_{i,j}$. From (3.2.7) and (3.2.9), it holds that

$$\mathbf{b}_n(t,x) = \frac{x-y_i}{t} \qquad \text{for all } x\in\mathscr{V}^t_i. \tag{3.2.11}$$

In particular, $\mathbf{b}_n(t,\cdot)$ is in $BV_{loc}(\mathbb{R}^d)$ with

$$D\mathbf{b}_n(t,\cdot) = \frac{\mathbf{I}_d}{t}\cdot\mathscr{L}^d+\sum_{i=1}^{p(\Omega)}\mathbf{b}_n(t,\cdot)\otimes\nu_i\mathscr{H}^{d-1}_{\llcorner\Omega\bigcap\partial\mathscr{V}^t_i}$$

and

$$\operatorname{div}\mathbf{b}_n(t,\cdot) = \frac{d}{t}\cdot\mathscr{L}^d+\sum_{i=1}^{p(\Omega)}\langle\mathbf{b}_n(t,\cdot),\nu_i\rangle\mathscr{H}^{d-1}_{\llcorner\Omega\bigcap\partial\mathscr{V}^t_i}$$

where $\nu_i$ is the inner normal vector to $\mathscr{V}^t_i$.

**Proposition 44.** *For every $t>0$ and for every open and bounded set $\Omega\subset\mathbb{R}^d$ of finite perimeter, setting $\gamma_M := \lambda_{\left(\max_{|q|\le\Lambda_M}|DL(q)|\right)}$, it holds that*

$$|D\mathbf{b}_n(t,\cdot)|(\Omega) \le \frac{1}{\gamma_M}\cdot\left(\|\mathbf{b}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)}+\frac{\operatorname{diam}(\Omega)}{t}\right)\cdot\mathscr{H}^{d-1}(\partial\Omega)+\frac{\sqrt{d}}{t}\cdot|\Omega|. \tag{3.2.12}$$

*Proof.* The proof is divided in to the following steps.

**1.** First we rewrite

$$D\mathbf{b}_n(t,\cdot) = \frac{\mathbf{I}_d}{t}\cdot\mathscr{H}^d + \frac{1}{2t}\cdot\sum_{i\neq j\in\{1,\cdots,p(\Omega)\}}(y_j-y_i)\otimes\nu_i\mathscr{H}^{d-1}_{\llcorner\Omega\bigcap\partial\mathscr{V}_i^t\bigcap\partial\mathscr{V}_j^t}$$

and

$$\operatorname{div}\mathbf{b}_n(t,\cdot) = \frac{d}{t}\cdot\mathscr{H}^d + \frac{1}{2t}\cdot\sum_{i\neq j\in\{1,\cdots,p(\Omega)\}}\left\langle y_j-y_i,\nu_i\right\rangle\mathscr{H}^{d-1}_{\llcorner\Omega\bigcap\partial\mathscr{V}_i^t\bigcap\partial\mathscr{V}_j^t} \tag{3.2.13}$$

with

$$\nu_i(x) = \frac{DL\left(\frac{x-y_j}{t}\right)-DL\left(\frac{x-y_i}{t}\right)}{\left|DL\left(\frac{x-y_j}{t}\right)-DL\left(\frac{x-y_i}{t}\right)\right|}\qquad\text{for }\mathscr{H}^{d-1}\ a.e.\ x\in\Omega\bigcap\partial\mathscr{V}_i^t\bigcap\partial\mathscr{V}_j^t.$$

In particular, this implies

$$\left|D\mathbf{b}_n(t,\cdot)-\frac{\mathbf{I}_d}{t}\right|(\Omega) \leq \frac{1}{2t}\sum_{i\neq j\in\{1,\cdots,p(\Omega)\}}|y_j-y_i|\cdot\mathscr{H}^{d-1}(\Omega\bigcap\partial\mathscr{V}_i^t\bigcap\partial\mathscr{V}_j^t). \tag{3.2.14}$$

For a fixed $x\in\Omega\bigcap\partial\mathscr{V}_i^t\bigcap\partial\mathscr{V}_j^t$, setting $p_i := DL\left(\frac{x-y_i}{t}\right)$ and $p_j := DL\left(\frac{x-y_j}{t}\right)$, we have

$$\nu_i(x) = \frac{p_j-p_i}{|p_j-p_i|}\qquad\text{and}\qquad y_j-y_i = DH(p_i)-DH(p_j).$$

From (3.2.10), it holds that $|p_i|\leq\max_{|q|\leq\Lambda_M}|DL(q)|$. Thus, the assumptions (3.2.3) and (3.2.5) yield

$$|y_i-y_j| \leq -\frac{1}{\gamma_M}\cdot\left\langle y_j-y_i,\nu_i(x)\right\rangle. \tag{3.2.15}$$

Recalling (3.2.13) and (3.2.14), we get

$$\left|D\mathbf{b}_n(t,\cdot)-\frac{\mathbf{I}_d}{t}\right|(\Omega) \leq \frac{1}{\gamma_M}\cdot\left|\operatorname{div}\mathbf{b}_n(t,\cdot)-\frac{d}{t}\right|(\Omega). \tag{3.2.16}$$

**2.** Let us now provide a bound on $\left|\operatorname{div}\mathbf{b}_n(t,\cdot)-\frac{d}{t}\right|(\Omega)$. Pick a point $x_0\in\Omega$. From (3.2.11), (3.2.13) and (3.2.15), the function $\mathbf{d}_n(t,x) := \frac{x-x_0}{t}-\mathbf{b}_n(t,x)$ is in $BV_{\text{loc}}\left(\mathbb{R}^d\right)$ and

$$\operatorname{div}\mathbf{d}_n(t,\cdot) = \frac{1}{2t}\cdot\sum_{i\neq j\in\{1,\cdots,p(\Omega)\}}\left\langle y_i-y_j,\nu_i\right\rangle\mathscr{H}^{d-1}_{\llcorner\Omega\bigcap\partial\mathscr{V}_i^t\bigcap\partial\mathscr{V}_i^t}$$

is a positive Radon measure. In particular, this implies that

$$|\text{div}\, \mathbf{d}_n(t,\cdot)|(\Omega) = \int_\Omega \text{div}\, \mathbf{d}_n(t,\cdot).$$

Let $\rho_\varepsilon \in \mathscr{C}_c^\infty(\mathbb{R}^d)$ be a family of modifiers, i.e., $\rho_\varepsilon(x) = \varepsilon^{-d}\rho\left(\frac{x}{\varepsilon}\right)$ for $\rho \in \mathscr{C}_c^\infty(\mathbb{R}^d)$ satisfying $\rho(x) \geq 0$, $\rho(x) = \rho(-x)$, $\text{supp}(\rho) \subset B_d(0,1)$ and $\int_{\mathbb{R}^d} \rho(x)dx = 1$. For every test function $\varphi_\varepsilon = \chi_\Omega * \rho_\varepsilon$, it holds that

$$\int_{\mathbb{R}^d} \varphi_\varepsilon \text{div}\, \mathbf{d}_n(t,\cdot) = -\int_{\mathbb{R}^d} \mathbf{d}_n(t,x)\cdot\nabla\varphi_\varepsilon(x)dx \leq \|\mathbf{d}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)} \cdot \int_{\mathbb{R}^d} |\nabla\varphi_\varepsilon(x)|dx.$$

Thus, taking $\varepsilon \to 0+$, we get $\int_\Omega \text{div}\, \mathbf{d}_n(t,\cdot) \leq \|\mathbf{d}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)} \cdot \mathscr{H}^{d-1}(\partial\Omega)$ and (3.2.16) yields

$$|D\mathbf{b}_n(t,\cdot)|(\Omega) \leq \frac{1}{\gamma_M}\cdot\left(\|\mathbf{b}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)} + \frac{\text{diam}(\Omega)}{t}\right)\cdot\mathscr{H}^{d-1}(\partial\Omega) + \frac{\sqrt{d}}{t}\cdot|\Omega|.$$

This completes the proof. □

Using Proposition 44, we can easily prove Theorem 43.

**Proof of Theorem 43.** We first claim that $\mathbf{b}_n(t,\cdot)$ converges to $\mathbf{b}(t,\cdot)$ in $\mathbf{L}^1_{\text{loc}}$. Since the sequence $\mathbf{b}_n(t,\cdot)$ is bounded in $[\mathbf{L}^\infty(\mathbb{R}^d)]^d$ and the set $\left(\bigcup_{n\geq1}\Sigma_t^n \bigcup \Sigma_t\right)$ has zero Lebesgue measure, it is sufficient to show that

$$\lim_{n\to\infty} \mathbf{b}_n(t,x) = \mathbf{b}(t,x) \qquad \text{for all } x \in \mathbb{R}^d \setminus \left(\bigcup_{n\geq1}\Sigma_t^n \bigcup \Sigma_t\right).$$

Assume by a contradiction that there exists a subsequence $\mathbf{b}_{n_k}(t,x)$ converging to some $w \neq \mathbf{b}(t,x)$. Since $u_n(t,\cdot)$ converges uniformly to $u(t,\cdot)$ in any compact subset of $\mathbb{R}^d$, we have

$$\begin{aligned} u(t,x) = \lim_{n_k\to\infty} u_{n_k}(t,x) &= \lim_{n_k\to\infty} u_0(x - t\mathbf{b}_{n_k}(t,x)) + t\cdot L(\mathbf{b}_{n_k}(t,x)) \\ &= u_0(x-tw) + t\cdot L(w) = u_0(x-tw) + t\cdot L\left(\frac{x-(x-tw)}{t}\right). \end{aligned}$$

Thus, $\mathbf{b}(t,x)$ is not a singleton and this yields a contradiction. Finally, from (3.2.12) and [2, Proposition 3.13], the function $\mathbf{b}_n(t,\cdot)$ converges weakly to $\mathbf{b}(t,\cdot)$ in $BV(\Omega,\mathbb{R}^d)$. In particular,

$\mathbf{b}(t,\cdot)$ has locally bounded variation and we obtain

$$|D\mathbf{b}(t,\cdot)|(\Omega) \leq \liminf_{n\to\infty} |D\mathbf{b}_n(t,\cdot)|(\Omega)$$

$$\leq \frac{1}{\gamma_M}\cdot\left(\liminf_{n\to+\infty}\|\mathbf{b}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)}+\frac{\operatorname{diam}(\Omega)}{t}\right)\cdot\mathscr{H}^{d-1}(\partial\Omega)+\frac{\sqrt{d}}{t}\cdot|\Omega|.$$

To conclude the proof, we recall (3.2.10), which yields (3.2.6).

□

As a consequence of Theorem 43, the following holds.

**Corollary 45.** *Under the same assumptions as in Theorem 43, the set*

$$S_T\left(\mathscr{U}_{[m,M]}\right) = \left\{S_T(\bar{u}) : \bar{u}\in\mathscr{U}_{[m,M]}\right\}$$

*with*

$$\mathscr{U}_{[m,M]} := \left\{\bar{u}\in\mathbf{Lip}\big(\mathbb{R}^d\big) : |\bar{u}(0)|\leq m, \mathrm{Lip}[\bar{u}]\leq M\right\} \tag{3.2.17}$$

*is compact in* $\mathbf{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ *for every* $T>0$.

*Proof.* Let $(\bar{u}_n)_{n\geq 1}\subseteq\mathscr{U}_{[m,M]}$ be a sequence of initial data. Setting

$$v_n(x) := S_T(\bar{u}_n)(x) \qquad \text{for all } x\in\mathbb{R}^d, n\geq 1,$$

we have

$$\|DH(Dv_n)\|_{\mathbf{L}^\infty} \leq \Lambda_M \qquad \text{and} \qquad |v_n(0)| \leq m+MT\Lambda_M.$$

Moreover, for any given $R>0$, Theorem 43 implies that

$$|D(DH(Dv_n))|(B_d(0,R)) \leq C_R \quad \text{for some constant } C_R>0.$$

From Helly's theorem, we infer that there exist a subsequence $(v_{n_k})_{k\geq 1}$ and $w\in BV_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d)$ such that

- $v_{n_k}(0)$ converges to some $\bar{v}_0\in\mathbb{R}$;
- $DH(Dv_{n_k})$ converges to $w$ point-wise and

$$\lim_{k\to\infty}\|DH(Dv_{n_k})-w\|_{\mathbf{L}^1(B_d(0,R))} = 0.$$

This implies that $Dv_{n_k} = DL(DH(Dv_{n_k}))$ converges to $DL(w)$ point-wise and

$$\lim_{n_k\to+\infty} \|Dv_{n_k} - DL(w)\|_{\mathbf{L}^1(B_d(0,R))} = 0.$$

Thus, denoting by

$$\bar{v}^R_{n_k} := \frac{1}{|B_d(0,R)|}\cdot\int_{B_d(0,R)} v_{n_k}(x)dx$$

the average value of $v_{n_k}$ in $B_d(0,R)$, we have

$$\begin{aligned}\lim_{n_k\to\infty}\left(\bar{v}^R_{n_k} - v_{n_k}(0)\right) &= \lim_{n_k\to\infty}\frac{1}{|B_d(0,R)|}\cdot\int_0^1\int_{B_d(0,R)} Dv_{n_k}(sx)(x)\,dx\,ds\\ &= \frac{1}{|B_d(0,R)|}\cdot\int_0^1\int_{B_d(0,R)} DL(w(sx))(x)dxds := \bar{v}^R\end{aligned}$$

and this yields $\lim_{n_k\to\infty}\bar{v}^R_{n_k} = \bar{v}_0 + \bar{v}^R$. On the other hand, by Poincaré inequality, it holds that

$$\left(\left\|\left(v_{n_k} - \bar{v}^R_{n_k}\right) - \left(v_{n'_k} - \bar{v}^R_{n'_k}\right)\right\|_{\mathbf{L}^1(B_d(0,R))}\right) \le R\cdot\left\|Dv_{n_k} - Dv_{n'_k}\right\|_{\mathbf{L}^1(B_d(0,R))}.$$

Therefore, the sequence $\left(v_{n_k}\right)_{k\ge1}$ is a Cauchy sequence in $\mathbf{W}^{1,1}(B_d(0,R))$ for every $R>0$ and converges to $\bar{v}$ in $\mathbf{W}^{1,1}_{\mathrm{loc}}\left(\mathbb{R}^d\right)$. □

### 3.2.2 Metric entropy in $\mathbf{W}^{1,1}$ for $S_T$

Assuming that $H$ is in $\mathscr{C}^2(\mathbb{R}^d)$, we shall establish upper and lower estimates for the metric entropy of

$$S^R_T(\mathscr{U}_{[m,M]}) := \left\{v_{\llcorner\Box_R} \,:\, v\in S_T\left(\mathscr{U}_{[m,M]}\right)\right\}$$

in $\mathbf{W}^{1,1}(\Box_R)$ for given constants $T,R,m,M>0$. In order to do so, let us introduce the following continuous functions:

- $\Psi_M : [0,M]\to[0,\infty)$ with $\Psi(0)=0$ and

$$\Psi_M(s) = s\cdot\min_{|p-q|\ge s;\,p,q\in\overline{B}_d(0,M)}\frac{|DH(p)-DH(q)|}{|p-q|}\quad\text{for all } s\in(0,M]; \tag{3.2.18}$$

- $\Phi_M : [0,M]\to[0,\infty)$ with $\Phi(0)=0$ and

$$\Phi_M(s) = s\cdot\min_{p\in\overline{B}_d\left(0,M-\frac{s}{2}\right)}\left(\max_{q\in\overline{B}_d(p,\frac{s}{2})}\left\|D^2H(q)\right\|_\infty\right)\quad\text{for all } s\in(0,M] \tag{3.2.19}$$

with $\left\|D^2H(q)\right\|_\infty := \max_{|v|\le 1}\left|D^2H(q)(v)\right|$.

Notice that both maps $s\mapsto \Psi_M(s)$ and $s\mapsto \Phi_M(s)$ are strictly increasing and the strictly convex property of $H$ implies that

$$0 \;<\; \Psi_M(s) \;\le\; \Phi_M(s) \;<\; M\cdot \max_{p\in\overline{B}_d(0,M)} \left\|D^2H(p)\right\|_\infty \qquad \text{for all } s\in(0,M].$$

From Theorem 43, for any $v\in S_T^R(\mathscr{U}_{[m,M]})$, it holds that

$$|DH(Dv)|(\square_R) \;\le\; V_T \qquad \text{and} \qquad \|v\|_{\mathbf{L}^\infty(\mathbb{R}^d)} \;\le\; m_T \tag{3.2.20}$$

with

$$\begin{cases} V_T & := \dfrac{d2^dR^{d-1}}{\gamma_M}\cdot\left(\Lambda_M+\dfrac{2\sqrt{d}R}{T}\right)+\dfrac{\sqrt{d}2^dR^d}{T} \\[2ex] m_T & := m+\sqrt{d}MR+T\cdot\sup\limits_{|q|\le\Lambda_M} L\big(q\big). \end{cases}$$

Our main result is stated as follows.

**Theorem 46.** *Assume that $H\in\mathscr{C}^2\left(\mathbb{R}^d\right)$ is strictly convex, coercive and satisfies (3.2.3). Then, for every*

$$0<\varepsilon<\min\left\{R^+\Psi_M^{-1}\left(\min\left\{\frac{12RV_T^2}{3V_T+2\Lambda_M},4V_T\left(\frac{RV_T}{\Lambda_M}\right)^{1/d}\right\}\right),R^-\cdot\Phi_M^{-1}\left(\frac{\lambda_M}{2T}\right)\right\},$$

*it holds that*

$$\log_2\left(\left\lfloor\frac{\beta^-}{\varepsilon}\right\rfloor\right)+\Gamma^-\cdot\left(\Phi_M\left(\frac{\varepsilon}{R^-}\right)\right)^{-d} \;\le\; \mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\middle|\,\mathbf{W}^{1,1}(\square_R)\right)$$
$$\le\; \log_2\left(\frac{\beta^+}{\varepsilon}\right)+\Gamma^+\cdot\left(\Psi_M\left(\frac{\varepsilon}{R^+}\right)\right)^{-d} \tag{3.2.21}$$

*where the constants $\beta^\pm, R^\pm$ and $\Gamma^\pm$ are explicitly computed to be*

$$\beta^- \;=\; 2^dR^dm, \qquad \beta^+ \;=\; \left(2^{d+1}R^d+2\right)(3+\sqrt{d}R)m_T, \qquad R^- \;=\; \frac{\omega_d\cdot R^d}{(d+1)2^{d+9}}$$

$$R^+ \;=\; \left(2^dR^d+1\right)(3+\sqrt{d}R), \quad \Gamma^- \;=\; \frac{1}{8\ln 2}\cdot\left(\frac{8R\lambda_M}{3T}\right)^d, \quad \Gamma^+ \;=\; 48\sqrt{d}\cdot\left(12d\sqrt{d}RV_T\right)^d.$$

Before proving Theorem 46, we present some cases where (3.2.21) gives sharp estimates.

