

Metric entropy for functions of bounded total generalized variation

Rossana Capuani, Prerona Dutta, and Khai T. Nguyen

Department of Mathematics, North Carolina State University

e-mails: rcapuan@ncsu.edu, pdutta2@ncsu.edu, khai@math.ncsu.edu

November 18, 2020

Abstract

We establish a sharp estimate for a minimal number of binary digits (bits) needed to represent all bounded total generalized variation functions taking values in a general totally bounded metric space (E, ρ) up to an accuracy of $\varepsilon > 0$ with respect to the \mathbf{L}^1 -distance. Such an estimate is explicitly computed in terms of doubling and packing dimensions of (E, ρ) . The obtained result is applied to 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.

Keywords: Metric entropy, doubling dimension, total generalized variation

1 Introduction

The metric entropy (or ε -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 (see e.g in [11, 24, 38]). It provides a tool for characterizing the rate of mixing of sets of small measure. The notion of metric entropy (or ε -entropy) has been introduced by Kolmogorov and Tikhomirov [27] in 1959 as follows:

Definition 1.1. *Let (E, ρ) 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 ε -covering of K , i.e., a covering of K by balls in E with radius no greater than ε . Then the ε -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 \mathcal{F} of real-valued functions defined on E under the family of $\mathbf{L}^1(dP)$ where P is a probability distribution on E . Upper and lower bounds on the ε -entropy of \mathcal{F} in terms

of Vapnik-Chervonenkis, pseudo-dimension and the scale-sensitive dimension of the function class were established in [20, 24, 25, 29, 38] and in [29, 33].

Thanks to the Helly's theorem, a set of uniformly bounded variation functions is compact in \mathbf{L}^1 -space. Consequently, attempts were made to quantify the degree of compactness of such sets by using the ε -entropy. In [29], the authors showed that the ε -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. Later on, this result was also extended to multi-dimensional cases in [21]. Some 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 [11, 22]. In the multi-dimensional cases, the covering numbers of convex and uniformly bounded functions were studied in [23]. It was shown that the ε -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 [19] and for convex functions that are uniformly bounded and uniformly Lipschitz with a known Lipschitz constant in [14]. These results have direct implications in the study of rates of convergence of empirical minimization procedures (see in [12, 40]) as well as optimal convergence rates in the numerous convexity constrained function estimation problems (see in [10, 15, 41]).

From a different aspect, the ε -entropy has been used to measure the set of solutions of nonlinear partial differential equations. In this setting, it could provide a measure of the order of "resolution" and the "complexity" of a numerical scheme, as suggested in [30]. The first results on this topic were obtained in [3, 18] for the scalar conservation law with uniformly convex flux f (i.e. $f''(u) \geq c > 0$), in one-dimensional space

$$u_t(t, x) + f(u(t, x))_x = 0. \quad (1.1)$$

It was shown that the number of functions needed to represent an entropy admissible weak solution u at any time $t > 0$ with an accuracy of ε with respect to the \mathbf{L}^1 -distance is of the order $\frac{1}{\varepsilon}$. A similar estimate was also obtained for the system of hyperbolic conservation laws in [5, 6] and for Hamilton-Jacobi equations with uniformly convex Hamiltonian in [1, 2]. All these proofs strongly relied on the BV regularity properties of solutions. Thereafter, the results in [3, 18] 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 [4] where entropy admissible weak solutions may have unbounded total variation. In this case, the sharp estimate on the ε -entropy for sets of entropy admissible weak solutions was provided by exploiting the BV bound of the characteristic speed $f'(u)$ at any positive time [16]. On the other hand, it was shown in [9, Example 7.2]) that for fluxes having one inflection point where all derivatives vanish, the composition of the derivative of the flux with the solution of (1.1) fails in general to belong to the BV space and the analysis in [4] cannot be applied here. However, for *weakly genuinely nonlinear* fluxes, that is to say for fluxes with no affine parts, equibounded sets of entropy solutions of (1.1) at positive time are still relatively compact in \mathbf{L}^1 (see [39, Theorem 26]). Therefore, for fluxes of such classes that do not fulfill the assumptions in [4], it remains an open problem to provide a sharp estimate on the ε -entropy for the solution set of (1.1). A different approach from [4] must be pursued to study the ε -entropy for (1.1) with weakly genuinely nonlinear fluxes, perhaps exploiting

the uniform bound on total generalized variation of entropy admissible weak solutions studied in [34, Theorem 1].

From the above viewpoints, the present paper aims to study the ε -entropy of classes of uniformly bounded total generalized variation functions taking values in a general totally bounded metric space (E, ρ) . More precisely, for a given convex function $\Psi : [0, +\infty) \rightarrow [0, +\infty)$ with $\Psi(0) = 0$ and $\Psi(s) > 0$ for all $s > 0$, let $\mathcal{F}_{[L, V]}^\Psi$ be a set of functions $g : [0, L] \rightarrow E$ such that the Ψ -total variation of g over the interval $[0, L]$ is bounded by V , i.e.,

$$\sup_{N \in \mathbb{N}, 0 = x_0 < x_1 < \dots < x_N = L} \sum_{i=0}^{N-1} \Psi(\rho(g(x_i), g(x_{i+1}))) \leq V.$$

We establish upper and lower bounds on $\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L, V]}^\Psi \middle| \mathbf{L}^1([0, L], E) \right)$, the ε -entropy of $\mathcal{F}_{[L, V]}^\Psi$ with respect to the \mathbf{L}^1 -distance. For deriving sharp estimates explicitly, our idea is to use the notions of doubling and packing dimensions of (E, ρ) , denoted by $\mathbf{d}(E)$ and $\mathbf{p}(E)$ respectively, which were first introduced by Assouad in [7]. In Theorem 3.1, we prove that for every $\varepsilon > 0$ sufficiently small, the sharp bounds on $\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L, V]}^\Psi \middle| \mathbf{L}^1([0, L], E) \right)$ can be approximated in terms of $\mathbf{p}(E)$, $\mathbf{d}(E)$ and Ψ . In particular, if $\Psi(s) = s^\gamma$ for some $\gamma \geq 1$ and the metric space (E, ρ) is generated by a finite dimensional normed space $(\mathbb{R}^d, \|\cdot\|)$ then the ε -entropy of $\mathcal{F}_{[L, V]}^\Psi$ in $\mathbf{L}^1([0, L], \mathbb{R}^d)$ is of the order $\frac{d}{\varepsilon^\gamma}$, i.e.,

$$\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L, V]}^\Psi \middle| \mathbf{L}^1([0, L], \mathbb{R}^d) \right) \approx \frac{d}{\varepsilon^\gamma}.$$

The result is applied to provide an upper estimate on the ε -entropy of a set of entropy admissible weak solutions to scalar conservation laws (1.1) with general weakly genuinely nonlinear fluxes in Theorem 3.7, which partially extends the recent one in [4]. The estimate is sharp in the case of fluxes having finite inflection points with a polynomial degeneracy. However, a natural question regarding sharp estimates of the ε -entropy for such solution sets to (1.1) with general weakly genuinely nonlinear fluxes is still open.

This paper is organised as follows. In Section 2, we present some preliminary results on covering and packing numbers of a totally bounded metric space and also include necessary concepts related to functions of bounded total generalized variation. In Section 3, the first subsection focuses on finding the upper and lower estimates of the ε -entropy for a set of bounded total generalized variation functions, while the second subsection is an application of these estimates to scalar conservation laws with weakly genuinely nonlinear fluxes.