**Remark 47.** The Hamiltonian $H(p)=|p|^{2k}$ for some postive integer $k\geq 2$ is not uniformly convex but satisfies all assumptions in Theorem 43. Moreover, a direct computation yields

$$\alpha_1 s^{2k-1} \leq \Psi_M(s) \leq \Phi_M(s) \leq \alpha_2 s^{2k-1} \qquad \text{for all } s\in[0,M]$$

for some constants $\alpha_1,\alpha_2>0$ depending on $k$ and

$$\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\Big|\,\mathbf{W}^{1,1}(\Box_R)\right) \approx \varepsilon^{-(2k-1)d}.$$

**Remark 48.** If $H\in\mathscr{C}^2(\mathbb{R}^d)$ is uniformly convex ,then $\alpha_1 s < \Psi_M(s) \leq \Phi_M(s) < \alpha_2 s$ for some $0<\alpha_1<\alpha_2$ and (3.2.21) yields the same result in [3] that

$$\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\Big|\,\mathbf{W}^{1,1}(\Box_R)\right) \approx \varepsilon^{-d}.$$

**Remark 49.** For the one-dimensional case ($d=1$), every strictly convex $H\in\mathscr{C}^2(\mathbb{R})$ satisfies (3.2.3). In addition, assume that $H$ has polynomial degeneracy, i.e., the set $I_H=\{\omega\in\mathbb{R}: H''(\omega)=0\}\neq\emptyset$ is finite and for each $w\in I_H$, there exists a natural number $p_\omega\geq 2$ such that

$$H^{(p_\omega+1)}(\omega)\neq 0 \qquad \text{and} \qquad H^{(j)}(\omega) = 0 \quad \text{for all } j\in\{2,\dots,p_\omega\}.$$

The polynomial degeneracy of $H$ is defined by

$$\mathbf{p}_H := p_{\omega_H} = \max_{\omega\in I_H} p_\omega \quad \text{for some } \omega_H\in I_H.$$

For every $M>\omega_H$, there exist $0<\alpha_1<\alpha_2$ such that $\alpha_1\cdot s^{\mathbf{p}_H} < \Psi_M(s) \leq \Phi_M(s) < \alpha_2\cdot s^{\mathbf{p}_H}$ and (3.2.21) implies that

$$\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\Big|\,\mathbf{W}^{1,1}(\Box_R)\right) \approx \varepsilon^{-\mathbf{p}_H}.$$

#### 3.2.2.1 Upper estimate

Towards the upper estimate of $\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\Big|\,\mathbf{W}^{1,1}(\Box_R)\right)$ in (3.2.21), we first provide a bound on the $\mathbf{L}^1$-distance between elements $Du_1$ and $Du_2$ in terms of the $\mathbf{L}^1$-distance between $DH(Du_1)$ and $DH(Du_2)$ for every $u_1,u_2\in S_T^R(\mathscr{U}_{[m,M]})$ by using the function $\Psi$ defined in (3.2.18). Observing that the map $s\mapsto\dfrac{\Psi_M(s)}{s}$ is monotone increasing and

$$\Psi_M(|p-q|) \leq |DH(p)-DH(q)| \quad \text{for all } p,q\in\overline{B}_d(0,M), \tag{3.2.22}$$

we prove the following lemma.

**Lemma 50.** *For any* $u_1, u_2 \in S_T^R(\mathscr{U}_{[m,M]}$, *it holds that*

$$\|Du_1 - Du_2\|_{\mathbf{L}^1(\Box_R)} \leq \left(2^d R^d + 1\right) \cdot \Psi_M^{-1}\left(\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}\right) \tag{3.2.23}$$

*with* $\mathbf{b}_1 := DH(Du_1)$ *and* $\mathbf{b}_2 := DH(Du_2)$.

*Proof.* For simplicity, setting $\beta := \Psi_M^{-1}\left(\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}\right)$, we claim that

$$|Du_1(x) - Du_2(x)| \leq \beta \cdot \max\left\{1, \frac{|\mathbf{b}_1(x) - \mathbf{b}_2(x)|}{\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}}\right\} \quad \text{for } a.e.\ x \in \Box_R. \tag{3.2.24}$$

Indeed, assume that $|Du_1(x) - Du_2(x)| > \beta$. From (3.2.22), it holds that

$$\begin{aligned} |Du_1(x) - Du_2(x)| &= \frac{|Du_1(x) - Du_2(x)|}{|DH(Du_1(x)) - DH(Du_2(x))|} \cdot |\mathbf{b}_1(x) - \mathbf{b}_2(x)| \\ &\leq \frac{|Du_1(x) - Du_2(x)|}{\Psi_M(|Du_1(x) - Du_2(x)|)} \cdot |\mathbf{b}_1(x) - \mathbf{b}_2(x)|. \end{aligned}$$

By the monotone increasing property of the map $s \longmapsto \frac{\Psi_M(s)}{s}$, one has

$$|Du_1(x) - Du_2(x)| \leq \frac{\beta}{|\Psi_M(\beta)|} \cdot |\mathbf{b}_1(x) - \mathbf{b}_2(x)| = \beta \cdot \frac{|\mathbf{b}_1(x) - \mathbf{b}_2(x)|}{\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}}$$

and this implies (3.2.24). Therefore, the $\mathbf{L}^1$-distance between $Du_1$ and $Du_2$ is bounded by

$$\begin{aligned} \|Du_1 - Du_2\|_{\mathbf{L}^1(\Box_R)} &= \int_{\Box_R} |Du_1 - Du_2(x)|\,dx \\ &\leq \beta \cdot \int_{\Box_R} \left(1 + \frac{|\mathbf{b}_1(x) - \mathbf{b}_2(x)|}{\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}}\right) dx = (2^d R^d + 1)\beta \\ &= \left(2^d R^d + 1\right) \cdot \Psi_M^{-1}\left(\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}\right) \end{aligned} \tag{3.2.25}$$

and this yields (3.2.23). □

Now we recall a result from Section 3.1 before proceeding to find the upper estimate of $\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]}) \,\middle|\, \mathbf{W}^{1,1}(\Box_R)\right)$. For any given constants $R, M, V > 0$, we consider a class of uniformly bounded total variation functions on $\Box_R$ denoted by

$$\mathscr{F}_{[R,M,V]} = \left\{ f : \Box_R \to \mathbb{R}^d \,\middle|\, \|f\|_{\mathbf{L}^\infty(\Box_R)} \leq M, |Df|(\Box_R) \leq V \right\}. \tag{3.2.26}$$

By a slight modification in the proof of Theorem 41, we obtain the following upper bound on the $\varepsilon$-entropy of $\mathscr{F}_{[R,M,V]}$ in $\mathbf{L}^1(\Box_R)$.

**Corollary 51.** *For every* $0<\varepsilon<\min\left\{\frac{6RV^2}{3V+2M}, 2V\left(\frac{RV}{M}\right)^{\frac{1}{d}}\right\}$*, it holds that*

$$\mathscr{H}_\varepsilon\left(\mathscr{F}_{[R,M,V]}\,\Big|\,\mathbf{L}^1(\Box_R)\right) \;\leq\; 48\sqrt{d}\cdot\left(\frac{6d\sqrt{d}RV}{\varepsilon}\right)^d. \tag{3.2.27}$$

*Proof.* By the definition of $\varepsilon$-entropy, we have

$$\mathscr{H}_\varepsilon\left(\mathscr{F}_{[R,M,V]}\,\Big|\,\mathbf{L}^1(\Box_R)\right) \;\leq\; d\cdot\mathscr{H}_{\frac{\varepsilon}{d}}\left(\mathscr{F}^1_{[R,M,V]}\,\Big|\,\mathbf{L}^1(\Box_R)\right) \tag{3.2.28}$$

with

$$\mathscr{F}^1_{[R,M,V]} \;=\; \left\{f:\Box_R\to\mathbb{R}\,\Big|\,\|f\|_{\mathbf{L}^\infty(\Box_R)}\leq M, |Df|\big(\Box_R\big)\leq V\right\}.$$

Considering a class of bounded total variation real-valued functions

$$\mathscr{B}_{[R,M,V]} \;=\; \left\{g:[0,R]\to[0,M]\,\big|\,|Dg|([0,R])\leq V\right\},$$

from [26, Lemma 2.3], for every $0<\varepsilon<\frac{RV}{3}$, we have

$$\mathscr{N}_\varepsilon\left(\mathscr{B}_{[R,\frac{9}{8}V,V]}\Big|\mathbf{L}^1([0,R])\right) \;\leq\; 2^{\frac{17RV}{\varepsilon}}$$

and this implies that

$$\mathscr{N}_\varepsilon\left(\mathscr{B}_{[R,M,V]}\Big|\mathbf{L}^1([0,R])\right) \;\leq\; \frac{8M}{V}\cdot\mathscr{N}_\varepsilon\left(\mathscr{B}_{[R,\frac{9}{8}V,V]}\Big|\mathbf{L}^1([0,R])\right) \;\leq\; \frac{8M}{V}\cdot 2^{\frac{17RV}{\varepsilon}}.$$

In particular, for every $0<\varepsilon<\dfrac{RV^2}{3V+M}$ such that $\dfrac{8M}{V}\leq 2^{\frac{RV}{\varepsilon}}$, it holds that

$$\mathscr{H}_\varepsilon\left(\mathscr{B}_{[R,M,V]}\Big|\mathbf{L}^1([0,R])\right) \;=\; \log_2\left(\mathscr{N}_\varepsilon\left(\mathscr{B}_{[R,M,V]}\Big|\mathbf{L}^1([0,R])\right)\right) \;\leq\; \frac{18RV}{\varepsilon}.$$

Using the above estimate, one can follow the same argument in the proof of [26, Theorem 3.1] to obtain that for every $0<\varepsilon<\min\left\{\frac{6RV^2}{3V+2M}, 2V\left(\frac{RV}{M}\right)^{\frac{1}{d}}\right\}$, it holds that

$$\mathscr{H}_\varepsilon\left(\mathscr{F}^1_{[R,M,V]}\,\Big|\,\mathbf{L}^1(\Box_R)\right) \;\leq\; \frac{48}{\sqrt{d}}\cdot\left(\frac{6\sqrt{d}RV}{\varepsilon}\right)^d$$

and (3.1.5) yields (3.2.27).

□

**Proof of the upper estimate of $\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\middle|\,\mathbf{W}^{1,1}(\square_R)\right)$ in Theorem 46 :**

**1.** Recalling the second inequality in (3.2.20), we have

$$\bar{v}^R \ :=\ \frac{1}{\mathrm{Vol}(\square_R)}\cdot\int_{\square_R} v(x)\,dx\ \in\ [-m_T, m_T] \qquad \text{for all } v\in S_T^R\left(\mathscr{U}_{[m,M]}\right).$$

For any $\varepsilon'>0$, we cover $[-m_T,m_T]$ by $K_{\varepsilon'}=\left\lfloor\dfrac{m_T}{\Psi_M^{-1}(\varepsilon')}\right\rfloor+1$ small intervals with length $2\Psi_M^{-1}(\varepsilon')$, i.e.,

$$[-m_T,m_T]\ \subseteq\ \bigcup_{i=1}^{K_{\varepsilon'}} B\left(a_i,\Psi_M^{-1}(\varepsilon')\right)\quad \text{for some } a_i\in[-m_T,m_T]$$

and then decompose the set $S_T^R\left(\mathscr{U}_{[m,M]}\right)$ into $K_{\varepsilon'}$ subsets as

$$S_T^R\left(\mathscr{U}_{[m,M]}\right)\ \subseteq\ \bigcup_{i=1}^{K_{\varepsilon'}} S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)$$

where

$$S_T^{R,i}(\mathscr{U}_{[m,M]})\ :=\ \left\{v\in S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right): \bar{v}^R\in B\left(a_i,\Psi_M^{-1}(\varepsilon')\right)\right\}.$$

Thus, for all $\varepsilon>0$, it holds that

$$\mathscr{N}_\varepsilon\left(S_T^R\left(\mathscr{U}_{[m,M]}\right)\middle|\mathbf{W}^{1,1}(\square_R)\right)\ \le\ \sum_{i=1}^{K_{\varepsilon'}}\mathscr{N}_\varepsilon\left(S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)\middle|\mathbf{W}^{1,1}(\square_R)\right). \tag{3.2.29}$$

**2.** Given $i\in\{1,2,\dots,K_{\varepsilon'}\}$, we shall provide an upper bound on the covering number $\mathscr{N}_\varepsilon\left(S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)\middle|\mathbf{W}^{1,1}(\square_R)\right)$ by introducing the set

$$\mathscr{B}_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)\ =\ \left\{DH(Dv): v\in S_T^{R,i}(\mathscr{U}_{[m,M]})\right\}.$$

From (3.2.28) and (3.2.20),

$$\mathscr{B}_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)\ \subseteq\ \mathscr{F}_{[R,\Lambda_M,V_T]}$$

and Corollary 51 yields

$$\mathscr{H}_{\varepsilon'/2}\left(\mathscr{B}_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)\middle|\mathbf{L}^1(\square_R)\right)\ \le\ 48\sqrt{d}\cdot\left(\frac{12d\sqrt{d}RV_T}{\varepsilon'}\right)^d\ =\ \Gamma^+\cdot(\varepsilon')^{-d} \tag{3.2.30}$$

for all $0 < \varepsilon' < \min\left\{\frac{12RV_T^2}{3V_T+2\Lambda_M}, 4V_T\left(\frac{RV_T}{\Lambda_M}\right)^{1/d}\right\}$. Hence, there exists a set of initial data $\mathscr{U}^{\varepsilon'}_{[m,M]}$ with its image under the map $S_T$ defined by

$$\mathbf{S}^i_{\varepsilon'} := \left\{\mathbf{v}_1,\dots,\mathbf{v}_{\beta_1^{\varepsilon'}}\right\} \subset S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right) \qquad \text{where} \qquad \beta_1^{\varepsilon'} := \mathrm{Card}\left(\mathscr{U}^{\varepsilon'}_{[m,M]}\right) \leq 2^{\Gamma^+\cdot(\varepsilon')^{-d}}$$

such that the following inclusion holds

$$\mathscr{B}_T^{R,i}\left(\mathscr{U}_{[m,M]}\right) \subseteq \bigcup_{j=1}^{\beta_1^{\varepsilon'}} B_{\mathbf{L}^1}(\mathbf{b}_j,\varepsilon') \qquad \text{with} \qquad \mathbf{b}_j := DH(D\mathbf{v}_j).$$

In particular, for any given $v \in S_T^R\left(\mathscr{U}_{[m,M]}\right)$, it holds that

$$\|DH(Dv)-\mathbf{b}_{j_0}\|_{\mathbf{L}^1(\Box_R)} < \varepsilon' \quad \text{for some } j_0 \in \overline{1,\beta_1^{\varepsilon'}}.$$

Recalling Lemma 50, we obtain

$$\begin{aligned}\|Dv - D\mathbf{v}_{j_0}\|_{\mathbf{L}^1(\Box_R)} &\leq \left(2^dR^d+1\right)\cdot\Psi_M^{-1}\left(\|DH(Dv)-\mathbf{b}_{j_0}\|_{\mathbf{L}^1(\Box_R)}\right)\\ &\leq \left(2^dR^d+1\right)\cdot\Psi_M^{-1}(\varepsilon')\end{aligned}$$

and Poincaré inequality yields

$$\left\|\left(v-\bar{v}^R\right)-\left(\mathbf{v}_{j_0}-\bar{\mathbf{v}}^R_{j_0}\right)\right\|_{\mathbf{L}^1(\Box_R)} \leq \sqrt{d}R\cdot\|Dv-D\mathbf{v}_{j_0}\|_{\mathbf{L}^1(\Box_R)} \leq \sqrt{d}R\left(2^dR^d+1\right)\cdot\Psi_M^{-1}(\varepsilon').$$

On the other hand, since $v,\mathbf{v}_{i_0} \in S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)$, we have

$$\left|\bar{v}^R-\bar{\mathbf{v}}^R_{j_0}\right| \leq \left|\bar{v}^R-a_i\right|+\left|\bar{\mathbf{v}}^R_{j_0}-a_i\right| \leq 2\Psi_M^{-1}(\varepsilon').$$

The $\mathbf{W}^{1,1}$-distance between $v$ and $\mathbf{v}_{j_0}$ can be estimated by

$$\begin{aligned}\|v-\mathbf{v}_{j_0}\|_{\mathbf{W}^{1,1}(\Box_R)} &\leq \|Dv-D\mathbf{v}_{j_0}\|_{\mathbf{L}^1(\Box_R)}+\left\|\left(v-\bar{v}^R\right)-\left(\mathbf{v}_{j_0}-\bar{\mathbf{v}}^R_{j_0}\right)\right\|_{\mathbf{L}^1(\Box_R)}\\ &\quad+\left|\bar{v}^R-\bar{\mathbf{v}}^R_{j_0}\right|\cdot\left|\Box_R\right| \leq \left(2^dR^d+1\right)(3+\sqrt{d}R)\cdot\Psi_M^{-1}(\varepsilon') = R^+\cdot\Psi_M^{-1}(\varepsilon').\end{aligned}$$

Thus, for any $0<\varepsilon<R^+\Psi_M^{-1}\left(\min\left\{\dfrac{12RV_T^2}{3V_T+2\Lambda_M}, 4V_T\left(\dfrac{RV_T}{\Lambda_M}\right)^{1/d}\right\}\right)$, if we choose $\varepsilon' = \Psi_M\left(\dfrac{\varepsilon}{R^+}\right)$ such that

$$0<\varepsilon'<\min\left\{\frac{12RV_T^2}{3V_T+2\Lambda_M}, 4V_T\left(\frac{RV_T}{\Lambda_M}\right)^{1/d}\right\}$$

then the set $S_T^{R,i}\left(\mathscr{U}_{[m,M]}\right)$ is covered by $\beta_1^{\varepsilon'}$ open balls in $\mathbf{W}^{1,1}\left(\Box_R\right)$ centered at $\mathbf{v}_i$ of radius $\varepsilon$, i.e.

$$S_T^{R,i}(\mathscr{U}_{[m,M]}) \subseteq \bigcup_{i=1}^{\beta_1^{\varepsilon'}} B_{\mathbf{W}^{1,1}}(\mathbf{v}_i,\varepsilon)$$

and thus

$$\mathscr{N}_\varepsilon\left(S_T^{R,i}(\mathscr{U}_{[m,M]}\Big|\mathbf{W}^{1,1}(\Box_R)\right) \leq \beta_1^{\varepsilon'} = 2^{\Gamma^+\cdot\left(\Psi_M\left(\frac{\varepsilon}{R^+}\right)\right)^{-d}}.$$

Finally, recalling (3.2.29), we get

$$\begin{aligned}\mathscr{N}_\varepsilon\left(S_T^{R}(\mathscr{U}_{[m,M]}\Big|\mathbf{W}^{1,1}(\Box_R)\right) &\leq \frac{\left(2^{d+1}R^d+2\right)(3+\sqrt{d}R)m_T}{\varepsilon}\cdot 2^{\Gamma^+\cdot\left(\Psi_M\left(\frac{\varepsilon}{R^+}\right)\right)^{-d}}\\ &= \frac{\beta^+}{\varepsilon}\cdot 2^{\Gamma^+\cdot\left(\Psi_M\left(\frac{\varepsilon}{R^+}\right)\right)^{-d}}\end{aligned}$$

and this yields the second inequality in (3.2.21). □

**Remark 52.** To obtain the upper bound of $\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\big|\,\mathbf{W}^{1,1}(\Box_R)\right)$ in (3.2.21), we only require that $H$ belongs to $\mathscr{C}^1(\mathbb{R}^d)$.

#### 3.2.2.2 Lower estimate

We will now prove the first inequality in (3.2.21). In order to do so, for any given $p\in\mathbb{R}^d$, let $\Phi(\cdot,p):[0,\infty)\to[0,\infty)$ be the strictly increasing continuous function defined by $\Phi(0,p)=0$ and

$$\Phi(s,p) = s\cdot\left(\max_{p'\in\overline{B}_d(p,\frac{s}{2})}\left\|D^2H(p')\right\|_\infty\right)\qquad\text{for all } s>0.$$

From the definition of $\Phi_M$ in (3.2.19), it holds that

$$\Phi_M(s) = \min_{p\in\overline{B}_d\left(0,M-\frac{s}{2}\right)}\Phi(s,p)\qquad\text{for all } s\in[0,M]. \tag{3.2.31}$$

The following proposition shows that a solution to (3.2.1) with a semiconvex initial condition preserves the semiconvexity on a given time interval, provided the semiconvexity constant of the initial data is sufficiently small in absolute value.

**Proposition 53.** *Given $T,M,r>0$ and $\bar{p}\in\overline{B}_d\left(0,M-\frac{r}{2}\right)$, let $\bar{u}$ be a semiconvex function with semiconvexity constant $-K$ such that*

$$D^-\bar{u}(\mathbb{R}^d)\subseteq\overline{B}_d\left(\bar{p},\frac{r}{2}\right)\qquad\text{and}\qquad K\leq\frac{\lambda_M}{4T}\cdot\frac{r}{\Phi(r,\bar{p})} \tag{3.2.32}$$

*with $\lambda_M$ as defined in (3.2.3). Then, the map $(t,x)\mapsto S_t(\bar{u})(x)$ is a classical solution for $0<t\leq T$ and*

$$DS_t(\bar{u})(x)\in\overline{B}_d\left(\bar{p},\frac{r}{2}\right)\qquad \textit{for all}\,(t,x)\in(0,T]\times\mathbb{R}^d.$$

*Proof.* For simplicity, we set $u(t,x):=S_t(\bar{u})(x)$ for all $(t,x)\in[0,\infty)\times\mathbb{R}^d$. It is well-known from [18, Theorem 5.3.8] that $u(t,\cdot)$ is locally semiconcave for every $t>0$. Thus, by Proposition 38, it is sufficient to show that $u(t,\cdot)$ is semiconvex with a semiconvexity constant $-C<0$ for all $t\in[0,T]$, i.e., for any fixed $(t,x)\in[0,T)\times\mathbb{R}^d$, it holds that

$$u(t,x+h)+u(t,x-h)-2u(t,x)\geq -C\cdot|h|^2\quad \text{for all } h\in\mathbb{R}^d. \tag{3.2.33}$$

By the Lipschitz continuity of $u(t,\cdot)$, we can assume that $u(t,\cdot)$ is differentiable at $x\pm h$. In this case, $\mathbf{b}(t,x\pm h)$ reduce to a single value denoted by $\mathbf{b}^\pm=DH(\mathbf{p}^\pm)$ with $\mathbf{p}^\pm=Du(t,x\pm h)$ and satisfy the relations

$$\begin{cases}\mathbf{p}^\pm\in D^-\bar{u}(x\pm h-t\mathbf{b}^\pm)\subseteq\overline{B}_d\left(\bar{p},\frac{r}{2}\right)\subseteq\overline{B}_d(0,M),\\ \\ u(t,x\pm h)=\bar{u}(x\pm h-t\mathbf{b}^\pm)+t\cdot L(\mathbf{b}^\pm).\end{cases} \tag{3.2.34}$$

Since $\bar{u}$ is semiconvex with semiconvexity constant $-K$, denoting $x^\pm:=x\pm h$, one has from Corollary 8 that $\left\langle\mathbf{p}^+-\mathbf{p}^-,x^+-x^--t(\mathbf{b}^+-\mathbf{b}^-)\right\rangle\geq -K\cdot|2h-t(\mathbf{b}^+-\mathbf{b}^-)|^2$ and

$$\begin{aligned}\left\langle\mathbf{p}^+-\mathbf{p}^-,\mathbf{b}^+-\mathbf{b}^-\right\rangle &\leq \frac{K}{t}\cdot\left|2h-t\left(\mathbf{b}^+-\mathbf{b}^-\right)\right|^2+\frac{2|h|}{t}\cdot|\mathbf{p}^+-\mathbf{p}^-|\\ &\leq 2Kt|\mathbf{b}^+-\mathbf{b}^-|^2+\frac{8K|h|^2}{t}+\frac{2|h|}{t}\cdot|\mathbf{p}^+-\mathbf{p}^-|\\ &\leq 2KT|DH(\mathbf{p}^+)-DH(\mathbf{p}^-)|^2+\frac{8K|h|^2}{t}+\frac{2|h|}{t}\cdot|\mathbf{p}^+-\mathbf{p}^-|.\end{aligned}$$

Since $\mathbf{p}^\pm\in\overline{B}_d(\bar{p},\frac{r}{2})$, it holds that

$$|DH(\mathbf{p}^+)-DH(\mathbf{p}^-)|\leq\frac{\Phi(r,\bar{p})}{r}\cdot|\mathbf{p}^+-\mathbf{p}^-|.$$

Thus, recalling (3.2.32) and (3.2.34), we estimate

$$\begin{aligned}2KT|DH(\mathbf{p}^+)-DH(\mathbf{p}^-)|^2 &\leq 2KT\cdot\frac{\Phi(r,\bar{p})}{r}\cdot|DH(\mathbf{p}^+)-DH(\mathbf{p}^-)|\cdot|\mathbf{p}^+-\mathbf{p}^-|\\ &\leq\frac{\lambda_M}{2}\cdot|DH(\mathbf{p}^+)-DH(\mathbf{p}^-)|\cdot|\mathbf{p}^+-\mathbf{p}^-|=\frac{\lambda_M}{2}\cdot|\mathbf{b}^+-\mathbf{b}^-|\cdot|\mathbf{p}^+-\mathbf{p}^-|\end{aligned}$$

and

$$\left\langle \mathbf{p}^+ - \mathbf{p}^-, \mathbf{b}^+ - \mathbf{b}^- \right\rangle \leq \frac{\lambda_M}{2} \cdot |\mathbf{b}^+ - \mathbf{b}^-| \cdot |\mathbf{p}^+ - \mathbf{p}^-| + \frac{8K|h|^2}{t} + \frac{2|h|}{t} \cdot |\mathbf{p}^+ - \mathbf{p}^-|. \tag{3.2.35}$$

On the other hand, from (3.2.3) we deduce that

$$\begin{aligned} \left\langle \mathbf{p}^+ - \mathbf{p}^-, \mathbf{b}^+ - \mathbf{b}^- \right\rangle &= \left\langle \mathbf{p}^+ - \mathbf{p}^-, DH(\mathbf{p}^+) - DH(\mathbf{p}^-) \right\rangle \geq \lambda_M \cdot |DH(\mathbf{p}^+) - DH(\mathbf{p}^-)| \cdot |\mathbf{p}^+ - \mathbf{p}^-| \\ &= \lambda_M \cdot |\mathbf{b}^+ - \mathbf{b}^-| \cdot |\mathbf{p}^+ - \mathbf{p}^-| \end{aligned}$$

and (3.2.35) yields

$$\frac{\lambda_M}{2} \cdot |t(\mathbf{b}^+ - \mathbf{b}^-)| \cdot |\mathbf{p}^+ - \mathbf{p}^-| \leq 8K|h|^2 + 2|h| \cdot |\mathbf{p}^+ - \mathbf{p}^-|. \tag{3.2.36}$$

We consider the following two cases.

- If $|\mathbf{p}^+ - \mathbf{p}^-| \leq K|h|$ then

$$|\mathbf{b}^+ - \mathbf{b}^-| = |DH(\mathbf{p}^+) - DH(\mathbf{p}^-)| \leq \frac{\Phi(r, \bar{p})}{r} \cdot |\mathbf{p}^+ - \mathbf{p}^-| \leq \frac{K\Phi(r, \bar{p})}{r} \cdot |h|.$$

- Otherwise, (3.2.36) implies that $\frac{\lambda_M}{2} \cdot |t(\mathbf{b}^+ - \mathbf{b}^-)| \leq 10|h|$.

Hence,

$$|t(\mathbf{b}^+ - \mathbf{b}^-)| \leq \left( \frac{KT\Phi(r, \bar{p})}{r} + \frac{20}{\lambda_M} \right) \cdot |h|. \tag{3.2.37}$$

By Hopf-Lax representation formula, we have $u(t, x \pm h) = \bar{u}(x \pm h - t\mathbf{b}^{\pm}) + t \cdot L(\mathbf{b}^{\pm})$ and

$$u(t, x) \leq 2\bar{u}\left( x - t \cdot \frac{\mathbf{b}^+ + \mathbf{b}^-}{2} \right) + t \cdot L\left( \frac{\mathbf{b}^+ + \mathbf{b}^-}{2} \right).$$

Using the convexity of $L$ and semiconvexity of $\bar{u}$, we estimate

$$\begin{aligned} u(t, x+h) + u(t, x-h) - 2u(t, x) \geq\ & t \cdot \left[ L(\mathbf{b}^+) + L(\mathbf{b}^-) - 2L\left( \frac{\mathbf{b}^+ + \mathbf{b}^-}{2} \right) \right] \\ & + \bar{u}(x + h - t\mathbf{b}^+) + \bar{u}(x - h - t\mathbf{b}^-) - 2\bar{u}\left( x - t \cdot \frac{\mathbf{b}^+ + \mathbf{b}^-}{2} \right) \\ \geq\ & -K \cdot \left| 2h - t\left( \mathbf{b}^+ - \mathbf{b}^- \right) \right|^2 \geq -8K|h|^2 - 2K|t(\mathbf{b}^+ - \mathbf{b}^-)|^2 \\ & \geq -2K \cdot \left[ 4 + \left( \frac{KT\Phi(r, \bar{p})}{r} + \frac{20}{\lambda_M} \right)^2 \right] \cdot |h|^2 \end{aligned}$$

and this yields (3.2.33). □

Relying on the above Proposition and Lemma 9, we prove the first inequality in (3.2.21).

**Proof of the lower estimate of $\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]})\,\Big|\,\mathbf{W}^{1,1}(\Box_R)\right)$ in Theorem 46:**

**1.** For any given $r>0$ and $p\in\mathbb{R}^d$, we denote by

$$\mathscr{S}\mathscr{C}_r^p \;:=\; \left\{\varphi = v+\langle p,\cdot\rangle : v\in SC_{[\frac{r}{2},K_r]}\right\} \qquad \text{with} \qquad K_r \;=\; \frac{\lambda_M}{4T}\cdot\frac{r}{\Phi_M(r)}$$

where $SC_{[\frac{r}{2},K_r]}$ is defined in (2.4.1). From (3.2.31), there exists $p_r\in\overline{B}_d(0,M-\frac{r}{2})$ satisfying

$$\Phi(r,p_r) \;=\; \Phi_M(r) \;=\; \min_{p\in\overline{B}\left(0,M-\frac{r}{2}\right)}\Phi(r,p).$$

Consider the operator $\mathscr{T}:\mathscr{S}\mathscr{C}_r^{p_r}\to\mathbf{Lip}\big(\mathbb{R}^d\big)$ defined by

$$\mathscr{T}(\varphi) \;=\; \varphi+S_T(\varphi_-)(0) \quad \text{with} \quad \varphi_-(x)=-\varphi(-x) \quad \text{for all } \varphi\in SC_r^{p_r}. \tag{3.2.38}$$

We show that

$$\mathscr{T}\left(\mathscr{S}\mathscr{C}_r^{p_r}\right) \;\subseteq\; S_T\left(\mathscr{U}_{[0,M]}\right) \quad \text{with} \quad \mathscr{U}_{[0,M]} \text{ as defined in (3.2.17).} \tag{3.2.39}$$

In order to find an initial data for a given function $\varphi\in\mathscr{S}\mathscr{C}_r^{p_r}$, we only need to reverse the equation. Since

$$\varnothing\neq D^+\varphi(x)\;\subset\;\overline{B}_d\left(0,\frac{r}{2}\right) \qquad \text{for all } x\in\mathbb{R}^d,$$

the following function

$$w_0(\cdot)\;:=\;-\mathscr{T}(\varphi)(-\cdot)\;=\;\varphi_-(\cdot)-S_T(\varphi_-)(0)$$

is semiconvex with a semiconvexity constant $-K_r$ and

$$D^-w_0(x)\;=\;p_r+D^+\varphi(-x)\;\subseteq\;\overline{B}_d\left(p_r,\frac{r}{2}\right) \qquad \text{for all } x\in\mathbb{R}^d.$$

Let $w(t,x)=S_t(w_0)(x)$ be the unique viscosity solution of (3.2.1) with initial datum $w_0$. Recalling Proposition 53 and property (ii) in Proposition 39, we have that $w$ is a $\mathscr{C}^1$ classical solution of (3.2.1) in $(0,T]\times\mathbb{R}^d$ and

$$D_xw(T,x)\;\subseteq\;\overline{B}_d\left(p_r,\frac{r}{2}\right)\;\subseteq\;\overline{B}_d(0,M) \qquad \text{for all } x\in\mathbb{R}^d.$$

Moreover, the translation property (iii) of the semigroup $S_t$ (as defined by 2.7.9) implies that

$$w(T,0) = S_T(w_0)(0) = S_T\big(\varphi_- - S_T(\varphi_-)(0)\big)(0) = 0.$$

Thus, the continuous function $u : [0,T]\times\mathbb{R}^d \to \mathbb{R}$, defined by

$$u(t,x) = -w(T-t,-x) \qquad \text{for all } (t,x)\in[0,T]\times\mathbb{R}^d,$$

is also a $\mathscr{C}^1$ classical solution of (3.2.1) in $(0,T)\times\mathbb{R}^d$ with

$$u(T,\cdot) \equiv \mathscr{T}(\varphi)(\cdot) \quad \text{and} \quad u(0,\cdot) = -w(T,-\cdot) \in \mathscr{U}_{[0,M]}.$$

In particular, $u(t,x)$ is a viscosity solution of (3.2.1) in $[0,T]\times\mathbb{R}^d$, so that by the uniqueness property of the semigroup map $S_t$, we get

$$S_T(u_0)(\cdot) \equiv \mathscr{T}(\varphi)(\cdot), \qquad u_0(\cdot) \equiv -w(T,-\cdot)$$

and this implies $\mathscr{T}(\varphi)(\cdot)\in S_T\big(\mathscr{U}_{[0,M]}\big)$.