2 Notations and preliminaries

Let E be a metric space with distance ρ and I be an interval in \mathbb{R} . Throughout the paper we shall denote by:

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

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

- $\text{diam}(F) = \sup_{x,y \in F} \rho(x,y)$, the diameter of the set F in (E, ρ) ;
- $\mathbf{L}^1(I, E)$, the Lebesgue metric space of all (equivalence classes of) summable functions $f : I \rightarrow E$, 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 $f, g \in \mathbf{L}^1(I, E)$;

- $\mathbf{L}^1(\mathbb{R})$, the Lebesgue space of all (equivalence classes of) summable functions on \mathbb{R} , equipped with the usual norm $\|\cdot\|_{\mathbf{L}^1}$;
- $\mathbf{L}^\infty(\mathbb{R})$, the space of all essentially bounded functions on \mathbb{R} , equipped with the usual norm $\|\cdot\|_{\mathbf{L}^\infty}$;
- $\text{Supp}(u)$, the essential support of a function $u \in \mathbf{L}^\infty(\mathbb{R})$;
- $B_{\mathbf{L}^1(I, E)}(\varphi, r)$, the open ball of radius r and center φ in $\mathbf{L}^1(I, E)$, with respect to the metric $\rho_{\mathbf{L}^1}$ on $\mathbf{L}^1(I, E)$, i.e.,

$$B_{\mathbf{L}^1(I, E)}(\varphi, r) = \{g \in \mathbf{L}^1(I, E) \mid \rho_{\mathbf{L}^1}(\varphi, g) < r\};$$

- $\mathcal{B}(I, [0, +\infty))$, a set of bounded functions from I to $[0, +\infty)$;
- $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$, space of smooth functions having derivatives of all orders;
- $TV(g, I)$, total variation of g over the interval I ;
- $TV^\Psi(g, I)$, Ψ -total variation of g over the interval I ;
- $TV^{\frac{1}{\gamma}}(g, I)$, γ -total variation of g over the interval I , i.e., Ψ -total variation of g with Ψ defined by $\Psi(s) = |s|^\gamma$;
- $\chi_I(x) = \begin{cases} 1 & \text{if } x \in I \\ 0 & \text{if } x \in \mathbb{R}^n \setminus I \end{cases}$ the characteristic function of I ;
- $\text{Card}(S)$, the number of elements in any finite set S ;
- $[x] := \max\{z \in \mathbb{Z} \mid z \leq x\}$, the integer part of x ;
- $\overline{1, N}$, the set of natural numbers from 1 to N ;
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, number of ways in which k objects can be chosen from among n objects.

2.1 Covering, packing and metric dimension

Let us first recall the concepts of covering number and packing number in a totally bounded metric space (E, ρ) . For any $K \subseteq E$ and $\alpha > 0$, we say that

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

Definition 2.1. *The α -covering and α -packing numbers of K in (E, ρ) are defined by*

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

and

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

respectively.

Since E is totally bounded, $\mathcal{N}_\alpha(K|E)$ is finite for every $\alpha > 0$. Moreover, the maps $\alpha \mapsto \mathcal{N}_\alpha(K|E)$ and $\alpha \mapsto \mathcal{M}_\alpha(K|E)$ are non-increasing. The relation between $\mathcal{N}_\alpha(K|E)$ and $\mathcal{M}_\alpha(K|E)$ is described by the following simple double inequality:

Lemma 2.2. *For any $\alpha > 0$, one has*

$$\mathcal{M}_{2\alpha}(K|E) \leq \mathcal{N}_\alpha(K|E) \leq \mathcal{M}_\alpha(K|E).$$

Proof. For the proof see e.g in [27]. □

Let us now introduce a commonly used notion of dimension for a metric space (E, ρ) , as proposed in [7, §4].

Definition 2.3. *The doubling and packing dimensions of (E, ρ) are respectively defined by*

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

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

We conclude this subsection with the following result relating α -covering and α -packing.

Lemma 2.4. *Given $R \geq 2\alpha > 0$, let k and m be natural numbers such that*

$$2 \cdot 7^k \leq \frac{R}{\alpha} \leq 2^m.$$

The following hold

$$\mathcal{N}_\alpha \left(B_\rho(z, R) \mid E \right) \leq 2^{m\mathbf{d}(E)} \quad (2.1)$$

and

$$\mathcal{M}_\alpha \left(B_\rho(z, R) \mid E \right) \geq 2^{(k+1)\mathbf{p}(E)} \quad (2.2)$$

for all $z \in E$.

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

$$\mathcal{N}_\alpha \left(B_\rho(z, 2^n \alpha) \mid E \right) \leq 2^{n\mathbf{d}(E)} \quad \text{for all } z \in E. \quad (2.3)$$

Assume that (2.3) holds for $n = i \geq 0$. For any given $z_0 \in E$, from Definition 2.3, one has

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

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

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

and

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

Thus, (2.3) holds for $n = i + 1$ and the method of induction yields (2.3) for all $n \geq 0$. In particular, the non-decreasing property of the map $r \mapsto \mathcal{N}_\alpha \left(B_\rho(z, r) \mid E \right)$ implies that

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

2. To achieve the inequality in (2.2), we prove that

$$\mathcal{M}_\alpha \left(B_\rho(z, 2 \cdot 7^n \alpha) \mid E \right) \geq 2^{(n+1)\mathbf{p}(E)} \quad \text{for all } z \in E. \quad (2.4)$$

It is clear from Definition 2.3 that (2.4) holds for $n = 0$. Assume that (2.4) holds for $n = i \geq 1$. For any given $z_0 \in E$, from Definition 2.3, one has

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

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

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

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

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

Since $B_\rho(x_j, 2 \cdot 7^i \alpha) \subseteq B_\rho(z_0, 2 \cdot 7^{i+1} \alpha)$ for all $j \in \{1, 2, \dots, 2^{\mathbf{p}(E)}\}$, one then has

$$\begin{aligned} \mathcal{M}_\alpha \left(B_\rho(z_0, 2 \cdot 7^{i+1} \alpha) \mid E \right) &\geq \sum_{j=1}^{2^{\mathbf{p}(E)}} \mathcal{M}_\alpha \left(B_\rho(x_j, 2 \cdot 7^i \alpha) \mid E \right) \geq 2^{\mathbf{p}(E)} \cdot 2^{(i+1)\mathbf{p}(E)} \\ &= 2^{(i+2)\mathbf{p}(E)}. \end{aligned}$$

Thus, by the method of induction, (2.4) holds for all $n \geq 0$. In particular, the non-decreasing property of the map $r \mapsto \mathcal{M}_\alpha \left(B_\rho(z, r) \mid E \right)$ implies that

$$\mathcal{M}_\alpha \left(B_\rho(z, R) \mid E \right) \geq \mathcal{M}_\alpha \left(B_\rho(z, 2 \cdot 7^k \alpha) \mid E \right) \geq 2^{(k+1)\mathbf{p}(E)}.$$

□

As a consequence of Lemma 2.2 and Lemma 2.4, one has that

$$\left(\frac{R}{4\alpha} \right)^{\log_7(2) \cdot \mathbf{p}(E)} \leq \mathcal{N}_\alpha \left(B_\rho(z, R) \mid E \right) \leq \left(\frac{2R}{\alpha} \right)^{\mathbf{d}(E)} \quad (2.5)$$

and

$$\left(\frac{R}{2\alpha} \right)^{\log_7(2) \cdot \mathbf{p}(E)} \leq \mathcal{M}_\alpha \left(B_\rho(z, R) \mid E \right) \leq \left(\frac{4R}{\alpha} \right)^{\mathbf{d}(E)}. \quad (2.6)$$

2.2 Functions of bounded total generalized variation

In this subsection, we now introduce the concept of total generalized variation of the function $g : [a, b] \rightarrow E$ which was well-studied in [35] for the case $E = \mathbb{R}$. Consider a convex function $\Psi : [0, +\infty) \rightarrow [0, +\infty)$ such that

$$\Psi(0) = 0 \quad \text{and} \quad \Psi(s) > 0 \quad \text{for all } s > 0. \quad (2.7)$$

Definition 2.5. *The Ψ -total variation of g over $[a, b]$ is defined as*

$$TV^\Psi(g, [a, b]) = \sup_{n \in \mathbb{N}, a=x_0 < x_1 < \dots < x_n=b} \sum_{i=0}^{n-1} \Psi(\rho(g(x_i), g(x_{i+1}))). \quad (2.8)$$

If the supremum is finite then we say that g has bounded Ψ -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 shall denote by

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

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

For any function $g \in BV^\Psi([a, b], E)$, it is easy to show by a contradiction argument 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, denoted by

$$g(x_0-) := \lim_{x \rightarrow x_0-} g(x) \quad \text{and} \quad g(x_0+) := \lim_{x \rightarrow x_0+} g(x).$$

Moreover, the set of discontinuities of g

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

is at most countable. In particular, one has the following:

Lemma 2.6. *For any function $g \in BV^\Psi([a, b], E)$, the following function*

$$\tilde{g}(b) = g(b), \quad \tilde{g}(x) := g(x+) \quad \text{for all } x \in [a, b)$$

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

$$\rho_{\mathbf{L}^1}(\tilde{g}, g) = 0 \quad \text{and} \quad TV^\Psi(\tilde{g}, [a, b]) \leq TV^\Psi(g, [a, b]). \quad (2.9)$$

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

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

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

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

and this yields the second inequality in (2.9). \square

The following remark is used in the proof of the upper estimate in Theorem 3.1.

Remark 2.7. *Under the assumption (2.7), the function Ψ is strictly increasing on $[0, +\infty)$ and*

$$\Psi(s) \leq \frac{s}{t} \cdot \Psi(t) \quad \text{for all } 0 \leq s < t. \quad (2.10)$$

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

Proof. By the convexity of Ψ and (2.7),

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

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

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

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

3 The ε -entropy for a class of BV^Ψ functions

3.1 Main results

Throughout this subsection, the metric space (E, ρ) is assumed to be totally bounded. For convenience, we use the notation

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

where $\mathbf{N}_\alpha := \mathcal{N}_\alpha(E|E)$ and $\mathbf{M}_\alpha := \mathcal{M}_\alpha(E|E)$ are the α -covering and the α -packing numbers of E in (E, ρ) 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 ε -entropy of a class of uniformly bounded Ψ -total variation functions defined on $[0, L]$ and taking values in (E, ρ) ,

$$\mathcal{F}_{[L, V]}^\Psi := \{f \in BV^\Psi([0, L], E) \mid TV^\Psi(f, [0, L]) \leq V\}, \quad (3.1)$$

in $\mathbf{L}^1([0, L], E)$.

Theorem 3.1. *Assume that the function $\Psi : [0, +\infty) \rightarrow [0, +\infty)$ is convex and satisfies the condition (2.7). Then, for every $0 < \varepsilon \leq 2L\Psi^{-1}(\frac{V}{4})$, it holds*

$$\frac{\mathbf{p}V}{2 \log_2(7) \cdot \Psi\left(\frac{256\varepsilon}{L}\right)} + \mathbf{K}_{\frac{258\varepsilon}{L}} \leq \mathcal{H}_\varepsilon\left(\mathcal{F}_{[L, V]}^\Psi \mid \mathbf{L}^1([0, L], E)\right) \leq [3\mathbf{d} + \log_2(5e)] \cdot \frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + \mathbf{H}_{\frac{\varepsilon}{4L}}. \quad (3.2)$$

As a consequence, the minimal number of functions needed to represent a function in $\mathcal{F}_{[L, V]}^\Psi$ up to an accuracy ε with respect to \mathbf{L}^1 -distance is of the order $\frac{1}{\Psi(O(\varepsilon))}$. Indeed, from (2.5) and (2.6), it holds that

$$\begin{cases} \mathbf{H}_\alpha & \leq \mathbf{d} \cdot \log_2\left(\text{diam}(E) \cdot \frac{2}{\alpha}\right) \\ \mathbf{K}_\alpha & \geq \mathbf{p} \cdot (\log_7 2) \cdot \log_2\left(\text{diam}(E) \cdot \frac{1}{2\alpha}\right) \end{cases} \quad \text{for all } \alpha > 0,$$

and (3.2) implies

$$\begin{aligned} \frac{\mathbf{p}V}{2 \log_2(7) \cdot \Psi\left(\frac{256\varepsilon}{L}\right)} + \mathbf{p} \cdot \log_7\left(\text{diam}(E) \cdot \frac{L}{516\varepsilon}\right) & \leq \mathcal{H}_\varepsilon\left(\mathcal{F}_{[L, V]}^\Psi \mid \mathbf{L}^1([0, L], E)\right) \\ & \leq [3\mathbf{d} + \log_2(5e)] \frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + \mathbf{d} \cdot \log_2\left(\text{diam}(E) \cdot \frac{8L}{\varepsilon}\right). \end{aligned} \quad (3.3)$$

On the other hand, one also obtains a sharp estimate on the ε -entropy for a class of uniformly bounded γ -total variation functions, i.e. $\Psi(x) = |x|^\gamma$, for all $\gamma \geq 1$. More precisely, let us denote by

$$\mathcal{F}_{[L, V]}^\gamma = \left\{f \in BV^{\frac{1}{\gamma}}([0, L], E) \mid TV^{\frac{1}{\gamma}}(f, [0, L]) \leq V\right\}, \quad (3.4)$$

it follows directly from Theorem 3.1 that

Corollary 3.2. *For every $0 < \varepsilon \leq 2^{\frac{\gamma-2}{\gamma}} LV^{\frac{1}{\gamma}}$,*

$$\begin{aligned} \frac{\mathbf{p}}{2^{8\gamma+1} \log_2(7)} \cdot \frac{L^\gamma V}{\varepsilon^\gamma} + \mathbf{p} \cdot \log_7 \left(\text{diam}(E) \cdot \frac{L}{516\varepsilon} \right) &\leq \mathcal{H}_\varepsilon \left(\mathcal{F}_{[L,V]}^\gamma \mid \mathbf{L}^1([0, L], E) \right) \\ &\leq 2^{\gamma+1} \cdot [3\mathbf{d} + \log_2(5e)] \frac{L^\gamma V}{\varepsilon^\gamma} + \mathbf{d} \cdot \log_2 \left(\text{diam}(E) \cdot \frac{8L}{\varepsilon} \right). \end{aligned} \quad (3.5)$$

In particular, as ε tends to $0+$, one derives that

$$\begin{aligned} \frac{\mathbf{p}}{2^{8\gamma+1} \log_2(7)} &\leq \liminf_{\varepsilon \rightarrow 0+} \left[\frac{\varepsilon^\gamma}{L^\gamma V} \cdot \mathcal{H}_\varepsilon \left(\mathcal{F}_{[L,V]}^\gamma \mid \mathbf{L}^1([0, L], E) \right) \right] \\ &\leq \limsup_{\varepsilon \rightarrow 0+} \left[\frac{\varepsilon^\gamma}{L^\gamma V} \cdot \mathcal{H}_\varepsilon \left(\mathcal{F}_{[L,V]}^\gamma \mid \mathbf{L}^1([0, L], E) \right) \right] \leq 2^{\gamma+1} [3\mathbf{d} + \log_2(5e)]. \end{aligned}$$

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

Finally, in order to apply our result to study the ε -entropy 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, ρ) is generated by a finite dimensional normed space $(\mathbb{R}^d, \|\cdot\|)$, i.e.,

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

Given an additional constant $M > 0$, the following provides upper and lower estimates for the ε -entropy of a class of uniformly bounded Ψ -total variation functions taking values in the open ball $B^d(0, M) \subset \mathbb{R}^d$,

$$\mathcal{F}_{[L,M,V]}^\Psi := \left\{ f \in BV^\Psi \left([0, L], B^d(0, M) \right) \mid TV^\Psi(f, [0, L]) \leq V \right\}, \quad (3.6)$$

in the normed space $\mathbf{L}^1(\mathbb{R}^d)$.

Corollary 3.3. *Under the same assumptions in Theorem 3.1, it holds*

$$\begin{aligned} \frac{Vd}{2 \log_2(7) \cdot \Psi \left(\frac{256\varepsilon}{L} \right)} + d \cdot \log_7 \left(\frac{LM}{258\varepsilon} \right) &\leq \mathcal{H}_\varepsilon \left(\mathcal{F}_{[L,M,V]}^\Psi \mid \mathbf{L}^1([0, L], \mathbb{R}^d) \right) \\ &\leq [3d \log_2 5 + \log_2(5e)] \cdot \frac{2V}{\Psi \left(\frac{\varepsilon}{2L} \right)} + d \cdot \log_2 \left(\frac{8LM}{\varepsilon} + 1 \right) \end{aligned} \quad (3.7)$$

for every $0 < \varepsilon \leq 2L\Psi^{-1} \left(\frac{V}{4} \right)$.