**2.** For every $\varepsilon>0$, we select a finite subset $A_\varepsilon \subseteq [-m,m]$ such that

$$\mathrm{Card}(A_\varepsilon) = \left\lfloor \frac{2^d R^d m}{\varepsilon} \right\rfloor \quad \text{and} \quad |a_i - a_j| \geq \frac{2\varepsilon}{2^d R^d} \quad \text{for all } a_i \neq a_j \in A_\varepsilon. \tag{3.2.40}$$

From the translation property (iii) of the semigroup $S_t$, it holds that

$$S_T\big(\mathscr{U}_{[m,M]}\big) \supseteq \bigcup_{a\in A_\varepsilon} S_T\big(a+\mathscr{U}_{[0,M]}\big) = A_\varepsilon + S_T\big(\mathscr{U}_{[0,M]}\big)$$

and (3.2.39) implies that

$$S_T\big(\mathscr{U}_{[m,M]}\big) \supseteq A_\varepsilon + \mathscr{T}\big(\mathscr{S}\mathscr{C}_r^{p_r}\big). \tag{3.2.41}$$

By Lemma 9, there exists a subset $\mathscr{G}_r^R$ of $\mathscr{S}\mathscr{C}_r^{p_r}$ such that

$$\mathrm{Card}\big(\mathscr{G}_r^R\big) \geq 2^{\beta_{[R,r]}\cdot\varepsilon^{-d}} \qquad \text{with} \qquad \beta_{[R,r]} = \frac{1}{3^d 2^{d^2+4d+3}\ln 2}\cdot\left(\frac{\omega_d R^{d+1} K_r}{(d+1)}\right)^d$$

and

$$\Big\| D\varphi_{\llcorner\square_R} - D\phi_{\llcorner\square_R} \Big\|_{\mathbf{L}^1(\square_R)} \geq 2\varepsilon \qquad \text{for all } \varphi\neq\phi\in\mathscr{G}_r^R$$

provided that

$$0 < \varepsilon \leq \min\left\{\frac{r}{2}, K_r\right\}\cdot\frac{\omega_d\cdot R^d}{(d+1)2^{d+8}}. \tag{3.2.42}$$

Since $D\mathscr{T}(\varphi)(x) = D\varphi(x)$ for all $x \in \mathbb{R}^d$, we get

$$\left\| D\mathscr{T}(\varphi)_{\llcorner\Box_R} - D\mathscr{T}(\phi)_{\llcorner\Box_R} \right\|_{\mathbf{W}^{1,1}(\Box_R)} \geq 2\varepsilon \qquad \text{for all } \varphi \neq \phi \in \mathscr{G}_r^R.$$

Recalling (3.2.40), we have $\left\| f_{\llcorner\Box_R} - g_{\llcorner\Box_R} \right\|_{\mathbf{W}^{1,1}(\Box_R)} \geq 2\varepsilon$ for all $f \neq g \in A_\varepsilon + \mathscr{T}\left(\mathscr{S}\mathscr{C}_r^{p_r}\right)$ and

$$\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]}\middle|\mathbf{W}^{1,1}(\Box_R)\right) \geq \log_2\left(\mathrm{Card}(A_\varepsilon)\cdot\mathrm{Card}\left(\mathscr{G}_r^R\right)\right) = \log_2\left(\left\lfloor \frac{2^d R^d m}{\varepsilon} \right\rfloor\right) + \frac{\beta_{[R,r]}}{\varepsilon^d}.$$

Finally, by choosing $r = \dfrac{\varepsilon}{R^-}$ with $R^- = \dfrac{\omega_d \cdot R^d}{(d+1)2^{d+9}}$, we compute that

$$K_r = \frac{\lambda_M}{4TR^-}\cdot\frac{\varepsilon}{\Phi_M\left(\frac{\varepsilon}{R^-}\right)} \qquad \text{and} \qquad \beta_{[R,r]} = \frac{1}{8\ln 2}\cdot\left(\frac{8R\lambda_M}{3T}\right)^d\cdot\left(\frac{\varepsilon}{\Phi_M\left(\frac{\varepsilon}{R^-}\right)}\right)^d.$$

Thus, for every $0 < \varepsilon \leq R^-\cdot\Phi_M^{-1}\left(\frac{\lambda_M}{2T}\right)$ such that (3.2.42) holds, we obtain

$$\mathscr{H}_\varepsilon\left(S_T^R\left(\mathscr{U}_{[0,M]}\right)\middle|\mathbf{W}^{1,1}(\Box_R)\right) \geq \frac{1}{8\ln 2}\cdot\left(\frac{8R\lambda_M}{3T}\right)^d\cdot\left(\Phi_M\left(\frac{\varepsilon}{R^-}\right)\right)^{-d} + \log_2\left(\left\lfloor \frac{2^d R^d m}{\varepsilon}\right\rfloor\right)$$

and this yields the first inequality in (3.2.21). □

**Remark 54.** With the same argument, the lower bound of $\mathscr{H}_\varepsilon\left(S_T^R(\mathscr{U}_{[m,M]}) \,\middle|\, \mathbf{W}^{1,1}(\Box_R)\right)$ in (3.2.21) can be obtained for $H \in \mathscr{C}^{1,1}(\mathbb{R}^d)$ by defining

$$\Phi_M(s) = s\cdot \inf_{p\in\overline{B}_d\left(0,M-\frac{s}{2}\right)}\left(\sup_{p_1\neq p_2\in\overline{B}_d(p,\frac{s}{2})}\frac{|DH(p_1)-DH(p_2)|}{|p_1-p_2|}\right)$$

for all $s > 0$.

### 3.2.3 A counter-example

In this subsection, we show that if the strictly convex and coercive Hamiltonian $H \in \mathscr{C}^2(\mathbb{R}^2)$ does not satisfy the uniform directional convexity condition (3.2.3) then Theorem 43 fails in general. Indeed, let us consider the following Hamiltonian

$$H(p) = \frac{3^3}{4^4}\cdot p_1^4 + p_2^2 \qquad \text{for all } p = (p_1, p_2) \in \mathbb{R}^2.$$

The associated Lagrangian $L$ of $H$ is computed by $L(q) = |q_1|^{\frac{4}{3}} + q_2^2$ for all $q = (q_1, q_2) \in \mathbb{R}^2$.

For any given $\bar{q} = (\bar{q}_1, \bar{q}_2) \in \mathbb{R}^2$, we have

$$L(q) = L(q - \bar{q}) \qquad \Longleftrightarrow \qquad |q_1|^{4/3} + q_2^2 = |q_1 - \bar{q}_1|^{4/3} + (q_2 - \bar{q}_2)^2$$

$$\Longleftrightarrow \qquad q_2 = \frac{|q_1 - \bar{q}_1|^{4/3} - |q_1|^{4/3}}{2\bar{q}_2} + \frac{\bar{q}_2}{2}.$$

Let $\gamma_{\bar{q}} : \mathbb{R} \to \mathbb{R}$ be such that

$$\gamma_{\bar{q}}(s) = \frac{|s - \bar{q}_1|^{4/3} - |s|^{4/3}}{2\bar{q}_2} + \frac{\bar{q}_2}{2} \qquad \text{for all } s \in \mathbb{R}.$$

In particular, assuming that $\bar{q}_2 = |\bar{q}_1|^{2/3}$ with $|\bar{q}_1| = \delta$ for some $\delta > 0$, it holds that

$$\gamma_{\bar{q}}(0) = \bar{q}_2, \qquad \gamma_{\bar{q}}(\bar{q}_1) = 0$$

and the following curve which connects two points $(0, \bar{q}_2)$ and $(\bar{q}_1, 0)$

$$\Gamma_{\bar{q}} = \begin{cases} \{(s, \gamma_{\bar{q}}(s)) : s \in [0, \delta]\} \subset [0, \delta] \times [0, \delta^{2/3}] & \text{if } \bar{q}_1 = \delta > 0 \\ \\ \{(s, \gamma_{\bar{q}}(s)) : s \in [-\delta, 0]\} \subset [-\delta, 0] \times [0, \delta^{2/3}] & \text{if } \bar{q}_1 = -\delta < 0 \end{cases} \tag{3.2.43}$$

has length $> \delta^{2/3}$. From this observation, we shall construct a uniformly Lipschitz initial datum $\bar{u}$ such that both $Du(1, \cdot)$ and $\mathbf{b}(1, \cdot) = DH(Du(1, \cdot))$ do not have locally bounded variation where $u = S_t(\bar{u})$ is the solution of (3.2.1) with $u(0, \cdot) \equiv \bar{u}$. Our construction is divided into two main steps.

**1.** Given $0 < \ell < 1$, first we construct a uniformly Lipschitz function $\bar{u} : \mathbb{R}^2 \to [0, \infty)$ with a Lipschitz constant which does not depend on $\ell$ such that

$$\text{supp}(\bar{u}) \subset [-2\ell, 2\ell] \quad \text{and} \quad |\mathbf{b}(1, \cdot)|([-\ell, \ell]^2), |Du(1, \cdot)|([-\ell, \ell]^2) \geq 1 \tag{3.2.44}$$

where

$$u(t, \cdot) = S_t(\bar{u}) \qquad \text{for all } t \geq 0.$$

For every $0 < \delta < \ell$, we consider the periodic lattice

$$y_\iota = \left(\iota_1 \delta, \iota_2 \delta^{2/3}\right) \quad \text{with} \quad \iota \in \mathscr{Z}_2 := \left\{(\iota_1', \iota_2') \in \mathbb{Z}^2 : \iota_1' + \iota_2' \in 2\mathbb{Z}\right\}$$

and the corresponding regions

$$\begin{aligned}\Omega_\iota &= \{x\in\mathbb{R}^2 : L(x-y_\iota)<L(x-y_{\iota'}) \quad \text{for all } \iota'\neq\iota\}\\ &= y_\iota+\left\{q\in\mathbb{R}^2 : L(q)<L(q+y_\iota-y_{\iota'}) \quad \text{for all } \iota'\neq\iota\right\}\\ &\subseteq y_\iota+[-\delta,\delta]\times[-\delta^{2/3},\delta^{-2/3}]\end{aligned}$$

with $\partial\Omega_\iota = \left[y_\iota+(\Gamma_{\bar q^+}\cup\Gamma_{\bar q^-})\right]\bigcup\left[y_{\iota-(1,1)}+\Gamma_{\bar q^+}\right]\bigcup\left[y_{\iota-(-1,1)}+\Gamma_{\bar q^-}\right]$, $\bar q^\pm=(\pm\delta,\delta^{2/3})$.

This corresponds to the function

$$g_1(x) = L(x-y_\iota) \quad \text{for all } x\in\Omega_\iota, \iota\in\mathscr{Z}_2.$$

The dual solution is

$$g_0(y) = \max_{x\in\mathbb{R}^2}\left\{g_1(x)-L(x-y)\right\} \quad \text{for all } y\in\mathbb{R}^2 .$$

By the definition of $\Omega_\iota$, both $g_0$ and $g_1$ are Lipschitz with a Lipschitz constant

$$M_\delta = \sup_{q\in[-\delta,\delta]\times[-\delta^{2/3},\delta^{-2/3}]}|DL(q)| = O(\delta^{1/3}).$$

For $\delta>0$ sufficiently small, we can construct a Lipschitz initial datum $\bar u$ with a Lipschitz constant $M_\delta$ and

$$\operatorname{supp}(\bar u) \subset [-2\ell,2\ell], \qquad \bar u(y) = g_0(y) \text{ for all } y\in\left[-\frac{3\ell}{2},\frac{3\ell}{2}\right]^2.$$

Let $u(t,\cdot)=S_t(\bar u)$ be the solution. At time $t=1$, we have

$$u(1,x) = \min_{y\in\mathbb{R}^2}\left\{\bar u(y)+L(x-y)\right\} = \bar u(y_x)+L(x-y_x)$$

for some $y_x\in\overline{B}(x,\Lambda_{M_\delta})$ with $\Lambda_{M_\delta}=\max\{|q|:L(q)\le M_\delta\cdot|q|\}=O(\delta^{1/3})$ being the maximal characteristic speed. Thus, if $M_\delta\le\frac{\ell}{2}$ then for all $x\in[-\ell,\ell]^2\cap\Omega_\iota$, $\iota\in\mathscr{Z}_2$,

$$\begin{aligned}u(1,x) &= \min_{y\in\left[-\frac{3\ell}{2},\frac{3\ell}{2}\right]^2}\left\{\bar u(y)+L(x-y)\right\} = \min_{y\in\left[-\frac{3\ell}{2},\frac{3\ell}{2}\right]^2}\left\{g_0(y)+L(x-y)\right\}\\ &= \min_{y\in\mathbb{R}^2}\left\{g_0(y)+L(x-y)\right\} = g_1(x) = L(x-y_\iota)\end{aligned}$$

and the slope of backward optimal rays through $(1,x)$ is

$$\mathbf{b}(1,x) = DH(Du(1,x)) = x - y_\iota.$$

For any two adjacent $y_\iota, y_{\iota'}$ with $\Omega_\iota, \Omega_{\iota'} \subset [-\ell,\ell]^2$ and $x \in \partial\Omega_\iota \cap \partial\Omega_{\iota'}$, we compute that

$$\begin{cases} Du(1,x) & = [DL(x-y_\iota) - DL(x-y_{\iota'})] \otimes \mathbf{n}(x)\mathscr{H}^1_{\llcorner \partial\Omega_\iota \cap \partial\Omega_{\iota'}} \\ \\ D\mathbf{b}(1,x) & = (y_\iota - y_{\iota'}) \otimes \mathbf{n}(x)\mathscr{H}^1_{\llcorner \partial\Omega_\iota \cap \partial\Omega_{\iota'}} \end{cases}$$

and this implies

$$|D\mathbf{b}(1,\cdot)|(\Omega_\iota \cup \Omega_{\iota'}), |Du(1,\cdot)|(\Omega_\iota \cup \Omega_{\iota'}) \geq \delta^{2/3} \cdot \mathscr{H}^1(\partial\Omega_\iota \cap \partial\Omega_{\iota'}) \geq \delta^{4/3}.$$

Since the number of open regions $\Omega_\iota \subset [-\ell,\ell]^2$ is of the order $\dfrac{\ell^2}{\delta^{5/2}}$, we have

$$|D\mathbf{b}(1,\cdot)|([-\ell,\ell]^2), |D(u(1,\cdot))|([-\ell,\ell]^2) \geq C \cdot \frac{\ell^2}{\delta^{5/2}} \cdot \delta^{4/3} = C \cdot \frac{\ell^2}{\delta^{1/3}}$$

for some constant $C > 0$. Thus, choosing $\delta > 0$ sufficiently small, we obtain (3.2.44).

**2.** Let us consider a sequence of disjoint squares $\square_n = c_n + [0, 2^{-n}] \times [0, 2^{-n}]$ such that

$$\bigcup_{n \geq 1} \square_n \subset [0,1]^2.$$

From the previous step, for any $n \geq 1$ we can construct a sequence of Lipschitz functions $\bar{u}_n : \mathbb{R}^2 \to [0,\infty)$ and $\mathrm{supp}(\bar{u}_n) \subset \square_n$ such that the solution $u_n$ of (3.2.1) with initial data $\bar{u}_n$ satisfies

$$|Du_n(1,\cdot)|\left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right), |DH(Du_n(1,\cdot))|\left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right) \geq 1$$

and

$$L(x-z) \geq \min_{y \in \square_n} \left\{ \bar{u}_n(y) + L(x-y) \right\} \quad \text{for all } x \in \left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right), z \in \mathbb{R}^2 \backslash \square_n.$$

Let us set $\bar{u} = \sum_{n=1}^{\infty} u_{0,n}$. The solution $u$ of (3.2.1) with the initial datum $\bar{u}$ satisfies

$$u(1,x) = u_n(1,x) \quad \text{for all } x \in \left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right)$$

and this implies

$$|D\mathbf{b}(1,\cdot)|([0,1]^2) \geq \sum_{n=1}^{\infty} |DH(Du_n(1,\cdot))|(\square_n) \geq \sum_{n=1}^{\infty} 1 = +\infty.$$

Similarly, we have that $|Du(1,\cdot)|([0,1]^2) = +\infty$. $\square$

CHAPTER

# 4

# METRIC ENTROPY FOR CONSERVATION LAWS

This chapter focuses on establishing sharp estimates in Section 4.1 for the metric entropy of a class of bounded total generalized variation functions taking values in a general totally bounded metric space $(E,\rho)$ up to an accuracy of $\varepsilon$ with respect to the $\mathbf{L}^1$- distance, using the notions of doubling and packing dimensions of $(E,\rho)$ introduced in Chapter 2. We also demonstrate how to apply the obtained result in Section 4.2 and provide an upper bound on the metric entropy for a set of entropy admissible weak solutions to scalar conservation laws in one-dimensional space with weakly genuinely nonlinear fluxes.

The first results on $\varepsilon$-entropy in the context of conservation laws were obtained in [5, 23] for the scalar conservation law in one-dimensional space

$$u_t(t,x)+f(u(t,x))_x \;=\; 0 \tag{4.0.1}$$

with uniformly convex flux $f$ (i.e. $f''(u)\geq c>0$) and a similar estimate was obtained for the system of hyperbolic conservation laws in [7, 6]. Thereafter, the results in [5, 23] were extended to scalar conservation laws with a smooth flux function $f$ that is either strictly (but not necessarily uniformly) convex or has a single inflection point with a polynomial degeneracy [8] where entropy admissible weak solutions may have unbounded total variation. In [11, Example 7.2]) it was shown that for fluxes having one inflection point where all derivatives vanish, the composition of the derivative of the flux with the solution of (4.0.1) fails in general to belong to the BV space. However, for *weakly genuinely nonlinear* fluxes, equibounded sets of entropy solutions of (4.0.1) at positive time are still relatively compact in $\mathbf{L}^1$ ([46, Theorem 26]). Thus for fluxes of such classes, a new approach utilizing the uniform bound on total generalized variation of entropy admissible weak solutions derived in [41, Theorem 1], is required to study the $\varepsilon$-entropy for (4.0.1) with weakly genuinely nonlinear fluxes.

Considering the above points of view, we embarked on the following study of $\varepsilon$-entropy of classes of uniformly bounded total generalized variation functions taking values in a general totally bounded metric space $(E,\rho)$.

## 4.1 Metric entropy for generalized BV functions

Throughout this section, the metric space $(E,\rho)$ is assumed to be totally bounded. For convenience, we use the notation

$$\mathbf{H}_\varepsilon := \log_2 \mathbf{N}_\varepsilon \quad \text{and} \qquad \mathbf{K}_\varepsilon := \log_2 \mathbf{M}_\varepsilon$$

where $\mathbf{N}_\varepsilon := \mathcal{N}_\varepsilon(E|E)$ and $\mathbf{M}_\varepsilon := \mathcal{M}_\varepsilon(E|E)$ are the $\varepsilon$-covering and the $\varepsilon$-packing numbers of $E$ in $(E,\rho)$ and

$$\begin{cases} \mathbf{d} & := \mathbf{d}(E) \ \text{ the doubling dimension of } E, \\ \\ \mathbf{p} & := \mathbf{p}(E) \ \text{ the packing dimension of } E. \end{cases}$$

Given two constants $L, V > 0$, we shall establish both upper and lower estimates on the $\varepsilon$-entropy in $\mathbf{L}^1([0,L],E)$ of a class of uniformly bounded $\Psi$-total variation functions defined on $[0,L]$ and taking values in $(E,\rho)$. We denote this class of functions by

$$\mathscr{F}^{\Psi}_{[L,V]} := \left\{ f \in BV^{\Psi}([0,L],E) \;\middle|\; TV^{\Psi}(f,[0,L]) \leq V \right\}. \tag{4.1.1}$$

Our main theorem in this section is as stated below.