Proof. It is well-known (see e.g in [27]) that

$$d \cdot \log_2 \left(\frac{r}{\alpha} \right) \leq \mathcal{H}_\alpha \left(B^d(0, r) \mid \mathbb{R}^d \right) \leq d \cdot \log_2 \left(\frac{2r}{\alpha} + 1 \right)$$

for any $\alpha > 0$ and open ball $B^d(0, r) \subset \mathbb{R}^d$. In particular, recalling that

$$\mathbf{H}_\alpha = \log_2 \mathcal{N}_\alpha \left(B^d(0, M) \mid \mathbb{R}^d \right) \quad \text{and} \quad \mathbf{K}_\alpha = \log_2 \mathcal{M}_\alpha \left(B^d(0, M) \mid \mathbb{R}^d \right),$$

we have

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

and from Definition 2.3, it holds that

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

Using the above estimates in (3.2), one obtains (3.7). \square

In the next two subsections, we will present the proof of Theorem 3.1.

3.1.1 Upper estimate

Towards the proof of the upper bound on $\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L,V]}^\Psi \mid \mathbf{L}^1([0, L], E) \right)$ in Theorem 3.1, let us extend a result on the ε -entropy for a class of bounded total variation real-valued functions in the scalar case [8] or in [21, Lemma 2.3]. In order to obtain a sharp upper bound, one needs to utilize the doubling dimension of the metric space E and go beyond the particular cases in [8, 21] to estimate the ε -entropy for a more general case in E . More precisely, considering a set of bounded total variation functions taking values in E , which we denote by

$$\mathcal{F}_{[L,V]} = \left\{ f \in BV([0, L], E) \mid TV(f, [0, L]) \leq V \right\}, \quad (3.8)$$

the following holds.

Proposition 3.1. *For every $0 < \varepsilon \leq \frac{LV}{2}$ sufficiently small, it holds that*

$$\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L,V]} \mid \mathbf{L}^1([0, L], E) \right) \leq [3\mathbf{d} + \log_2(5e)] \cdot \frac{2LV}{\varepsilon} + \mathbf{H}_{\frac{\varepsilon}{2L}}.$$

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 $h_1 := \frac{L}{N_1}$ such that $I_{N_1-1} = [(N_1 - 1)h_1, L]$ and

$$I_i = [ih_1, (i+1)h_1) \quad \text{for all } i \in \overline{0, N_1 - 2};$$

- pick an optimal h_2 -covering $A = \{a_1, a_2, \dots, a_{\mathbf{N}_{h_2}}\}$ 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 2.1).

A function $f \in \mathcal{F}_{[L,V]}$ can be approximated by a piecewise constant function $f^\# : [0, L] \rightarrow A$ defined as follows:

$$f^\#(s) = a_{f,i} \quad \text{for all } s \in I_i, \quad i \in \overline{0, N_1 - 1}$$

for some $a_{f,i} \in A$ such that $f(t_i) \in B_\rho(a_{f,i}, h_2)$ with $t_i := \frac{2i+1}{2}h_1$. Notice that $a_{f,i}$ is not a unique choice. With this construction, the \mathbf{L}^1 -distance between f and $f^\#$ can be bounded above by

$$\begin{aligned} \rho_{\mathbf{L}^1}(f, f^\#) &\leq \sum_{i=0}^{N_1-1} \int_{I_i} \rho(f(s), f^\#(s)) ds = \sum_{i=0}^{N_1-1} \int_{I_i} \rho(f(s), a_{f,i}) ds \\ &\leq \sum_{i=0}^{N_1-1} \int_{I_i} [\rho(f(s), f(t_i)) + \rho(f(t_i), a_{f,i})] ds < \sum_{i=0}^{N_1-1} \int_{I_i} [\rho(f(s), f(t_i)) + h_2] ds \\ &\leq \left(\sum_{i=0}^{N_1-1} \frac{|I_i|}{2} \cdot [TV(f, [ih_1, t_i]) + TV(f, [t_i, (i+1)h_1])] \right) + Lh_2 \\ &= \frac{h_1}{2} \cdot TV(f, [0, L]) + Lh_2 \leq \frac{LV}{2N_1} + Lh_2 \end{aligned}$$

and the total variation of $f^\#$ over $[0, L]$ can be estimated by

$$\begin{aligned} TV(f^\#, [0, L]) &= \sum_{i=0}^{N_1-2} \rho(a_{f,i}, a_{f,i+1}) \\ &\leq \sum_{i=0}^{N_1-2} [\rho(a_{f,i+1}, f(t_{i+1})) + \rho(f(t_i), a_{f,i}) + \rho(f(t_{i+1}), f(t_i))] \\ &\leq \sum_{i=0}^{N_1-2} [2h_2 + \rho(f(t_{i+1}), f(t_i))] \leq 2(N_1 - 1) \cdot h_2 + V. \end{aligned}$$

Consider the following set of piecewise constant functions

$$\begin{aligned} \mathcal{F}_{[N_1, h_2]}^\# &= \left\{ \varphi : [0, L] \rightarrow A \mid \varphi(s) = \varphi(t_i) \text{ for all } s \in I_i, i \in \overline{0, N_1 - 1} \right. \\ &\quad \left. \text{and } TV(\varphi, [0, L]) \leq 2(N_1 - 1) \cdot h_2 + V \right\}. \end{aligned}$$

The set $\mathcal{F}_{[L,V]}$ is covered by a finite collection of closed balls centered at $\varphi \in \mathcal{F}_{[N_1, h_2]}^\#$ of radius $\frac{LV}{2N_1} + Lh_2$ in $\mathbf{L}^1([0, L], E)$, i.e.,

$$\mathcal{F}_{[L,V]} \subseteq \bigcup_{\varphi \in \mathcal{F}_{[N_1, h_2]}^\#} \overline{B}_{\mathbf{L}^1([0, L], E)} \left(\varphi, \frac{LV}{2N_1} + Lh_2 \right)$$

and the Definition 1.1 yields

$$\mathcal{H}_{\left[\frac{LV}{2N_1} + Lh_2\right]} \left(\mathcal{F}_{[L,V]} \mid \mathbf{L}^1([0, L], E) \right) \leq \log_2 \text{Card} \left(\mathcal{F}_{[N_1, h_2]}^\# \right). \quad (3.9)$$

2. In order to provide an upper bound on $\text{Card} \left(\mathcal{F}_{[N_1, h_2]}^\# \right)$, we introduce a discrete metric $\rho^\# : A \times A \rightarrow \mathbb{N}$ associated to ρ as follows:

$$\rho^\#(x, y) := \begin{cases} 0 & \text{if } x = y, \\ q + 1 & \text{if } \frac{\rho(x, y)}{h_2} \in (q, q + 1] \quad \text{for some } q \in \mathbb{N}, \end{cases} \quad (3.10)$$

for every $x, y \in A$. Since A is an optimal h_2 -covering of E , one has

$$\text{Card} \left(A \cap B_\rho(a, r) \right) \leq \mathcal{N}_{h_2} (B_\rho(a, r + h_2) | E) \quad \text{for all } a \in A, r > 0$$

and the second inequality in (2.5) yields

$$\text{Card} \left(A \cap B_\rho(a, r) \right) \leq \left(2 \cdot \left(\frac{r}{h_2} + 1 \right) \right)^{\mathbf{d}}.$$

Hence, for every $\ell \geq 1$ and $x \in A$, it holds

$$\begin{aligned} \text{Card} \left(\overline{B}_{\rho^\#}(x, \ell - 1) \right) &= \text{Card} \left(\{y \in A \mid \rho^\#(x, y) \leq \ell - 1\} \right) \\ &= \text{Card} \left(A \cap B_\rho(x, (\ell - 1)h_2) \right) \leq (2\ell)^{\mathbf{d}}. \end{aligned} \quad (3.11)$$

For any given $f^\# \in \mathcal{F}_{[N_1, h_2]}^\#$, the following increasing step function $\varphi_{f^\#} : [0, L] \rightarrow \mathbb{N}$ defined by

$$\varphi_{f^\#}(s) = \begin{cases} 0 & \text{for all } s \in I_0 \\ \sum_{\ell=0}^{i-1} \rho^\#(f^\#(t_\ell), f^\#(t_{\ell+1})) + i - 1 & \text{for all } s \in I_i, i \in \overline{1, N_1 - 1} \end{cases} \quad (3.12)$$

measures the total of jumps of $f^\#$ up to time t_i . From (3.10), one has

$$\begin{aligned} \sup_{t \in [0, L]} |\varphi_{f^\#}(t)| &\leq \sum_{\ell=0}^{N_1-2} \rho^\#(f^\#(t_\ell), f^\#(t_{\ell+1})) + N_1 - 2 \\ &\leq \sum_{\ell=0}^{N_1-2} \left(\frac{\rho(f^\#(t_\ell), f^\#(t_{\ell+1}))}{h_2} + 1 \right) + N_1 - 2 \leq \frac{TV(f^\#, [0, L])}{h_2} + 2N_1 - 3 \\ &\leq \frac{1}{h_2} \cdot (2(N_1 - 1) \cdot h_2 + V) + 2N_1 - 3 = 4N_1 - 5 + \frac{V}{h_2}. \end{aligned} \quad (3.13)$$

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 $\varphi_{f^\#}$ in (3.12) satisfies

$$\varphi_{f^\#}(s) = \varphi_{f^\#}(t_i) \in \{0, 1, 2, \dots, \Gamma_{[N_1, h_2]} - 1\} \quad \text{for all } s \in I_i, i \in \overline{0, N_1 - 1}.$$

Thus, if we consider the map $T : \mathcal{F}_{[N_1, h_2]}^\# \rightarrow \mathcal{B}([0, L], [0, +\infty))$ such that

$$T(f^\#) = \varphi_{f^\#} \quad \text{for all } f^\# \in \mathcal{F}_{[N_1, h_2]}^\#,$$

then

$$T \left(\mathcal{F}_{[N_1, h_2]}^\# \right) = \left\{ \varphi_{f^\#} \mid f^\# \in \mathcal{F}_{[N_1, h_2]}^\# \right\} \subseteq \mathcal{I}_{[N_1, h_2]}.$$

Here, $\mathcal{I}_{[N_1, h_2]}$ is the set of increasing step functions $\phi : [0, L] \rightarrow \{0, 1, 2, \dots, \Gamma_{[N_1, h_2]} - 1\}$ such that

$$\phi(0) = 0 \quad \text{and} \quad \phi(s) = \phi(t_i) \quad \text{for all } i \in \overline{0, N_1 - 1}, s \in I_i.$$

Since the cardinality of $\mathcal{I}_{[N_1, h_2]}$ is equal to $\binom{\Gamma_{[N_1, h_2]}}{N_1 - 1}$, one has

$$\text{Card} \left(T \left(\mathcal{F}_{[N_1, h_2]}^\# \right) \right) \leq \text{Card}(\mathcal{I}_{[N_1, h_2]}) = \binom{\Gamma_{[N_1, h_2]}}{N_1 - 1}. \quad (3.14)$$

3. To complete the proof, we need to establish an upper estimate on the cardinality of $T^{-1}(\varphi_{f^\#})$, the set of functions in $\mathcal{F}_{[N_1, h_2]}^\#$ that have the same total length of jumps as that of $f^\#$ at any time t_i . In order to do so, for any given $f^\# \in \mathcal{F}_{[N_1, h_2]}^\#$, we set

$$k_i^\# := \rho^\# \left(f^\#(t_i), f^\#(t_{i+1}) \right) \quad \text{for all } i \in \overline{0, N_1 - 2}.$$

As in (3.13), we have

$$\sum_{i=0}^{N_1-2} k_i^\# = \sum_{i=0}^{N_1-2} \rho^\# \left(f^\#(t_i), f^\#(t_{i+1}) \right) \leq 3(N_1 - 1) + \frac{V}{h_2}$$

and

$$\begin{aligned} T^{-1}(\varphi_{f^\#}) &= \left\{ g \in \mathcal{F}_{[N_1, h_2]}^\# \mid \rho^\#(g(t_{i+1}), g(t_i)) = k_i^\# \text{ for all } i \in \overline{0, N_1 - 2} \right\} \\ &\subseteq \left\{ g \in \mathcal{F}_{[N_1, h_2]}^\# \mid g(t_{i+1}) \in \overline{B_{\rho^\#}} \left(g(t_i), k_i^\# \right) \text{ for all } i \in \overline{0, N_1 - 2} \right\}. \end{aligned}$$

Observe from (3.11) that if $g(t_i)$ is already chosen then there are at most $(2k_i^\#)^{\mathbf{d}}$ choices for $g(t_{i+1})$. Since we have \mathbf{N}_{h_2} choices of the starting point $g(0)$, the cardinality of $T^{-1}(\varphi_{f^\#})$ can be estimated as follows

$$\begin{aligned} \text{Card} \left(T^{-1}(\varphi_{f^\#}) \right) &\leq \mathbf{N}_{h_2} \cdot \prod_{i=0}^{N_1-2} (2k_i^\#)^{\mathbf{d}} \leq \mathbf{N}_{h_2} \cdot \left(\frac{\sum_{i=0}^{N_1-2} 2k_i^\#}{N_1 - 1} \right)^{\mathbf{d}(N_1-1)} \\ &\leq \mathbf{N}_{h_2} \cdot \left(\frac{2 \left(3(N_1 - 1) + \frac{V}{h_2} \right)}{N_1 - 1} \right)^{\mathbf{d}(N_1-1)} = \mathbf{N}_{h_2} \cdot \left(6 + \frac{2}{N_1 - 1} \cdot \frac{V}{h_2} \right)^{\mathbf{d}(N_1-1)}. \quad (3.15) \end{aligned}$$

Recalling (3.14)-(3.15) 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

$$\begin{aligned}
\text{Card} \left(\mathcal{F}_{[N_1, h_2]}^\# \right) &\leq \mathbf{N}_{h_2} \cdot \left(6 + \frac{2}{N_1 - 1} \cdot \frac{V}{h_2} \right)^{\mathbf{d}(N_1 - 1)} \cdot \binom{\Gamma_{[N_1, h_2]}}{N_1 - 1} \\
&= \mathbf{N}_{h_2} \cdot \left(6 + \frac{2}{N_1 - 1} \cdot \frac{V}{h_2} \right)^{\mathbf{d}(N_1 - 1)} \cdot \frac{(\Gamma_{[N_1, h_2]} - N_1 + 2) \cdots \Gamma_{[N_1, h_2]}}{(N_1 - 1)!} \\
&\leq \frac{\mathbf{N}_{h_2}}{\sqrt{2\pi(N_1 - 1)}} \cdot \left(6 + \frac{2}{N_1 - 1} \cdot \frac{V}{h_2} \right)^{\mathbf{d}(N_1 - 1)} \cdot \left(\frac{\Gamma_{[N_1, h_2]}}{N_1 - 1} \right)^{N_1 - 1} \cdot e^{N_1 - 1} \\
&\leq \mathbf{N}_{h_2} \cdot \left(6 + \frac{2}{N_1 - 1} \cdot \frac{V}{h_2} \right)^{\mathbf{d}(N_1 - 1)} \cdot \left(4e + \frac{V}{h_2} \cdot \frac{e}{N_1 - 1} \right)^{(N_1 - 1)}.
\end{aligned}$$

Thus, (3.9) yields

$$\begin{aligned}
\mathcal{H}_{\left[\frac{LV}{2N_1} + Lh_2 \right]} \left(\mathcal{F}_{[L, V]} \mid \mathbf{L}^1([0, L], E) \right) &\leq \mathbf{d} \cdot (N_1 - 1) \cdot \log_2 \left(6 + \frac{V}{h_2} \cdot \frac{2}{N_1 - 1} \right) \\
&\quad + (N_1 - 1) \cdot \log_2 \left(4e + \frac{V}{h_2} \cdot \frac{e}{N_1 - 1} \right) + \mathbf{H}_{h_2}. \quad (3.16)
\end{aligned}$$

4. For every $0 < \varepsilon \leq \frac{LV}{2}$, by choosing $N_1 \in \mathbb{Z}^+$ and $h_2 > 0$ such that

$$\frac{3LV}{2\varepsilon} < N_1 - 1 = \left\lfloor \frac{3LV}{2\varepsilon} \right\rfloor + 1 \leq \frac{2LV}{\varepsilon}, \quad h_2 = \frac{V}{N_1 - 1},$$

we have

$$\frac{LV}{2N_1} + Lh_2 \leq \frac{LV}{2N_1} + \frac{LV}{N_1 - 1} \leq \frac{3LV}{2(N_1 - 1)} < \varepsilon \quad \text{and} \quad h_2 \geq \frac{\varepsilon}{2L}.$$

Thus, (3.16) implies that

$$\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L, V]} \mid \mathbf{L}^1([0, L], E) \right) \leq [3\mathbf{d} + \log_2(5e)] \cdot \frac{2LV}{\varepsilon} + \mathbf{H}_{\frac{\varepsilon}{2L}}$$

and this completes the proof. \square

Using Proposition 3.1, we now proceed to provide a proof for the upper estimate of the ε -entropy for the set $\mathcal{F}_{[L, V]}^\Psi$ in $\mathbf{L}^1([0, L], E)$.