**Theorem 55.** *Assume that the function* $\Psi:[0,+\infty)\to[0,+\infty)$ *is convex and satisfies the condition* (2.5.4)*. Then, for every* $0<\varepsilon\le 2L\Psi^{-1}\left(\frac{V}{4}\right)$*, it holds that*

$$\frac{\mathbf{p}V}{2\log_2(7)\cdot\Psi\left(\frac{256\varepsilon}{L}\right)}+\mathbf{K}_{\frac{258\varepsilon}{L}}\ \le\ \mathscr{H}_\varepsilon\left(\mathscr{F}^{\Psi}_{[L,V]}\Big|\mathbf{L}^1([0,L],E)\right)\ \le\ \left[3\mathbf{d}+\log_2(5e)\right]\cdot\frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)}+\mathbf{H}_{\frac{\varepsilon}{4L}}\,. \quad (4.1.2)$$

As a consequence, the minimal number of functions needed to represent a function in $\mathscr{F}^{\Psi}_{[L,V]}$ up to an accuracy $\varepsilon$ with respect to $\mathbf{L}^1$-distance is of the order $\frac{1}{\Psi(O(\varepsilon))}$. Indeed, from (2.3.5) and (2.3.6), it holds that

$$\begin{cases}\mathbf{H}_\varepsilon & \le\ \mathbf{d}\cdot\log_2\left(\mathrm{diam}(E)\cdot\dfrac{2}{\varepsilon}\right)\\ \mathbf{K}_\varepsilon & \ge\ \mathbf{p}\cdot(\log_7 2)\cdot\log_2\left(\mathrm{diam}(E)\cdot\dfrac{1}{2\varepsilon}\right)\end{cases}\qquad \text{for all } \varepsilon>0,$$

and (4.1.2) implies

$$\frac{\mathbf{p}V}{2\log_2(7)\cdot\Psi\left(\frac{256\varepsilon}{L}\right)}+\mathbf{p}\cdot\log_7\left(\mathrm{diam}(E)\cdot\frac{L}{516\varepsilon}\right)\ \le\ \mathscr{H}_\varepsilon\left(\mathscr{F}^{\Psi}_{[L,V]}\ \Big|\ \mathbf{L}^1([0,L],E)\right)$$
$$\le\ \left[3\mathbf{d}+\log_2(5e)\right]\frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)}+\mathbf{d}\cdot\log_2\left(\mathrm{diam(E)}\cdot\frac{8L}{\varepsilon}\right). \quad (4.1.3)$$

On the other hand, as a direct implication of Theorem 55we can also obtain a sharp estimate on the $\varepsilon$-entropy for a class of uniformly bounded $\gamma$-total variation functions, i.e. $\Psi(x)=|x|^\gamma$, for all $\gamma\ge 1$. More precisely, let us denote this class by

$$\mathscr{F}^{\gamma}_{[L,V]}\ =\ \left\{f\in BV^{\frac{1}{\gamma}}([0,L],E)\ \Big|\ TV^{\frac{1}{\gamma}}(f,[0,L])\le V\right\}. \quad (4.1.4)$$

**Corollary 56.** *For every* $0<\varepsilon\le 2^{\frac{\gamma-2}{\gamma}}LV^{\frac{1}{\gamma}}$*,*

$$\frac{\mathbf{p}}{2^{8\gamma+1}\log_2(7)}\cdot\frac{L^\gamma V}{\varepsilon^\gamma}+\mathbf{p}\cdot\log_7\left(\mathrm{diam}(E)\cdot\frac{L}{516\varepsilon}\right)\ \le\ \mathscr{H}_\varepsilon\left(\mathscr{F}^{\gamma}_{[L,V]}\,|\,\mathbf{L}^1([0,L],E)\right)$$
$$\le\ 2^{\gamma+1}\cdot\left[3\mathbf{d}+\log_2(5e)\right]\frac{L^\gamma V}{\varepsilon^\gamma}+\mathbf{d}\cdot\log_2\left(\mathrm{diam(E)}\cdot\frac{8L}{\varepsilon}\right). \quad (4.1.5)$$

In particular, as $\varepsilon$ tends to 0+, we have

$$\frac{\mathbf{p}}{2^{8\gamma+1}\log_2(7)}\ \le\ \liminf_{\varepsilon\to 0+}\left[\frac{\varepsilon^\gamma}{L^\gamma V}\cdot\mathscr{H}_\varepsilon\left(\mathscr{F}^{\gamma}_{[L,V]}\,|\,\mathbf{L}^1([0,L],E)\right)\right]$$
$$\le\ \limsup_{\varepsilon\to 0+}\left[\frac{\varepsilon^\gamma}{L^\gamma V}\cdot\mathscr{H}_\varepsilon\left(\mathscr{F}^{\gamma}_{[L,V]}\,|\,\mathbf{L}^1([0,L],E)\right)\right]\ \le\ 2^{\gamma+1}\left[3\mathbf{d}+\log_2(5e)\right].$$

Thus, the $\varepsilon$-entropy of $\mathscr{F}^{\gamma}_{[L,V]}$ in $\mathbf{L}^1([0,L],E)$ is of the order $\varepsilon^{-\gamma}$.

Finally, to facilitate the application our result in finding $\varepsilon$-entropy estimates for entropy admissible weak solution sets to scalar conservation laws in one-dimensional space with weakly genuinely nonlinear fluxes, we consider the case where the metric space $(E,\rho)$ is generated by a finite dimensional normed space $(\mathbb{R}^d, \|\cdot\|)$, i.e.,

$$E = \mathbb{R}^d \qquad \text{and} \qquad \rho(x,y) = \|x-y\| \qquad \text{for all } x,y \in \mathbb{R}^d.$$

Given an additional constant $M>0$, the following provides upper and lower estimates for the $\varepsilon$-entropy of a class of uniformly bounded $\Psi$-total variation functions

$$\mathscr{F}^{\Psi}_{[L,M,V]} := \left\{ f \in BV^{\Psi}\big([0,L], B^d(0,M)\big) \;\middle|\; TV^{\Psi}(f,[0,L]) \leq V \right\}, \tag{4.1.6}$$

taking values within the open ball $B^d(0,M) \subset \mathbb{R}^d$ in the normed space $\mathbf{L}^1(\mathbb{R}^d)$.

**Corollary 57.** *Under the same assumptions in Theorem 55, for every* $0 < \varepsilon \leq 2L\Psi^{-1}\left(\frac{V}{4}\right)$,

$$\begin{aligned} \frac{Vd}{2\log_2(7)\cdot\Psi\left(\frac{256\varepsilon}{L}\right)} + d\cdot\log_7\left(\frac{LM}{258\varepsilon}\right) &\leq \mathscr{H}_{\varepsilon}\left(\mathscr{F}^{\Psi}_{[L,M,V]} \;\middle|\; \mathbf{L}^1([0,L],\mathbb{R}^d)\right) \\ &\leq \big[3d\log_2 5 + \log_2(5e)\big]\cdot\frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + d\cdot\log_2\left(\frac{8LM}{\varepsilon}+1\right). \end{aligned} \tag{4.1.7}$$

*Proof.* It is well-known (e.g in [33]) that for any $\varepsilon > 0$ and open ball $B^d(0,r) \subset \mathbb{R}^d$,

$$d\cdot\log_2\left(\frac{r}{\varepsilon}\right) \leq \mathscr{H}_{\varepsilon}\left(B^d(0,r)\middle|\mathbb{R}^d\right) \leq d\cdot\log_2\left(\frac{2r}{\varepsilon}+1\right).$$

In particular, recalling that $\mathbf{H}_{\varepsilon} = \log_2 \mathscr{N}_{\varepsilon}\left(B^d(0,M)\middle|\mathbb{R}^d\right)$ and $\mathbf{K}_{\varepsilon} = \log_2 \mathscr{M}_{\varepsilon}\left(B^d(0,M)\middle|\mathbb{R}^d\right)$, we have

$$\mathbf{H}_{\varepsilon} \leq d\cdot\log_2\left(\frac{2M}{\varepsilon}+1\right), \qquad \mathbf{K}_{\varepsilon} \geq \mathbf{H}_{\varepsilon} \geq d\cdot\log_2\left(\frac{M}{\varepsilon}\right)$$

and from Definition 4, it holds that

$$d \leq \mathbf{p}\big(\mathbb{R}^d\big) \leq \mathbf{d}\big(\mathbb{R}^d\big) \leq d\cdot\log_2 5.$$

Using the above estimates in (4.1.2), we obtain (4.1.7). □

In the next two subsections, we present the proof of Theorem 55.

### 4.1.1 Upper estimate

Towards the proof of the upper bound on $\mathscr{H}_\varepsilon\left(\mathscr{F}^{\Psi}_{[L,V]}\,\middle|\,\mathbf{L}^1([0,L],E)\right)$ in Theorem 55, let us extend a result on the $\varepsilon$-entropy for a class of bounded total variation real-valued functions in the scalar case [10] or in [26, Lemma 2.3]. In order to obtain a sharp upper bound, we need to utilize the doubling dimension of the metric space $E$ and go beyond the particular cases in [10, 26] to estimate the $\varepsilon$-entropy for a more general case in $E$. More precisely, considering a set of bounded total variation functions taking values in $E$, denoted by

$$\mathscr{F}_{[L,V]} = \left\{f \in BV([0,L],E)\,\middle|\, TV(f,[0,L]) \leq V\right\}, \tag{4.1.8}$$

the following holds.

**Proposition 58.** *For every* $0 < \varepsilon \leq \dfrac{LV}{2}$ *sufficiently small, it holds that*

$$\mathscr{H}_\varepsilon\left(\mathscr{F}_{[L,V]}\,\middle|\,\mathbf{L}^1([0,L],E)\right) \leq \left[3\mathbf{d}+\log_2(5e)\right]\cdot\frac{2LV}{\varepsilon}+\mathbf{H}_{\frac{\varepsilon}{2L}}.$$

*Proof.* The proof is divided into four steps.

**1.** Given two constants $N_1 \in \mathbb{Z}^+$ and $h_2 > 0$, let us

- divide $[0,L]$ into $N_1$ small intervals $I_i$ with length $h_1 := \dfrac{L}{N_1}$ such that $I_{N_1-1} = [(N_1-1)h_1, L]$ and $I_i = [ih_1,(i+1)h_1)$ for all $i \in \overline{0, N_1-2}$;
- pick an optimal $h_2$-covering $A = \left\{a_1, a_2, \dots, a_{\mathbf{N}_{h_2}}\right\}$ of $E$, i.e., $E \subseteq \bigcup_{i=1}^{\mathbf{N}_{h_2}} B_\rho(a_i,h_2)$, where $\mathbf{N}_{h_2}$ is the $h_2$-covering number of $E$ (see Definition 2).

A function $f \in \mathscr{F}_{[L,V]}$ can be approximated by a piecewise constant function $f^\sharp : [0,L] \to A$ defined as: $f^\sharp(s) = a_{f,i}$ for all $s \in I_i$, $i \in \overline{0,N_1-1}$ for some $a_{f,i} \in A$ chosen such that $f(t_i) \in B_\rho(a_{f,i},h_2)$ with $t_i := \dfrac{2i+1}{2}h_1$. It is to be noted that the choice of $a_{f,i}$ is not unique.

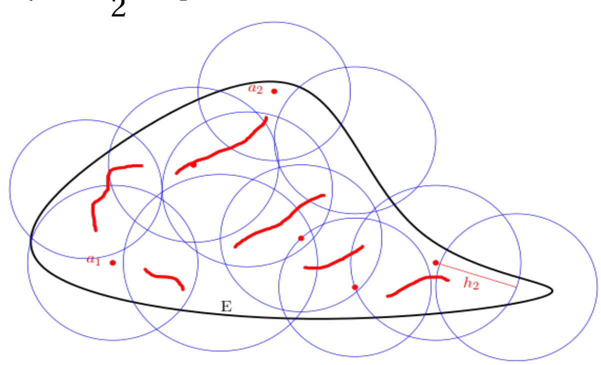


**Figure 4.1** $h_2$-covering of $E$

With this construction, the $\mathbf{L}^1$-distance between $f$ and $f^\sharp$ can be bounded above by

$$\begin{aligned}
\rho_{\mathbf{L}^1}(f,f^\sharp) &\le \sum_{i=0}^{N_1-1}\int_{I_i}\rho(f(s),f^\sharp(s))ds \;=\; \sum_{i=0}^{N_1-1}\int_{I_i}\rho(f(s),a_{f,i})ds \\
&\le \sum_{i=0}^{N_1-1}\int_{I_i}\Big[\rho(f(s),f(t_i))+\rho(f(t_i),a_{f,i})\Big]ds \;<\; \sum_{i=0}^{N_1-1}\int_{I_i}\Big[\rho(f(s),f(t_i))+h_2\Big]ds \\
&\le \left(\sum_{i=0}^{N_1-1}\frac{|I_i|}{2}\cdot\big[TV(f,[ih_1,t_i])+TV(f,[t_i,(i+1)h_1])\big]\right)+Lh_2 \\
&= \frac{h_1}{2}\cdot TV(f,[0,L])+Lh_2 \;\le\; \frac{LV}{2N_1}+Lh_2
\end{aligned}$$

and the total variation of $f^\sharp$ over $[0,L]$ can be estimated by

$$\begin{aligned}
TV\left(f^\sharp,[0,L]\right) &= \sum_{i=0}^{N_1-2}\rho(a_{f,i},a_{f,i+1}) \\
&\le \sum_{i=0}^{N_1-2}\Big[\rho(a_{f,i+1},f(t_{i+1}))+\rho(f(t_i),a_{f,i})+\rho\left(f(t_{i+1}),f(t_i)\right)\Big] \\
&\le \sum_{i=0}^{N_1-2}\Big[2h_2+\rho(f(t_{i+1}),f(t_i))\Big] \;\le\; 2(N_1-1)\cdot h_2+V.
\end{aligned}$$

We consider the following set of piecewise constant functions

$$\mathscr{F}^\sharp_{[N_1,h_2]} \;=\; \Big\{\varphi:[0,L]\to A \;\Big|\; \varphi(s)=\varphi(t_i)\ \text{ for all } s\in I_i,\ i\in\overline{0,N_1-1}$$

$$\text{and}\quad TV(\varphi,[0,L])\le 2(N_1-1)\cdot h_2+V\Big\}.$$

The set $\mathscr{F}_{[L,V]}$ is covered by a finite collection of closed balls centered at $\varphi\in\mathscr{F}^\sharp_{[N_1,h_2]}$ of radius $\frac{LV}{2N_1}+Lh_2$ in $\mathbf{L}^1([0,L],E)$, i.e.,

$$\mathscr{F}_{[L,V]} \subseteq \bigcup_{\varphi\in\mathscr{F}^\sharp_{[N_1,h_2]}} \overline{B}_{\mathbf{L}^1([0,L],E)}\left(\varphi,\frac{LV}{2N_1}+Lh_2\right)$$

and Definition 1 yields

$$\mathscr{H}_{\left[\frac{LV}{2N_1}+Lh_2\right]}\left(\mathscr{F}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right) \;\le\; \log_2\mathrm{Card}\left(\mathscr{F}^\sharp_{[N_1,h_2]}\right). \tag{4.1.9}$$

**2.** In order to provide an upper bound on $\mathrm{Card}\left(\mathscr{F}^\sharp_{[N_1,h_2]}\right)$, we introduce a discrete metric

$\rho^\sharp : A\times A \to \mathbb{N}$ associated to $\rho$ as follows:

$$\rho^\sharp(x,y) := \begin{cases} 0 & \text{if} \quad x = y, \\ \\ q+1 & \text{if} \quad \dfrac{\rho(x,y)}{h_2} \in \big(q, q+1\big] \quad \text{for some } q \in \mathbb{N}\,, \end{cases} \tag{4.1.10}$$

for every $x,\ y\in A$.