Proof of the upper estimate in Theorem 3.1. From Lemma 2.6, one has

$$\mathcal{H}_\varepsilon \left(\mathcal{F}_{[L, V]}^\Psi \mid \mathbf{L}^1([0, L], E) \right) = \mathcal{H}_\varepsilon \left(\tilde{\mathcal{F}}_{[L, V]}^\Psi \mid \mathbf{L}^1([0, L], E) \right) \quad (3.17)$$

with $\tilde{\mathcal{F}}_{[L, V]}^\Psi = \left\{ f \in \mathcal{F}_{[L, V]}^\Psi \mid f \text{ is continuous from the right on the interval } [0, L] \right\}$. Thus, it is sufficient to prove the second inequality in (3.2) for $\tilde{\mathcal{F}}_{[L, V]}^\Psi$ instead of $\mathcal{F}_{[L, V]}^\Psi$.

1. For a fixed constant $h > 0$ and $f \in \tilde{\mathcal{F}}_{[L, V]}^\Psi$, let $A_{f, h} = \{x_0, x_1, x_2, \dots, x_{N_{f, h}}\}$ be a partition of $[0, L]$ which is defined by induction as follows:

$$x_0 = 0, \quad x_{i+1} = \sup \left\{ x \in (x_i, L) \mid \rho(f(y), f(x_i)) \in [0, h] \text{ for all } y \in (x_i, x] \right\} \quad (3.18)$$

for all $i \in \overline{0, N_{f,h} - 1}$. Since f is continuous from the right on $[0, L)$, it holds

$$\rho(f(x_i), f(x_{i+1})) \geq h \quad \text{for all } i \in \overline{0, N_{f,h} - 2}.$$

Thus, the increasing property of Ψ implies that

$$V \geq TV^\Psi(f, [0, L]) \geq \sum_{i=0}^{N_{f,h}-2} \Psi(\rho(f(x_i), f(x_{i+1}))) \geq (N_{f,h} - 1) \cdot \Psi(h),$$

and this yields

$$N_{f,h} - 1 \leq \frac{TV^\Psi(f, [0, L])}{\Psi(h)} \leq \frac{V}{\Psi(h)} < +\infty. \quad (3.19)$$

Introduce a piecewise constant function $f_h : [0, L] \rightarrow E$ such that

$$f_h(x) = \begin{cases} f(x_i) & \text{for all } x \in [x_i, x_{i+1}), \quad i \in \overline{0, N_{f,h} - 2} \\ f(x_{N_{f,h}-1}) & \text{for all } x \in [x_{N_{f,h}-1}, L]. \end{cases}$$

From (3.18), the \mathbf{L}^1 -distance between f_h and f is bounded by

$$\begin{aligned} \rho_{\mathbf{L}^1}(f_h, f) &= \int_{[0, L]} \rho(f_h(x), f(x)) dx = \sum_{i=0}^{N_{f,h}-1} \int_{[x_i, x_{i+1})} \rho(f(x_i), f(x)) dx \\ &\leq h \cdot \sum_{i=0}^{N_{f,h}-1} (x_{i+1} - x_i) = Lh. \end{aligned} \quad (3.20)$$

On the other hand, by the convexity of Ψ we have

$$\begin{aligned} V &\geq \sum_{i=0}^{N_{f,h}-2} \Psi(\rho(f(x_i), f(x_{i+1}))) \geq (N_{f,h} - 1) \cdot \Psi\left(\frac{1}{N_{f,h} - 1} \cdot \sum_{i=0}^{N_{f,h}-2} \rho(f(x_i), f(x_{i+1}))\right) \\ &= (N_{f,h} - 1) \cdot \Psi\left(\frac{TV(f_h, [0, L])}{N_{f,h} - 1}\right) \end{aligned}$$

and the strictly increasing property of Ψ^{-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 2.7 and (3.19), 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 (3.20) and (3.8), the set $\tilde{\mathcal{F}}_{[L,V]}^\Psi$ is covered by a collection of closed balls centered at $g \in \mathcal{F}_{[L,V_h]}$ of radius Lh in $\mathbf{L}^1([0, L], E)$, i.e.,

$$\tilde{\mathcal{F}}_{[L,V]}^\Psi \subseteq \bigcup_{g \in \mathcal{F}_{[L,V_h]}} \bar{B}_{\mathbf{L}^1([0,L],E)}(g, Lh).$$

In particular, for every $\varepsilon > 0$, choosing $h = \frac{\varepsilon}{2L}$ we have

$$V_{\frac{\varepsilon}{2L}} = \frac{\varepsilon V}{2L \cdot \Psi\left(\frac{\varepsilon}{2L}\right)} \quad \text{and} \quad \tilde{\mathcal{F}}_{[L,V]}^\Psi \subseteq \bigcup_{g \in \mathcal{F}_{[L, V_{\frac{\varepsilon}{2L}}]}} \bar{B}_{\mathbf{L}^1([0,L],E)}\left(g, \frac{\varepsilon}{2}\right)$$

and this implies

$$\mathcal{H}_\varepsilon\left(\tilde{\mathcal{F}}_{[L,V]}^\Psi \mid \mathbf{L}^1([0, L], E)\right) \leq \mathcal{H}_{\frac{\varepsilon}{2}}\left(\mathcal{F}_{[L, V_{\frac{\varepsilon}{2L}}]}\right) \mid \mathbf{L}^1([0, L], E). \quad (3.21)$$

If $0 < \varepsilon \leq 2L\Psi^{-1}\left(\frac{V}{4}\right)$ then

$$\varepsilon \leq \varepsilon \cdot \frac{V}{4 \cdot \Psi\left(\frac{\varepsilon}{2L}\right)} = \frac{L}{2} \cdot \frac{\varepsilon V}{2L \cdot \Psi\left(\frac{\varepsilon}{2L}\right)} = \frac{L}{2} \cdot V_{\frac{\varepsilon}{2L}}.$$

In this case, one can apply Proposition 3.1 to get

$$\begin{aligned} \mathcal{H}_{\frac{\varepsilon}{2}}\left(\mathcal{F}_{[L, V_{\frac{\varepsilon}{2L}}]}\right) \mid \mathbf{L}^1([0, L], E) &\leq [3\mathbf{d} + \log_2(5e)] \cdot \frac{4LV_{\frac{\varepsilon}{2L}}}{\varepsilon} + \mathbf{H}_{\frac{\varepsilon}{4L}} \\ &= [3\mathbf{d} + \log_2(5e)] \cdot \frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + \mathbf{H}_{\frac{\varepsilon}{4L}} \end{aligned}$$

and thereafter, we use (3.17), (3.21) to obtain the second inequality in (3.2). \square

3.1.2 Lower estimate

To prove the first inequality in Theorem 3.1, let us provide a lower estimate on the ε -entropy in $\mathbf{L}^1([0, L], E)$ to

$$\mathcal{G}_{[L,V,h,x]}^\Psi := \left\{ g : [0, L] \rightarrow B_\rho(x, h) \mid TV^\Psi(g, [0, L]) \leq V \right\}, \quad (3.22)$$

a class of bounded Ψ -total variation functions over $[0, L]$ taking values in the ball centered at a point $x \in E$ of radius $h > 0$.

Lemma 3.4. *Assume that $\mathbf{p} \geq 1$. For every $\varepsilon > 0$, it holds*

$$\mathcal{M}_\varepsilon\left(\mathcal{G}_{[L, V, 2^{(4+2/\tilde{\mathbf{p}})} \cdot \frac{\varepsilon}{L}, x]}^\Psi \mid \mathbf{L}^1([0, L], E)\right) \geq 2^{\frac{\tilde{\mathbf{p}}V}{2^{2\Psi\left(2^{(4+2/\tilde{\mathbf{p}})} \cdot \frac{2\varepsilon}{L}\right)}}} \quad (3.23)$$

where $\tilde{\mathbf{p}} = \log_7(2) \cdot \mathbf{p}$.

Proof. The proof is divided into two steps:

1. We first recall from (2.6) that

$$\mathcal{M}_{2^{-(2+2/\tilde{p})} \cdot h}(B_\rho(x, h)|E) \geq \left(\frac{h}{2 \cdot 2^{-(2+2/\tilde{p})} \cdot h} \right)^{\tilde{p}} = 2^{\tilde{p}+2} \quad \text{for all } h > 0 .$$

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 $h_1 = \frac{L}{N_1}$ as in Proposition 3.1;
- take a $(2^{-(2+2/\tilde{p})} \cdot h)$ -packing $A_h = \{a_1, a_2, \dots, a_{2^{\tilde{p}+2}}\}$ of $B_\rho(x, h)$, i.e.,

$$A_h \subseteq B_\rho(x, h) \quad \text{and} \quad \rho(a_i, a_j) > 2^{-(2+2/\tilde{p})} \cdot h$$

for all $a_i \neq a_j \in A_h$.

Consider the set of indices

$$\Delta_{h, N_1} = \left\{ \delta = (\delta_i)_{i \in \{0, 1, \dots, N_1-1\}} \mid \delta_i \in A_h \right\}$$

and define a class of piecewise constant functions on $[0, L]$ as follows:

$$\mathcal{G}_{h, N_1} = \left\{ g_\delta = \sum_{i=0}^{N_1-1} \delta_i \cdot \chi_{I_i} \mid \delta \in \Delta_{h, N_1} \right\}.$$

For any $\delta \in \Delta_{h, N_1}$, the Ψ -total variation of g_δ 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

$$(N_1 - 1) \cdot \Psi(2h) \leq V, \tag{3.24}$$

the definition of $\mathcal{G}_{[L, V, h, x]}^\Psi$ in (3.22) implies that $g_\delta \in \mathcal{G}_{[L, V, h, x]}^\Psi$ for every $\delta \in \Delta_{h, N_1}$ and thus

$$\mathcal{G}_{h, N_1} \subseteq \mathcal{G}_{[L, V, h, x]}^\Psi.$$

In particular, we get

$$\mathcal{M}_\varepsilon \left(\mathcal{G}_{[L, V, h, x]}^\Psi \mid \mathbf{L}^1([0, L], E) \right) \geq \mathcal{M}_\varepsilon \left(\mathcal{G}_{h, N_1} \mid \mathbf{L}^1([0, L], E) \right) \quad \text{for all } \varepsilon > 0. \tag{3.25}$$

2. Let us provide a lower bound on the ε -packing number $\mathcal{M}_\varepsilon \left(\mathcal{G}_{h, N_1} \mid \mathbf{L}^1([0, L], E) \right)$. For any given $\delta, \tilde{\delta} \in \Delta_{h, N_1}$ and $\varepsilon > 0$, we define

$$\mathcal{I}_{\tilde{\delta}}(2\varepsilon) = \left\{ \delta \in \Delta_{h, N_1} \mid \rho_{\mathbf{L}^1}(g_\delta, g_{\tilde{\delta}}) \leq 2\varepsilon \right\}, \quad \eta(\delta, \tilde{\delta}) = \text{Card} \left(\left\{ i \in \overline{0, N_1-1} \mid \delta_i \neq \tilde{\delta}_i \right\} \right).$$

The \mathbf{L}^1 -distance between g_δ and $g_{\tilde{\delta}}$ is bounded below by

$$\begin{aligned}\rho_{\mathbf{L}^1}(g_\delta, g_{\tilde{\delta}}) &= \sum_{i=0}^{N_1-1} \int_{I_i} \rho(g_\delta(t), g_{\tilde{\delta}}(t)) dt = \sum_{i=0}^{N_1-1} \rho(\delta_i, \tilde{\delta}_i) \cdot |I_i| \\ &= \frac{L}{N_1} \cdot \sum_{i=0}^{N_1-1} \rho(\delta_i, \tilde{\delta}_i) > 2^{-(2+2/\tilde{\mathfrak{P}})} \cdot \frac{Lh}{N_1} \cdot \eta(\delta, \tilde{\delta})\end{aligned}$$

and this implies the inclusion

$$\mathcal{I}_{\tilde{\delta}}(2\varepsilon) \subseteq \left\{ \delta \in \Delta_{h, N_1} \mid \eta(\delta, \tilde{\delta}) < \frac{2^{3+2/\tilde{\mathfrak{P}}} N_1 \varepsilon}{Lh} \right\}. \quad (3.26)$$

On the other hand, for every $r \in \overline{0, N_1 - 1}$, we compute

$$\text{Card} \left(\left\{ \delta \in \Delta_{h, N_1} \mid \eta(\delta, \tilde{\delta}) = r \right\} \right) = \binom{N_1}{r} \cdot (2^{\tilde{\mathfrak{P}}+2} - 1)^r.$$

Thus, (3.26) implies that

$$\text{Card}(\mathcal{I}_{\tilde{\delta}}(2\varepsilon)) \leq \text{Card} \left(\left\{ \delta \in \Delta_{h, N_1} \mid \eta(\delta, \tilde{\delta}) < \frac{2^{3+2/\tilde{\mathfrak{P}}} N_1 \varepsilon}{Lh} \right\} \right) \leq \sum_{r=0}^{\left\lfloor \frac{2^{3+2/\tilde{\mathfrak{P}}} N_1 \varepsilon}{Lh} \right\rfloor} \binom{N_1}{r} \cdot (2^{\tilde{\mathfrak{P}}+2} - 1)^r.$$

In particular, for every $0 < \varepsilon \leq 2^{-(4+2/\tilde{\mathfrak{P}})} Lh$, we have

$$\begin{aligned}\text{Card}(\mathcal{I}_{\tilde{\delta}}(2\varepsilon)) &\leq \sum_{r=0}^{\left\lfloor \frac{N_1}{2} \right\rfloor} \binom{N_1}{r} \cdot (2^{\tilde{\mathfrak{P}}+2} - 1)^r \leq (2^{\tilde{\mathfrak{P}}+2} - 1)^{\frac{N_1}{2}} \cdot \sum_{r=0}^{\left\lfloor \frac{N_1}{2} \right\rfloor} \binom{N_1}{r} \\ &\leq 2^{(\tilde{\mathfrak{P}}+2)\frac{N_1}{2}} \cdot 2^{N_1} = 2^{N_1(2+\tilde{\mathfrak{P}}/2)}. \quad (3.27)\end{aligned}$$

Recalling Definition 2.1, we then obtain that

$$\mathcal{M}_\varepsilon(\mathcal{G}_{h, N_1} \mid \mathbf{L}^1([0, L], E)) \geq \frac{\text{Card}(\mathcal{G}_{h, N_1})}{\text{Card}(\mathcal{I}_{\tilde{\delta}}(2\varepsilon))} \geq \frac{2^{N_1(\tilde{\mathfrak{P}}+2)}}{2^{N_1(2+\tilde{\mathfrak{P}}/2)}} = 2^{N_1\tilde{\mathfrak{P}}/2}.$$

Finally, by choosing $h = 2^{(4+2/\tilde{\mathfrak{P}})} \cdot \frac{\varepsilon}{L}$ and $N_1 = \left\lfloor \frac{V}{\Psi(2^{(4+2/\tilde{\mathfrak{P}})} \cdot \frac{2\varepsilon}{L})} \right\rfloor + 1$ such that (3.24) holds, we derive

$$\mathcal{M}_\varepsilon(\mathcal{G}_{2^{(4+2/\tilde{\mathfrak{P}})} \cdot \frac{\varepsilon}{L}, N_1} \mid \mathbf{L}^1([0, L], E)) \geq 2^{\frac{\tilde{\mathfrak{P}}V}{2\Psi(2^{(4+2/\tilde{\mathfrak{P}})} \cdot \frac{2\varepsilon}{L})}}$$

and thereafter, (3.25) yields (3.23). \square

To complete this section, we prove the first inequality in (3.2).