Since $A$ is an optimal $h_2$-covering of $E$,

$$\text{Card}\big(A \bigcap B_\rho(a,r)\big) \ \leq\ \mathscr{N}_{h_2}\big(B_\rho(a, r+h_2)\big| E\big) \quad \text{for all } a\in A, r>0$$

and the second inequality in (2.3.5) yields

$$\text{Card}\big(A \bigcap B_\rho(a,r)\big) \ \leq\ \left(2\cdot\left(\frac{r}{h_2}+1\right)\right)^{\mathbf{d}}.$$

Hence, for every $\ell \geq 1$ and $x\in A$, it holds that

$$\begin{aligned} \text{Card}\big(\overline{B}_{\rho^\sharp}(x, \ell-1)\big) \ &=\ \text{Card}\big(\{y\in A \mid \rho^\sharp(x,y) \leq \ell - 1\}\big) \\ &=\ \text{Card}\big(A \bigcap B_\rho\,(x, (\ell-1)h_2)\big) \ \leq\ (2\ell)^{\mathbf{d}}. \end{aligned} \tag{4.1.11}$$

For any given $f^\sharp \in \mathscr{F}^\sharp_{[N_1,h_2]}$, the following increasing step function $\varphi_{f^\sharp} : [0,L] \to \mathbb{N}$ defined by

$$\varphi_{f^\sharp}(s) \ =\ \begin{cases} 0 \quad \text{for all } s\in I_0 \\ \\ \displaystyle\sum_{\ell=0}^{i-1} \rho^\sharp\big(f^\sharp(t_\ell), f^\sharp(t_{\ell+1})\big) + i - 1 \quad \text{for all } s\in I_i,\, i\in \overline{1, N_1-1} \end{cases} \tag{4.1.12}$$

measures the total of jumps of $f^\sharp$ up to time $t_i$. From (4.1.10), we have

$$\begin{aligned} \sup_{t\in[0,L]} \big|\varphi_{f^\sharp}(t)\big| \ &\leq\ \sum_{\ell=0}^{N_1-2} \rho^\sharp\big(f^\sharp(t_\ell), f^\sharp(t_{\ell+1})\big) + N_1 - 2 \\ &\leq\ \sum_{\ell=0}^{N_1-2}\left(\frac{\rho(f^\sharp(t_\ell), f^\sharp(t_{\ell+1}))}{h_2} + 1\right) + N_1 - 2 \ \leq\ \frac{TV(f^\sharp, [0,L])}{h_2} + 2N_1 - 3 \\ &\leq \frac{1}{h_2}\cdot (2(N_1-1)\cdot h_2 + V) + 2N_1 - 3 \ =\ 4N_1 - 5 + \frac{V}{h_2}\,. \end{aligned} \tag{4.1.13}$$

In particular, upon setting $\Gamma_{[N_1,h_2]} := 4N_1 - 4 + \left\lfloor \frac{V}{h_2} \right\rfloor$, a constant depending on $N_1$ and $h_2$, the function $\varphi_{f^\sharp}$ in (4.1.12) satisfies

$$\varphi_{f^\sharp}(s) = \varphi_{f^\sharp}(t_i) \in \left\{0,1,2,\dots,\Gamma_{[N_1,h_2]}-1\right\} \quad \text{for all } s \in I_i,\ i \in \overline{0, N_1-1}\,.$$

Thus, if we consider the map $T : \mathscr{F}^\sharp_{[N_1,h_2]} \to \mathscr{B}([0,L],[0,+\infty))$ such that

$$T(f^\sharp) = \varphi_{f^\sharp} \qquad \text{for all } f^\sharp \in \mathscr{F}^\sharp_{[N_1,h_2]}\,,$$

then

$$T\left(\mathscr{F}^\sharp_{[N_1,h_2]}\right) = \left\{\varphi_{f^\sharp} \,\middle|\, f^\sharp \in \mathscr{F}^\sharp_{[N_1,h_2]}\right\} \subseteq \mathscr{I}_{[N_1,h_2]}\,.$$

Here, $\mathscr{I}_{[N_1,h_2]}$ is the set of increasing step functions $\phi : [0,L] \to \left\{0,1,2,\dots,\Gamma_{[N_1,h_2]}-1\right\}$ such that

$$\phi(0) = 0 \quad \text{and} \quad \phi(s) = \phi(t_i) \quad \text{for all } i \in \overline{0, N_1-1},\ s \in I_i\,.$$

$$\mathrm{Card}\left(T\left(\mathscr{F}^\sharp_{[N_1,h_2]}\right)\right) \le \mathrm{Card}(\mathscr{I}_{[N_1,h_2]}) = \binom{\Gamma_{[N_1,h_2]}}{N_1-1}. \tag{4.1.14}$$

**3.** To complete the proof, we need to establish an upper estimate on the cardinality of $T^{-1}(\varphi_{f^\sharp})$, the set of functions in $\mathscr{F}^\sharp_{[N_1,h_2]}$ that have the same total length of jumps as that of $f^\sharp$ at any time $t_i$. In order to do so, for any given $f^\sharp \in \mathscr{F}^\sharp_{[N_1,h_2]}$, we set

$$k_i^\sharp := \rho^\sharp\left(f^\sharp(t_i), f^\sharp(t_{i+1})\right) \quad \text{for all } i \in \overline{0, N_1-2}\,.$$

As in (4.1.13), we have $\sum_{i=0}^{N_1-2} k_i^\sharp = \displaystyle\sum_{i=0}^{N_1-2} \rho^\sharp\left(f^\sharp(t_i), f^\sharp(t_{i+1})\right) \le 3(N_1-1) + \frac{V}{h_2}$ and

$$\begin{aligned} T^{-1}(\varphi_{f^\sharp}) &= \left\{g \in \mathscr{F}^\sharp_{[N_1,h_2]} \,\middle|\, \rho^\sharp\left(g(t_{i+1}), g(t_i)\right) = k_i^\sharp \text{ for all } i \in \overline{0, N_1-2}\right\} \\ &\subseteq \left\{g \in \mathscr{F}^\sharp_{[N_1,h_2]} \,\middle|\, g(t_{i+1}) \in \overline{B}_{\rho^\sharp}\left(g(t_i), k_i^\sharp\right) \text{for all } i \in \overline{0, N_1-2}\right\}. \end{aligned}$$

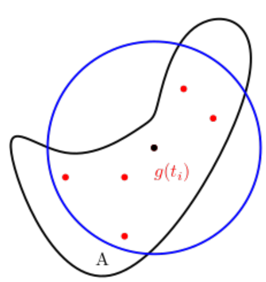


**Figure 4.2** Estimation of $\mathrm{Card}\left(T^{-1}(\varphi_{f^\sharp})\right)$

We observe from (4.1.11) that if $g(t_i)$ is already chosen then there are at most $(2k_i^\sharp)^{\mathbf{d}}$ choices for $g(t_{i+1})$. Since there are $\mathbf{N}_{h_2}$ choices of the starting point $g(0)$, the cardinality of $T^{-1}(\varphi_{f^\sharp})$ can be estimated as follows

$$\begin{aligned}\mathrm{Card}\big(T^{-1}(\varphi_{f^\sharp})\big) \;&\leq\; \mathbf{N}_{h_2}\cdot\Pi_{i=0}^{N_1-2}(2k_i^\sharp)^{\mathbf{d}} \;\leq\; \mathbf{N}_{h_2}\cdot\left(\frac{\sum_{i=0}^{N_1-2}2k_i^\sharp}{N_1-1}\right)^{\mathbf{d}(N_1-1)}\\ &\leq\; \mathbf{N}_{h_2}\cdot\left(\frac{2\big(3(N_1-1)+\frac{V}{h_2}\big)}{N_1-1}\right)^{\mathbf{d}(N_1-1)} \;=\; \mathbf{N}_{h_2}\cdot\left(6+\frac{2}{N_1-1}\cdot\frac{V}{h_2}\right)^{\mathbf{d}(N_1-1)}. \quad (4.1.15)\end{aligned}$$

Recalling (4.1.14)-(4.1.15) and the classical Stirling's approximation

$$(N_1-1)! \;\geq\; \sqrt{2\pi(N_1-1)}\cdot\left(\frac{N_1-1}{e}\right)^{N_1-1},$$

we estimate

$$\begin{aligned}\mathrm{Card}\Big(\mathscr{F}^\sharp_{[N_1,h_2]}\Big) \;&\leq\; \mathbf{N}_{h_2}\cdot\left(6+\frac{2}{N_1-1}\cdot\frac{V}{h_2}\right)^{\mathbf{d}(N_1-1)}\cdot\binom{\Gamma_{[N_1,h_2]}}{N_1-1}\\ &=\; \mathbf{N}_{h_2}\cdot\left(6+\frac{2}{N_1-1}\cdot\frac{V}{h_2}\right)^{\mathbf{d}(N_1-1)}\cdot\frac{\big(\Gamma_{[N_1,h_2]}-N_1+2\big)\ldots\Gamma_{[N_1,h_2]}}{(N_1-1)!}\\ &\leq\; \frac{\mathbf{N}_{h_2}}{\sqrt{2\pi(N_1-1)}}\cdot\left(6+\frac{2}{N_1-1}\cdot\frac{V}{h_2}\right)^{\mathbf{d}(N_1-1)}\cdot\left(\frac{\Gamma_{[N_1,h_2]}}{N_1-1}\right)^{N_1-1}\cdot e^{N_1-1}\\ &\leq\; \mathbf{N}_{h_2}\cdot\left(6+\frac{2}{N_1-1}\cdot\frac{V}{h_2}\right)^{\mathbf{d}(N_1-1)}\cdot\left(4e+\frac{V}{h_2}\cdot\frac{e}{N_1-1}\right)^{(N_1-1)}.\end{aligned}$$

Thus, (4.1.9) yields

$$\begin{aligned}\mathscr{H}_{\left[\frac{LV}{2N_1}+Lh_2\right]}\Big(\mathscr{F}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\Big) \;&\leq\; \mathbf{d}\cdot(N_1-1)\cdot\log_2\left(6+\frac{V}{h_2}\cdot\frac{2}{N_1-1}\right)\\ &\quad+(N_1-1)\cdot\log_2\left(4e+\frac{V}{h_2}\cdot\frac{e}{N_1-1}\right)+\mathbf{H}_{h_2}. \quad (4.1.16)\end{aligned}$$

**4.** For every $0<\varepsilon\leq\dfrac{LV}{2}$, by choosing $N_1\in\mathbb{Z}^+$ and $h_2>0$ such that

$$\frac{3LV}{2\varepsilon} \;<\; N_1-1 \;=\; \left\lfloor\frac{3LV}{2\varepsilon}\right\rfloor+1 \;\leq\; \frac{2LV}{\varepsilon}, \qquad h_2 \;=\; \frac{V}{N_1-1},$$

we have $\frac{LV}{2N_1}+Lh_2 \;\leq\; \frac{LV}{2N_1}+\frac{LV}{N_1-1} \;\leq\; \frac{3LV}{2(N_1-1)} \;<\; \varepsilon$ and $h_2 \;\geq\; \frac{\varepsilon}{2L}$.

Thus, (4.1.16) implies that

$$\mathcal{H}_\varepsilon\left(\mathcal{F}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right) \le \left[3\mathbf{d}+\log_2(5e)\right]\cdot\frac{2LV}{\varepsilon}+\mathbf{H}_{\frac{\varepsilon}{2L}}$$

and this completes the proof. □

Using Proposition 58, we now proceed to provide a proof for the upper estimate of the $\varepsilon$-entropy for the set $\mathcal{F}^{\Psi}_{[L,V]}$ in $\mathbf{L}^1([0,L],E)$.

**Proof of the upper estimate in Theorem 55.** From Lemma 19, we have

$$\mathcal{H}_\varepsilon\left(\mathcal{F}^{\Psi}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right) = \mathcal{H}_\varepsilon\left(\tilde{\mathcal{F}}^{\Psi}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right) \tag{4.1.17}$$

with $\tilde{\mathcal{F}}^{\Psi}_{[L,V]} = \left\{f\in\mathcal{F}^{\Psi}_{[L,V]}\,\Big|\, f \text{ is continuous from the right on the interval } [0,L)\right\}$. Thus, it is sufficient to prove the second inequality in (4.1.2) for $\tilde{\mathcal{F}}^{\Psi}_{[L,V]}$ instead of $\mathcal{F}^{\Psi}_{[L,V]}$.

**1.** For a fixed constant $h>0$ and $f\in\tilde{\mathcal{F}}^{\Psi}_{[L,V]}$, let $A_{f,h}=\left\{x_0,x_1,x_2,\dots,x_{N_{f,h}}\right\}$ be a partition of $[0,L]$ which is defined by induction as follows:

$$x_0 = 0, \qquad x_{i+1} = \sup\left\{x\in(x_i,L)\,\big|\,\rho(f(y),f(x_i))\in[0,h] \quad \text{for all } y\in(x_i,x]\right\} \tag{4.1.18}$$

for all $i\in\overline{0,N_{f,h}-1}$. Since $f$ is continuous from the right on $[0,L)$, it holds that

$$\rho(f(x_i),f(x_{i+1})) \ge h \qquad \text{for all } i\in\overline{0,N_{f,h}-2}.$$

Thus, the increasing property of $\Psi$ implies that

$$V \ge TV^{\Psi}(f,[0,L]) \ge \sum_{i=0}^{N_{f,h}-2}\Psi\big(\rho(f(x_i),f(x_{i+1}))\big) \ge (N_{f,h}-1)\cdot\Psi(h)$$

and this yields

$$N_{f,h}-1 \le \frac{TV^{\Psi}(f,[0,L])}{\Psi(h)} \le \frac{V}{\Psi(h)} < +\infty\,. \tag{4.1.19}$$

**2.** We introduce a piecewise constant function $f_h:[0,L]\to E$ such that

$$f_h(x) = \begin{cases} f(x_i) & \text{for all } x\in[x_i,x_{i+1})\,,\ i\in\overline{0,N_{f,h}-2} \\[2ex] f\left(x_{N_{f,h}-1}\right) & \text{for all } x\in\left[x_{N_{f,h}-1},L\right]. \end{cases}$$

From (4.1.18), the $\mathbf{L}^1$-distance between $f_h$ and $f$ is bounded by

$$\rho_{\mathbf{L}^1}(f_h,f) = \int_{[0,L]} \rho(f_h(x),f(x))dx = \sum_{i=0}^{N_{f,h}-1}\int_{[x_i,x_{i+1})}\rho(f(x_i),f(x))dx$$
$$\leq h\cdot\sum_{i=0}^{N_{f,h}-1}(x_{i+1}-x_i) = Lh. \quad (4.1.20)$$

On the other hand, by the convexity of $\Psi$ we have

$$V \geq \sum_{i=0}^{N_{f,h}-2}\Psi\big(\rho(f(x_i),f(x_{i+1}))\big) \geq \big(N_{f,h}-1\big)\cdot\Psi\left(\frac{1}{N_{f,h}-1}\cdot\sum_{i=0}^{N_{f,h}-2}\rho(f(x_i),f(x_{i+1}))\right)$$
$$= \big(N_{f,h}-1\big)\cdot\Psi\left(\frac{TV(f_h,[0,L))}{N_{f,h}-1}\right)$$

and the strictly increasing property of $\Psi^{-1}$ implies

$$TV(f_h,[0,L)) \leq (N_{f,h}-1)\cdot\Psi^{-1}\left(\frac{V}{N_{f,h}-1}\right).$$

From Remark 20 and (4.1.19), it holds that

$$\Psi^{-1}\left(\frac{V}{N_{f,h}-1}\right)\cdot\frac{N_{f,h}-1}{V} \leq \Psi^{-1}(\Psi(h))\cdot\frac{1}{\Psi(h)} = \frac{h}{\Psi(h)}$$

and this yields

$$TV(f_h,[0,L]) \leq \frac{h}{\Psi(h)}\cdot V =: V_h\,.$$

From (4.1.20) and (4.1.8), the set $\tilde{\mathscr{F}}^{\Psi}_{[L,V]}$ is covered by a collection of closed balls centered at $g\in\mathscr{F}_{[L,V_h]}$ of radius $Lh$ in $\mathbf{L}^1([0,L],E)$, i.e.,

$$\tilde{\mathscr{F}}^{\Psi}_{[L,V]} \subseteq \bigcup_{g\in\mathscr{F}_{[L,V_h]}}\overline{B}_{\mathbf{L}^1([0,L],E)}(g,Lh)\,.$$

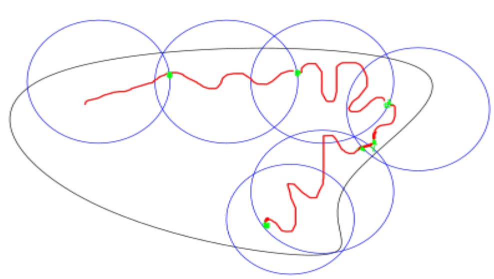

**Figure 4.3** $Lh$-covering of $\tilde{\mathscr{F}}^{\Psi}_{[L,V]}$ with balls centered at $g\in\mathscr{F}_{[L,V_h]}$

In particular, for every $\varepsilon > 0$, choosing $h = \frac{\varepsilon}{2L}$ we have

$$V_{\frac{\varepsilon}{2L}} = \frac{\varepsilon V}{2L \cdot \Psi\left(\frac{\varepsilon}{2L}\right)} \qquad \text{and} \qquad \tilde{\mathscr{F}}^{\Psi}_{[L,V]} \subseteq \bigcup_{g \in \mathscr{F}_{\left[L, V_{\frac{\varepsilon}{2L}}\right]}} \overline{B}_{\mathbf{L}^1([0,L],E)}\left(g, \frac{\varepsilon}{2}\right)$$

and this implies

$$\mathscr{H}_{\varepsilon}\left(\tilde{\mathscr{F}}^{\Psi}_{[L,V]} \,\Big|\, \mathbf{L}^1([0,L],E)\right) \le \mathscr{H}_{\frac{\varepsilon}{2}}\left(\mathscr{F}_{\left[L, V_{\frac{\varepsilon}{2L}}\right]} \,\Big|\, \mathbf{L}^1([0,L],E)\right). \tag{4.1.21}$$

If $0 < \varepsilon \le 2L\Psi^{-1}\left(\dfrac{V}{4}\right)$ then $\varepsilon \le \varepsilon \cdot \dfrac{V}{4 \cdot \Psi\left(\dfrac{\varepsilon}{2L}\right)} = \dfrac{L}{2} \cdot \dfrac{\varepsilon V}{2L \cdot \Psi\left(\dfrac{\varepsilon}{2L}\right)} = \dfrac{L}{2} \cdot V_{\frac{\varepsilon}{2L}}$.

In this case, we can apply Proposition 58 to get

$$\begin{aligned}
\mathscr{H}_{\frac{\varepsilon}{2}}\left(\mathscr{F}_{\left[L, V_{\frac{\varepsilon}{2L}}\right]} \,\Big|\, \mathbf{L}^1([0,L],E)\right) &\le \left[3\mathbf{d} + \log_2(5e)\right] \cdot \frac{4LV_{\frac{\varepsilon}{2L}}}{\varepsilon} + \mathbf{H}_{\frac{\varepsilon}{4L}} \\
&= \left[3\mathbf{d} + \log_2(5e)\right] \cdot \frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + \mathbf{H}_{\frac{\varepsilon}{4L}}
\end{aligned}$$

and thereafter, we use (4.1.17), (4.1.21) to obtain the second inequality in (4.1.2). □

### 4.1.2 Lower estimate

To prove the first inequality in Theorem 55, let us provide a lower estimate on the $\varepsilon$-entropy in $\mathbf{L}^1([0,L],E)$ to

$$\mathscr{G}^{\Psi}_{[L,V,h,x]} := \left\{ g : [0,L] \to B_{\rho}(x,h) \,\Big|\, TV^{\Psi}(g,[0,L]) \le V \right\}, \tag{4.1.22}$$

a class of bounded $\Psi$-total variation functions over $[0,L]$ taking values in the ball centered at a point $x \in E$ of radius $h > 0$.

**Lemma 59.** *Assume that* $\mathbf{p} \ge 1$. *For every* $\varepsilon > 0$*, it holds that*

$$\mathscr{M}_{\varepsilon}\left(\mathscr{G}^{\Psi}_{\left[L,V,2^{(4+2/\overline{\mathbf{p}})} \cdot \frac{\varepsilon}{L}, x\right]} \,\Big|\, \mathbf{L}^1([0,L],E)\right) \ge 2^{\frac{\overline{\mathbf{p}}V}{2\Psi\left(2^{(4+2/\overline{\mathbf{p}})} \cdot \frac{2\varepsilon}{L}\right)}} \tag{4.1.23}$$

*where* $\overline{\mathbf{p}} = \log_7(2) \cdot \mathbf{p}$.

*Proof.* The proof is divided into two steps.

**1.** First we recall from (2.3.6) that

$$\mathscr{M}_{2^{-(2+2/\overline{\mathbf{p}})}\cdot h}(B_\rho(x,h)|E) \geq \left(\frac{h}{2\cdot 2^{-(2+2/\overline{\mathbf{p}})}\cdot h}\right)^{\overline{\mathbf{p}}} = 2^{\overline{\mathbf{p}}+2} \qquad \text{for all } h>0\,.$$

Given two constants $h>0$ and $N_1\in\mathbb{Z}^+$, let us

- divide $[0,L]$ into $N_1$ small mutually disjoint intervals $I_i$ with length $h_1 = \dfrac{L}{N_1}$ as in Proposition 58;
- take a $\left(2^{-(2+2/\overline{\mathbf{p}})}\cdot h\right)$–packing $A_h = \{a_1,a_2,\dots,a_{2^{\overline{\mathbf{p}}+2}}\}$ of $B_\rho(x,h)$, i.e.,

  $$A_h \subseteq B_\rho(x,h) \qquad \text{and} \qquad \rho(a_i,a_j) > 2^{-(2+2/\overline{\mathbf{p}})}\cdot h$$

  for all $a_i\neq a_j\in A_h$ .

Considering the set of indices

$$\Delta_{h,N_1} = \left\{\delta = (\delta_i)_{i\in\{0,1,\,\cdots,\,N_1-1\}} \,\Big|\, \delta_i\in A_h\right\}$$

we define a class of piecewise constant functions on $[0,L]$ as

$$\mathscr{G}_{h,N_1} = \left\{g_\delta = \sum_{i=0}^{N_1-1}\delta_i\cdot\chi_{I_i} \,\Big|\, \delta\in\Delta_{h,N_1}\right\}.$$

For any $\delta\in\Delta_{h,N_1}$, the $\Psi$-total variation of $g_\delta$ is bounded by $TV^\Psi(g_\delta,[0,L]) \leq (N_1-1)\cdot\Psi(2h)$.

Hence, under the following condition on $h$ and $V$ given by

$$(N_1-1)\cdot\Psi(2h) \leq V, \tag{4.1.24}$$

the definition of $\mathscr{G}^\Psi_{[L,V,h,x]}$ in (4.1.22) implies that $g_\delta\in\mathscr{G}^\Psi_{[L,V,h,x]}$ for every $\delta\in\Delta_{h,N_1}$ and thus

$$\mathscr{G}_{h,N_1} \subseteq \mathscr{G}^\Psi_{[L,V,h,x]}\,.$$

In particular, we get

$$\mathscr{M}_\varepsilon\left(\mathscr{G}^\Psi_{[L,V,h,x]} \,\Big|\, \mathbf{L}^1([0,L],E)\right) \geq \mathscr{M}_\varepsilon\left(\mathscr{G}_{h,N_1} \,\Big|\, \mathbf{L}^1([0,L],E)\right) \quad \text{for all } \varepsilon>0\,. \tag{4.1.25}$$

**2.** Let us provide a lower bound on the $\varepsilon$-packing number $\mathscr{M}_\varepsilon\left(\mathscr{G}_{h,N_1} \,\Big|\, \mathbf{L}^1([0,L],E)\right)$. For any given $\delta,\tilde{\delta}\in\Delta_{h,N_1}$ and $\varepsilon>0$, we define

$$\mathscr{I}_{\tilde{\delta}}(2\varepsilon) = \Big\{\delta\in\Delta_{h,N_1}\ \Big|\ \rho_{\mathbf{L}^1}(g_\delta,g_{\tilde{\delta}})\le 2\varepsilon\Big\},\quad \eta(\delta,\tilde{\delta}) = \mathrm{Card}\Big(\big\{i\in\overline{0,N_1-1}\ \big|\ \delta_i\neq\tilde{\delta}_i\big\}\Big).$$

The $\mathbf{L}^1$-distance between $g_\delta$ and $g_{\tilde{\delta}}$ is bounded below by

$$\begin{aligned}\rho_{\mathbf{L}^1}(g_\delta,g_{\tilde{\delta}}) &= \sum_{i=0}^{N_1-1}\int_{I_i}\rho(g_\delta(t),g_{\tilde{\delta}}(t))\,dt = \sum_{i=0}^{N_1-1}\rho(\delta_i,\tilde{\delta}_i)\cdot|I_i|\\ &= \frac{L}{N_1}\cdot\sum_{i=0}^{N_1-1}\rho(\delta_i,\tilde{\delta}_i) > 2^{-(2+2/\overline{\mathbf{p}})}\cdot\frac{Lh}{N_1}\cdot\eta(\delta,\tilde{\delta})\end{aligned}$$

and this implies the inclusion

$$\mathscr{I}_{\tilde{\delta}}(2\varepsilon) \subseteq \left\{\delta\in\Delta_{h,N_1}\ \middle|\ \eta(\delta,\tilde{\delta})<\frac{2^{3+2/\overline{\mathbf{p}}}N_1\varepsilon}{Lh}\right\}. \tag{4.1.26}$$

On the other hand, for every $r\in\overline{0,N_1-1}$, we compute

$$\mathrm{Card}\Big(\Big\{\delta\in\Delta_{h,N_1}\ \Big|\ \eta(\delta,\tilde{\delta})=r\Big\}\Big) = \binom{N_1}{r}\cdot\big(2^{\overline{\mathbf{p}}+2}-1\big)^r.$$

Thus, (4.1.26) implies that

$$\mathrm{Card}(\mathscr{I}_{\tilde{\delta}}(2\varepsilon)) \le \mathrm{Card}\left(\left\{\delta\in\Delta_{h,N_1}\ \middle|\ \eta(\delta,\tilde{\delta})<\frac{2^{3+2/\overline{\mathbf{p}}}N_1\varepsilon}{Lh}\right\}\right) \le \sum_{r=0}^{\left\lfloor\frac{2^{3+2/\overline{\mathbf{p}}}N_1\varepsilon}{Lh}\right\rfloor}\binom{N_1}{r}\cdot\big(2^{\overline{\mathbf{p}}+2}-1\big)^r.$$

In particular, for every $0<\varepsilon\le 2^{-(4+2/\overline{\mathbf{p}})}Lh$, we have

$$\begin{aligned}\mathrm{Card}(\mathscr{I}_{\tilde{\delta}}(2\varepsilon)) &\le \sum_{r=0}^{\lfloor\frac{N_1}{2}\rfloor}\binom{N_1}{r}\cdot\big(2^{\overline{\mathbf{p}}+2}-1\big)^r \le \big(2^{\overline{\mathbf{p}}+2}-1\big)^{\frac{N_1}{2}}\cdot\sum_{r=0}^{\lfloor\frac{N_1}{2}\rfloor}\binom{N_1}{r}\\ &\le 2^{(\overline{\mathbf{p}}+2)\frac{N_1}{2}}\cdot 2^{N_1} = 2^{N_1(2+\overline{\mathbf{p}}/2)}.\end{aligned} \tag{4.1.27}$$

Recalling Definition 2, we then obtain

$$\mathscr{M}_\varepsilon\Big(\mathscr{G}_{h,N_1}\ \Big|\ \mathbf{L}^1([0,L],E)\Big) \ge \frac{\mathrm{Card}\big(\mathscr{G}_{h,N_1}\big)}{\mathrm{Card}(\mathscr{I}_{\tilde{\delta}}(2\varepsilon))} \ge \frac{2^{N_1(\overline{\mathbf{p}}+2)}}{2^{N_1(2+\overline{\mathbf{p}}/2)}} = 2^{N_1\overline{\mathbf{p}}/2}.$$

Finally, by choosing $h=2^{(4+2/\overline{\mathbf{p}})}\cdot\dfrac{\varepsilon}{L}$ and $N_1=\left\lfloor\dfrac{V}{\Psi(2^{(4+2/\overline{\mathbf{p}})}\cdot\frac{2\varepsilon}{L})}\right\rfloor+1$ such that (4.1.24) holds, we derive

$$\mathscr{M}_\varepsilon\left(\mathscr{G}_{2^{(4+2/\overline{\mathbf{p}})}\cdot\frac{\varepsilon}{L},N_1}\,\Big|\,\mathbf{L}^1([0,L],E)\right) \geq 2^{\frac{\overline{\mathbf{p}}V}{2\Psi\left(2^{(4+2/\overline{\mathbf{p}})}\cdot\frac{2\varepsilon}{L}\right)}}$$

and thereafter, (4.1.25) yields (4.1.23). □

We conclude this section by proving the first inequality in (4.1.2).

**Proof of the lower bound in Theorem 55.** For any $0 < 2h < h_2$, let $\{x_1, x_2, \dots, x_{\mathbf{M}_{h_2}}\} \subseteq E$ be an $h_2$-packing of $E$ with size $\mathbf{M}_{h_2}$, i.e.,

$$B_\rho\left(x_i, \frac{h_2}{2}\right)\bigcap B_\rho\left(x_j, \frac{h_2}{2}\right) = \emptyset \quad \text{for all } i \neq j \in \overline{1,\mathbf{M}_{h_2}}\,.$$

Recalling the definition of $\mathscr{G}^\Psi_{[L,V,h,x]}$ in (4.1.22), we have

$$\rho_{\mathbf{L}^1}(f_i,f_j) \;\geq\; \int_{[0,L]}\Big[\rho(x_i,x_j)-\rho(x_i,f_i(s))-\rho(x_j,f_j(s))\Big]ds \,\geq\, L\cdot(h_2-2h) =: L_{h,h_2}$$

for any $f_i \in \mathscr{G}^\Psi_{[L,V,h,x_i]}$ and $f_j \in \mathscr{G}^\Psi_{[L,V,h,x_j]}$ with $i \neq j \in \overline{1,\mathbf{M}_{h_2}}$.

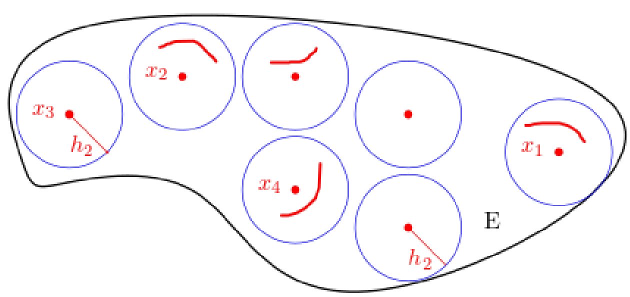


**Figure 4.4** $h_2$-packing of $E$

Thus, Lemma 3 implies that

$$\begin{aligned}
\mathscr{N}_{\frac{L_{h,h_2}}{2}}\left(\mathscr{F}^\Psi_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right) \;&\geq\; \mathscr{M}_{L_{h,h_2}}\left(\mathscr{F}^\Psi_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right)\\
&\geq\; \mathscr{M}_{L_{h,h_2}}\left(\bigcup_{i=1}^{\mathbf{M}_{h_2}}\mathscr{G}^\Psi_{[L,V,h,x_i]}\,\Big|\,\mathbf{L}^1([0,L],E)\right)\\
&=\; \sum_{i=1}^{\mathbf{M}_{h_2}}\mathscr{M}_{L_{h,h_2}}\left(\mathscr{G}^\Psi_{[L,V,h,x_i]}\,\Big|\,\mathbf{L}^1([0,L],E)\right).
\end{aligned}$$

We consider the following two cases.

• If $\mathbf{p}=0$ then by choosing $h=\dfrac{\varepsilon}{L}$ and $h_2=\dfrac{4\varepsilon}{L}$ such that $L_{h,h_2}=2\varepsilon$, we have

$$\mathcal{N}_{\varepsilon}\left(\mathscr{F}^{\Psi}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right)\ \geq\ \mathbf{M}_{\frac{4\varepsilon}{L}}$$

and this particularly implies the first inequality in (4.1.2).

• Otherwise if $\mathbf{p}\geq 1$, then for any $\varepsilon>0$, choosing $h=2^{(5+2/\overline{\mathbf{p}})}\cdot\dfrac{\varepsilon}{L}$ and $h_2=\left(2+2^{(6+2/\overline{\mathbf{p}})}\right)\cdot\dfrac{\varepsilon}{L}$ with $\overline{\mathbf{p}}=\log_7(2)\cdot\mathbf{p}$ such that $L_{h,h_2}=2\varepsilon$, we can apply (4.1.23) to $\mathscr{G}^{\Psi}_{[L,V,h,x_i]}$ for every $i\in\overline{1,\mathbf{M}_{h_2}}$ to obtain

$$\mathcal{N}_{\varepsilon}\left(\mathscr{F}^{\Psi}_{[L,V]}\,\Big|\,\mathbf{L}^1([0,L],E)\right)\ \geq\ \sum_{i=1}^{\mathbf{M}_{\left(2+2^{(6+2/\overline{\mathbf{p}})}\right)\cdot\frac{\varepsilon}{L}}}\mathscr{M}_{2\varepsilon}\left(\mathscr{G}^{\Psi}_{\left[L,V,2^{(4+2/\overline{\mathbf{p}})}\cdot\frac{2\varepsilon}{L},x_i\right]}\,\Big|\,\mathbf{L}^1([0,L],E)\right)$$

$$\geq\ \mathbf{M}_{\left(2+2^{(6+2/\overline{\mathbf{p}})}\right)\cdot\frac{\varepsilon}{L}}\cdot 2^{\overline{\frac{\overline{\mathbf{p}}V}{2\Psi\left(2^{(6+2/\overline{\mathbf{p}})}\cdot\frac{\varepsilon}{L}\right)}}}\ \geq\ \mathbf{M}_{\frac{258\varepsilon}{L}}\cdot 2^{\overline{\frac{\overline{\mathbf{p}}V}{2\Psi\left(\frac{256\varepsilon}{L}\right)}}}$$

and this yields the first inequality in (4.1.2).

□

## 4.2 Application to scalar conservation law

In this section, we use Theorem 55 and [41, Theorem 1] to establish an upper bound on the $\varepsilon$-entropy of a set of entropy admissible weak solutions for a scalar conservation law in one-dimensional space

$$u_t(t,x)+f(u(t,x))_x=0\qquad\text{for all }(t,x)\in(0,+\infty)\times\mathbb{R}\tag{4.2.1}$$

with weakly genuinely nonlinear flux $f\in\mathscr{C}^2(\mathbb{R})$, i.e., which is not affine on any open interval such that the set

$$\{u\in\mathbb{R}\mid f''(u)\neq 0\}\ \text{ is dense in }\ \mathbb{R}.\tag{4.2.2}$$

We recall that the equation (4.2.1) does not possess classical solutions since discontinuities arise in finite time even if the initial data are smooth. Hence, it is natural to consider weak solutions in the sense of distributions that, for the sake of uniqueness, satisfy an entropy admissibility criterion ([22, 34]) equivalent to the celebrated Oleinik E-condition ([44]) which generalizes the classical stability conditions mentioned in Section 2.6 that were introduced by Lax ([36]). This condition is stated as follows.

**Oleinik E-condition:** *A shock discontinuity located at $x$ and connecting a left state*

*$u^L := u(t,x-)$ with a right state $u^R := u(t,x+)$ is entropy admissible if and only if it holds that*

$$\frac{f(u^L)-f(u)}{u^L-u} \geq \frac{f(u^R)-f(u)}{u^R-u}$$

*for every $u$ between $u^L$ and $u^R$, where $u(t,x\pm)$ denote the one-sided limits of $u(t,\cdot)$ at $x$.*
It is well-known that the equation (4.2.1) generates an $\mathbf{L}^1$-contractive semigroup of solutions $(S_t)_{t\geq 0}$ that associates, to every given initial data $u_0 \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^\infty(\mathbb{R})$, the unique entropy admissible weak solution $S_t u_0 := u(t,\cdot)$ of the corresponding Cauchy problem (cfr. [22, 34]). For any given $T, L, M > 0$, we provide an upper bound for $\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]})\middle|\mathbf{L}^1(\mathbb{R})\right)$ with

$$\mathscr{U}_{[L,M]} := \left\{u_0 \in \mathbf{L}^\infty(\mathbb{R}) \,\middle|\, \mathrm{Supp}\,(u_0) \subset [-L,L]\,,\ \|u_0\|_{\mathbf{L}^\infty(\mathbb{R})} \leq M\right\},$$

the set of bounded, compactly supported initial data.

By monotonicity of the solution operator $S_t$ and recalling that $S_t u_0$ can be obtained as a limit of piecewise constant front tracking approximations ([16, Chapter 6]), it was shown in [8, Lemma 2.2] that

**Lemma 60.** *For every $L, M, T > 0$ and $u_0 \in \mathscr{U}_{[L,M]}$, it holds that*

$$\left\|S_T u_0\right\|_{\mathbf{L}^\infty(\mathbb{R})} \leq M \qquad \text{and} \qquad \mathrm{Supp}(S_T u_0) \subseteq \left[-\ell_{[L,M,T]}, \ell_{[L,M,T]}\right]$$

*where* $\quad \ell_{[L,M,T]} := L + T \cdot f'_M \qquad$ and $\qquad f'_M := \sup_{|v|\leq M} |f'(v)|.$

Let us introduce the function $\mathfrak{d} : [0,+\infty) \to [0,+\infty)$ such that

$$\mathfrak{d}(h) = \min_{a\in[-M,M-h]} \left(\inf_{g\in\mathscr{A}_{[a,a+h]}} \|f-g\|_{\mathbf{L}^\infty([a,a+h])}\right)$$

with $\mathscr{A}_{[a,a+h]}$ being the set of affine functions defined on $[a, a+h]$. The convex envelop $\Phi$ of $\mathfrak{d}$ is defined by

$$\Phi = \sup_{\varphi\in\mathscr{G}} \varphi \quad \text{with} \quad \mathscr{G} := \{\varphi : [0,+\infty) \to [0,+\infty) \mid \varphi \text{ is convex},\ \varphi(0) = 0,\ \varphi \leq \mathfrak{d}\}.$$

The following function

$$\Psi(x) := \Phi(x/2) \cdot x \qquad \text{for all } x \in [0+\infty)$$

is convex and satisfies the condition (2.5.4). As a consequence of [41, Theorem 1], the following holds.

**Lemma 61.** *For any $u_0 \in \mathscr{U}_{[L,M]}$, the function $S_T u_0$ has bounded $\Psi$-total variation on $\mathbb{R}$ and*

$$TV^{\Psi}(S_T u_0, \mathbb{R}) \leq \gamma_{[L,M,T]} := \gamma_{[L,M]}\left(1 + \frac{1}{T}\right)$$

*where $\gamma_{[L,M]}$ is a constant depending only on $L, M$ and $f$.*

Recalling Corollary 57 for $d = 1$ that for every $0 < \varepsilon \leq 2L\Psi^{-1}\left(\frac{V}{4}\right)$

$$\mathscr{H}_\varepsilon\left(\mathscr{F}^{\Psi}_{[L,M,V]} \,\Big|\, \mathbf{L}^1([0,L],\mathbb{R})\right) \leq \left[3\log_2 5 + \log_2(5e)\right] \cdot \frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + \log_2\left(\frac{8LM}{\varepsilon} + 1\right), \qquad (4.2.3)$$

we prove the following.

**Theorem 62.** *Assume that $f \in \mathscr{C}^2(\mathbb{R})$ satisfies (4.2.2). Then, for any constants $L, M, T > 0$,*

$$\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]}) \Big| \mathbf{L}^1(\mathbb{R})\right) \leq \log_2\left(\frac{16M(L + T \cdot f'_M)}{\varepsilon} + 1\right) + 2\left[3\log_2 5 + \log_2(5e)\right] \cdot \frac{\gamma_{[L,M]}\left(1 + \frac{1}{T}\right)}{\Psi\left(\frac{\varepsilon}{4L + 4T \cdot f'_M}\right)}$$

*for every $\varepsilon > 0$ sufficiently small.*

*Proof.* Let us define the following set

$$\tilde{S}_T(\mathscr{U}_{[L,M]}) := \left\{ v : \left[0, 2\ell_{[L,M,T]}\right] \to [-M, M] \,\Big|\, \exists\, u_0 \in \mathscr{U}_{[L,M]} \text{ such that } v(x) = S_T u_0\left(x - \ell_{[L,M,T]}\right) \text{ for all } x \in \left[0, 2\ell_{[L,M,T]}\right] \right\}.$$

From Lemma 60 and Lemma 61, it holds that

$$\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]}) \,\Big|\, \mathbf{L}^1(\mathbb{R})\right) = \mathscr{H}_\varepsilon\left(\tilde{S}_T(\mathscr{U}_{[L,M]}) \,\Big|\, \mathbf{L}^1\left(\left[0, 2\ell_{[L,M,T]}\right], \mathbb{R}\right)\right) \qquad (4.2.4)$$

and

$$\tilde{S}_T(\mathscr{U}_{[L,M]}) \subseteq \mathscr{F}^{\Psi}_{\left[2\ell_{[L,M,T]}, M, \gamma_{[L,M,T]}\right]}$$

where

$$\mathscr{F}^{\Psi}_{\left[2\ell_{[L,M,T]}, M, \gamma_{[L,M,T]}\right]} = \left\{ g \in BV^{\Psi}\left(\left[0, 2\ell_{[L,M,T]}\right], [-M, M]\right) \,\Big|\, TV^{\Psi}(g, [0, 2\ell_{[L,M,T]}]) \leq \gamma_{[L,M,T]} \right\}$$

is defined as in Corollary 57. By (4.2.3) and (4.2.4), we obtain

$$\begin{aligned}\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]})\,\Big|\,\mathbf{L}^1(\mathbb{R})\right) &= \mathscr{H}_\varepsilon\left(\tilde{S}_T(\mathscr{U}_{[L,M]})\,\Big|\,\mathbf{L}^1\left(\left[0,2\ell_{[L,M,T]}\right],\mathbb{R}\right)\right)\\ &\le \mathscr{H}_\varepsilon\left(\mathscr{F}^{\Psi}_{\left[2\ell_{[L,M,T]},M,\gamma_{[L,M,T]}\right]}\,\Big|\,\mathbf{L}^1\left(\left[0,2\ell_{[L,M,T]}\right],\mathbb{R}\right)\right)\\ &\le \left[3\log_2 5+\log_2(5e)\right]\cdot\frac{2\gamma_{[L,M,T]}}{\Psi\left(\frac{\varepsilon}{4\ell_{[L,M,T]}}\right)}+\log_2\left(\frac{16M\ell_{[L,M,T]}}{\varepsilon}+1\right).\end{aligned}$$

This completes the proof.

□

**Remark 63.** In general, the upper estimate of $\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]})\,\Big|\,\mathbf{L}^1(\mathbb{R})\right)$ in Theorem 62 is not optimal.

To complete our work on the topic explored this section, let us consider (4.2.1) with a smooth flux $f$ having polynomial degeneracy, i.e., the set $I_f=\{u\in\mathbb{R}\mid f''(u)=0\}$ is finite and for each $w\in I_f$, there exists a natural number $p\ge 2$ such that

$$f^{(j)}(w) = 0 \qquad \text{for all } j\in\overline{2,p} \qquad \text{and} \qquad f^{(p+1)}(w)\neq 0.$$

For every $w\in I_f$, let $p_w$ be the minimal $p\ge 2$ such that $f^{(p+1)}(w)\neq 0$. The polynomial degeneracy of $f$ is defined by

$$p_f := \max_{w\in I_f} p_w\,.$$

Recalling [41, Theorem 3], we have that $S_T u_0\in BV^{\frac{1}{p_f}}(\mathbb{R},\mathbb{R})$ and

$$TV^{\frac{1}{p_f}}(S_T u_0,\mathbb{R}) \le \tilde{\gamma}_{[L,M]}\left(1+\frac{1}{T}\right) = \tilde{\gamma}_{[L,M,T]}$$

for a constant $\tilde{\gamma}_{[L,M]}$ depending only on $L, M$ and $f$. This yields

$$\tilde{S}_T(\mathscr{U}_{[L,M]}) \subseteq \mathscr{F}^{p_f}_{\left[2\ell_{[L,M,T]},M,\tilde{\gamma}_{[L,M,T]}\right]}\,,$$

where the set

$$\mathscr{F}^{p_f}_{\left[2\ell_{[L,M,T]},M,\tilde{\gamma}_{[L,M,T]}\right]} = \left\{g\in BV^{\frac{1}{p_f}}\left(\left[0,2\ell_{[L,M,T]}\right],[-M,M]\right)\,\Big|\,TV^{\frac{1}{p_f}}(g,[0,L])\le\tilde{\gamma}_{[L,M,T]}\right\}$$

is defined as in (4.1.4). Using (4.2.3) we directly obtain an extended result on the upper estimate of the $\varepsilon$-entropy of solutions in [8, Theorem 1.5] for general fluxes having polynomial degeneracy.

**Proposition 64.** *Assume that $f$ is smooth, having polynomial degeneracy $p_f$. Then, given the constants $L, M, T > 0$, for every $\varepsilon > 0$ sufficiently small, it holds that*

$$\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]}) \,\Big|\, \mathbf{L}^1(\mathbb{R})\right) \;\leq\; \frac{\Gamma_{[T,L,M,f]}}{\varepsilon^{p_f}} + \log_2\left(\frac{16(L+Tf'_M)M}{\varepsilon} + 1\right),$$

*where*

$$\Gamma_{[T,L,M,f]} \;=\; 2^{2p_f+1}\big[3\log_2 5 + \log_2(5e)\big]\tilde{\gamma}_{[L,M]}\big(L + T\cdot f'_M\big)^{p_f}\left(1+\frac{1}{T}\right).$$

**Remark 65.** The above estimate is sharp in this special case. Indeed, we may exactly follow the same argument as in the proof of [8, Theorem 1.5] to show that

$$\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]}) \,\Big|\, \mathbf{L}^1(\mathbb{R})\right) \;\geq\; \Lambda_{T,L,M,f}\cdot\frac{1}{\varepsilon^{p_f}}\,,$$

where $\Lambda_{T,L,M,f} > 0$ is a constant depending on $L, M, T$ and $f$. Hence, $\mathscr{H}_\varepsilon\left(S_T(\mathscr{U}_{[L,M]}) \,\Big|\, \mathbf{L}^1(\mathbb{R})\right)$ is of the order $\frac{1}{\varepsilon^{p_f}}$ .

# BIBLIOGRAPHY


[1] Acosta, G. & Dúran, R.C. "An optimal Poincaré inequality in $\mathbf{L}^1$ for convex domains". *Proc. Amer. Math. Soc.* **132.1** (2003), pp. 195–202.

[2] Ambrosio, L., Fusco, N. & Pallara, D. *Functions of bounded variation and free discontinuity problems.* Oxford Science Publications, Clarendon Press, Oxford, UK, 2000.

[3] Ancona, F., Cannarsa, P. & Nguyen, K.T. "Quantitative compactness estimates for Hamilton-Jacobi equations". *Arch. Rat. Mech. Anal.* **219.2** (2016), pp. 793–828.

[4] Ancona, F., Cannarsa, P. & Nguyen, K.T. "The compactness estimates for Hamilton-Jacobi Equations depending on space". *Bulletin of the Institute of Mathematics, Academia Sinica* **11.1** (2016), pp. 63–113.

[5] Ancona, F., Glass O. & Nguyen, K.T. "Lower compactness estimates for scalar balance laws". *Comm. Pure Appl. Math* **65.9** (2012), pp. 1303–1329.

[6] Ancona, F., Glass O. & Nguyen, K.T. "On quantitative compactness estimates for hyperbolic conservation laws". *Proceedings of the 14th International Conference on Hyperbolic Problems (HYP2012)*, AIMS, Springfield, MO, 2014.

[7] Ancona, F., Glass O. & Nguyen, K.T. "On lower compactness estimates for general nonlinear hyperbolic systems". *Ann. Inst. H. Poincaré Anal. Non Linéaire* **32.6** (2015), pp. 1229–1257.

[8] Ancona, F., Glass O. & Nguyen, K.T. "On Kolmogorov entropy compactness estimates for scalar conservation laws without uniform convexity". *SIAM J. Math. Anal.*, **51.4** (2019), pp. 3020–3051.

[9] Assouad, P. "Plongements lipschitziens dans $\mathbb{R}^d$". *Bull. Soc. Math. France* **111** (1983), pp. 429–448.

[10] Bartlett, P.L., Kulkarni, S.R. & Posner, S.E. "Covering numbers for real-valued function classes". *IEEE Trans. Inform. Theory* **43.5** (1997), pp. 1721–1724.

[11] Bianchini, S. & Marconi, E. "On the structure of $\mathbf{L}^\infty$- entropy solutions to scalar conservation laws in one-space dimension". *Arch. Rat. Mech. Anal.* **226** (2017), pp. 441–493.

[12] Bianchini. S., Dutta P. & Nguyen, K.T. "Metric entropy for Hamilton-Jacobi equation with uniformly directionally convex Hamiltonian". submitted to *SIAM J. Math. Anal.* (2021), https://arxiv.org/pdf/2012.10577.pdf.

[13] Birgé, L. "Approximation dans les espaces metriques et theorie de l'estimation". *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **65** (1983), pp. 181–237.

[14] Birgé, L. "Estimating a density under order restrictions: nonasymptotic minimal risk". *Ann. Stat.* **15** (1987), pp. 995–1012.

[15] Birgé, L. & Marssart, P. "Rates of convergence for minimum contrast estimators". *Probab. Theory Related Fields* **97** (1993), pp. 113–150.

[16] Bressan, A. *Hyperbolic systems of conservation laws.* Oxford Lecture Series in Mathematics and Its Applications, Oxford University Press, Oxford, 2000.

[17] Bronshtein, E.M. "$\varepsilon$-entropy of convex sets and functions". *Siberian Math J.* **17** (1976), pp. 393–398.

[18] Cannarsa, P. & Sinestrari, C. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control.* Progress in Nonlinear Differential Equations and their Applications **58**, Birkhäuser Boston, 2004.

[19] Capuani, R., Dutta, P. & Nguyen, K.T. "Metric entropy for functions of bounded total generalized variation". *SIAM J. Math. Anal.*, **53.1** (2021), pp. 1168–1190.

[20] Cheng, K.-S. "A regularity theorem for a nonconvex scalar conservation law". *J. Differential Equations* **61.1** (1986), pp. 79–127.

[21] Crandall, M.G. & Lions, P.-L. "Viscosity solutions of Hamilton-Jacobi equations". *Trans. Amer. Math. Soc.* **277.1** (1983), pp. 1–42.

[22] Dafermos, C.M. *Hyperbolic conservation laws in continuum physics.* Grundlehren Math. Wissenschaften Series **325**, Fourth ed. Berlin: Springer-Verlag, 2016.

[23] De Lellis, C. & Golse, F. "A Quantitative Compactness Estimate for Scalar Conservation Laws". *Comm. Pure Appl. Math.* **58.7** (2005), pp. 989–998.

[24] Dryanov, D. "Kolmogorov entropy for classes of convex functions". *Constructive Approx.* **30** (2009), pp. 137–153.

[25] Dudley, R.M. "Central limits theorems for empirical measure". *Ann. Probability* **6** (1978), pp. 899–929.

[26] Dutta, P. & Nguyen, K.T. "Covering numbers for bounded variation functions". *J. Math. Anal. Appl.* **468.2** (2018), pp. 1131–1143.

[27] Groeneboom, P. *Some current developments in density of estimation.* CWI Monographs, North Holland, 1986.

[28] Guntuboyina, A. & Sen, B. "Covering Numbers for Convex Functions". *IEEE Transactions On Information Theory* **59.4** (2013), pp. 1957–1965.

[29] Haussler, D. "Decision theoretic generalizations of the PAC model for neural net and other learning applications". *Information and Computation* **100.1** (1992), pp. 78–150.

[30] Haussler, D. "Sphere packing numbers for subsets of the Boolean $n$-cube with bounded Vapnik-Chervonenkis". *Journal of Combinatorial Theorem, Series A* **69.2** (1995), pp. 217–232.

[31] Helly, E. "Über lineare Funktionaloperationen". *Wien. Ber.*, **121** (1912), pp. 265–297 (German).

[32] Hoeffding, W. "Probability inequalities for sums of bounded random variables". *J. Amer. Statist. Assoc.* **58** (1963), pp. 13–30.

[33] Kolmogorov, A.N. & Tikhomirov, V.M. "$\varepsilon$-entropy and $\varepsilon$-capacity of sets in functional spaces". *Uspekhi Mat. Nauk* **14** (1959), pp. 3–86.

[34] Kruvzkov, S.N. "First order quasilinear equations with several independent variables". *Mat. Sb. (N.S.)* **81(123)** (1970), pp. 228–255 (Russian). English translation in *Math. USSR Sbornik* **10.2** (1970), pp. 217–243.

[35] Kulkarni, S.R., Mitter, S.K. & Tsitsiklis, J.N. "Active learning using arbitrary binary-valued queries". *Machine Learning* **11** (1993), pp. 23–35.

[36] Lax, P.D. "Hyperbolic systems of conservation laws II". *Comm. on Pure and Applied Math.* **10** (1957), pp. 537–566.

[37] Lax, P.D. "Accuracy and resolution in the computation of solutions of linear and nonlinear equations". *Recent advances in numerical analysis,* Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison. Publ. Math. Res. Center Univ. Wisconsin (1978), pp. 107–117. Academic Press, New York, 1978.

[38] Lax, P.D. *Course on hyperbolic systems of conservation laws*, XXVII Scuola Estiva di Fis. Mat., Ravello, 2002.

[39] Le Cam, L. (1973) Convergence of estimates under dimensionality restrictions, *Ann. Statist.*, **1**, 38–53.

[40] Lee, W.S., Bartlett, P.L. & Willamson, R.C. "On efficient learning of linear combinations of basic function". *Proceedings of the Eight Annual Conference on Computational learning theory*, ACM Press (1995), pp. 369–376.

[41] Marconi, E. "Regularity estimates for scalar conservation laws in one space dimension". *J. Hyperbolic Differential Equations* **15.4** (2018), pp. 623–691.

[42] Musielak, J. & Orlicz, W. "On generalized variations". *I. Studia Math.* **18** (1959), pp. 11–41.

[43] Oleinik, O.A. "Discontinuous solutions of non-linear differential equations". *Uspehi Mat. Nauk (N.S.)* **12.3(75)** (1957), pp. 3–73 (Russian). English translation in *Ann. Math. Soc. Trans. Ser. 2* **26** (1963), pp. 95–172.

[44] Oleinik, O.A. "Uniqueness and stability of the generalized solution of the Cauchy problem for a quasi-linear equation". *Uspehi Mat. Nauk* **14.2(86)** (1959), pp.165–170 (Russian).

[45] Pollard, D. *Convergence of Stochastic Processes.* Springer, New York, 1984.

[46] Tartar, L. "Compensated compactness and applications to partial differential equations". *Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV*, Res. Notes in Math. **39**, Pitman, Boston, Mass.-London (1979), pp. 136–212.

[47] Van de Geer, S. *Applications of empirical process theory.* Cambridge Univ. Press, Cambridge, U.K., 2000.

[48] Yang, Y. & Barron, A. "Information-theoretic determination of minimax rates of convergence". *Ann. Statist.* **27** (1999), pp. 1564–1599.