Proof of the lower bound in Theorem 3.1. 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.,

$$B_\rho\left(x_i, \frac{h_2}{2}\right) \cap B_\rho\left(x_j, \frac{h_2}{2}\right) = \emptyset \quad \text{for all } i \neq j \in \overline{1, \mathbf{M}_{h_2}}.$$

Recalling the definition of $\mathcal{G}_{[L,V,h,x]}^\Psi$ in (3.22), we have

$$\rho_{\mathbf{L}^1}(f_i, f_j) \geq \int_{[0,L]} \left[\rho(x_i, x_j) - \rho(x_i, f_i(s)) - \rho(x_j, f_j(s)) \right] ds \geq L \cdot (h_2 - 2h) =: L_{h,h_2}$$

for any $f_i \in \mathcal{G}_{[L,V,h,x_i]}^\Psi$ and $f_j \in \mathcal{G}_{[L,V,h,x_j]}^\Psi$ with $i \neq j \in \overline{1, \mathbf{M}_{h_2}}$. Thus, Lemma 2.2 implies that

$$\begin{aligned} \mathcal{N}_{\frac{L_{h,h_2}}{2}}\left(\mathcal{F}_{[L,V]}^\Psi \mid \mathbf{L}^1([0,L], E)\right) &\geq \mathcal{M}_{L_{h,h_2}}\left(\mathcal{F}_{[L,V]}^\Psi \mid \mathbf{L}^1([0,L], E)\right) \\ &\geq \mathcal{M}_{L_{h,h_2}}\left(\bigcup_{i=1}^{\mathbf{M}_{h_2}} \mathcal{G}_{[L,V,h,x_i]}^\Psi \mid \mathbf{L}^1([0,L], E)\right) \\ &= \sum_{i=1}^{\mathbf{M}_{h_2}} \mathcal{M}_{L_{h,h_2}}\left(\mathcal{G}_{[L,V,h,x_i]}^\Psi \mid \mathbf{L}^1([0,L], E)\right). \end{aligned}$$

Two cases are considered:

- If $\mathbf{p} = 0$ then by choosing $h = \frac{\varepsilon}{L}$ and $h_2 = \frac{4\varepsilon}{L}$ such that $L_{h,h_2} = 2\varepsilon$, we have

$$\mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,V]}^\Psi \mid \mathbf{L}^1([0,L], E)\right) \geq \mathbf{M}_{\frac{4\varepsilon}{L}}$$

and this particularly implies the first inequality in (3.2).

- Otherwise if $\mathbf{p} \geq 1$, then for any $\varepsilon > 0$, choosing $h = 2^{(5+2/\tilde{\mathbf{p}})} \cdot \frac{\varepsilon}{L}$ and $h_2 = \left(2 + 2^{(6+2/\tilde{\mathbf{p}})}\right) \cdot \frac{\varepsilon}{L}$ with $\tilde{\mathbf{p}} = \log_7(2) \cdot \mathbf{p}$ such that $L_{h,h_2} = 2\varepsilon$, we can apply (3.23) to $\mathcal{G}_{[L,V,h,x_i]}^\Psi$ for every $i \in \overline{1, \mathbf{M}_{h_2}}$ to obtain

$$\begin{aligned} \mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,V]}^\Psi \mid \mathbf{L}^1([0,L], E)\right) &\geq \sum_{i=1}^{\mathbf{M}_{(2+2^{(6+2/\tilde{\mathbf{p}})}) \cdot \frac{\varepsilon}{L}}} \mathcal{M}_{2\varepsilon}\left(\mathcal{G}_{[L,V,2^{(4+2/\tilde{\mathbf{p}})} \cdot \frac{2\varepsilon}{L}, x_i]}^\Psi \mid \mathbf{L}^1([0,L], E)\right) \\ &\geq \mathbf{M}_{(2+2^{(6+2/\tilde{\mathbf{p}})}) \cdot \frac{\varepsilon}{L}} \cdot 2^{\frac{\tilde{\mathbf{p}}V}{2^{(6+2/\tilde{\mathbf{p}})} \cdot \frac{\varepsilon}{L}}} \geq \mathbf{M}_{\frac{258\varepsilon}{L}} \cdot 2^{2\Psi\left(\frac{\tilde{\mathbf{p}}V}{L}\right)} \end{aligned}$$

and this yields the first inequality in (3.2). \square

3.2 An application to scalar conservation laws with weakly nonlinear fluxes

In this subsection, we use Theorem 3.1 and [34, Theorem 1] to establish an upper bound on the ε -entropy of a set of entropy admissible weak solutions for a scalar conservation law in one-dimensional space

$$u_t(t, x) + f(u(t, x))_x = 0 \quad \text{for all } (t, x) \in (0, +\infty) \times \mathbb{R} \quad (3.28)$$

with weakly genuinely nonlinear flux $f \in \mathcal{C}^2(\mathbb{R})$, i.e., which is not affine on any open interval such that the set

$$\{u \in \mathbb{R} \mid f''(u) \neq 0\} \text{ is dense in } \mathbb{R}. \quad (3.29)$$

We recall that the equation (3.28) 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 [17, 28] equivalent to the celebrated Oleinik E-condition [37] which generalizes the classical stability conditions introduced by Lax [32]:

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 there holds*

$$\frac{f(u^L) - f(u)}{u^L - u} \geq \frac{f(u^R) - f(u)}{u^R - u}$$

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 (3.28) 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. [17, 28]). For any given $T, L, M > 0$, we provide an upper bound for $\mathcal{H}_\varepsilon(S_T(\mathcal{U}_{[L,M]})|\mathbf{L}^1(\mathbb{R}))$ with

$$\mathcal{U}_{[L,M]} := \left\{ u_0 \in \mathbf{L}^\infty(\mathbb{R}) \mid \text{Supp}(u_0) \subset [-L, L], \|u_0\|_{\mathbf{L}^\infty(\mathbb{R})} \leq M \right\},$$

the set of bounded, compactly supported initial data.

By the 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 [13, Chapter 6], one can show that

Lemma 3.5. *For every $L, M, T > 0$ and $u_0 \in \mathcal{U}_{[L,M]}$, it holds*

$$\|S_T u_0\|_{\mathbf{L}^\infty(\mathbb{R})} \leq M \quad \text{and} \quad \text{Supp}(S_T u_0) \subseteq [-\ell_{[L,M,T]}, \ell_{[L,M,T]}]$$

where

$$\ell_{[L,M,T]} := L + T \cdot f'_M \quad \text{and} \quad f'_M := \sup_{|v| \leq M} |f'(v)|.$$

Proof. For the proof see [4, Lemma 2.2]. □

Let us introduce the function $\mathfrak{d} : [0, +\infty) \rightarrow [0, +\infty)$ such that

$$\mathfrak{d}(h) = \min_{a \in [-M, M-h]} \left(\inf_{g \in \mathcal{A}_{[a, a+h]}} \|f - g\|_{\mathbf{L}^\infty([a, a+h])} \right)$$

with $\mathcal{A}_{[a, a+h]}$ being the set of affine functions defined on $[a, a+h]$. The convex envelop Φ of \mathfrak{d} is defined by

$$\Phi = \sup_{\varphi \in \mathcal{G}} \varphi \quad \text{with} \quad \mathcal{G} := \{\varphi : [0, +\infty) \rightarrow [0, +\infty) \mid \varphi \text{ is convex, } \varphi(0) = 0, \varphi \leq \mathfrak{d}\}.$$

The following function

$$\Psi(x) := \Phi(x/2) \cdot x \quad \text{for all } x \in [0, +\infty)$$

is convex and satisfies the condition (2.7). As a consequence of [34, Theorem 1], the following holds:

Lemma 3.6. For any $u_0 \in \mathcal{U}_{[L,M]}$, the function $S_T u_0$ has bounded Ψ -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 3.3 for $d = 1$ that for every $0 < \varepsilon \leq 2L\Psi^{-1}\left(\frac{V}{4}\right)$

$$\mathcal{H}_\varepsilon\left(\mathcal{F}_{[L,M,V]}^\Psi \mid \mathbf{L}^1([0, L], \mathbb{R})\right) \leq [3 \log_2 5 + \log_2(5e)] \cdot \frac{2V}{\Psi\left(\frac{\varepsilon}{2L}\right)} + \log_2\left(\frac{8LM}{\varepsilon} + 1\right), \quad (3.30)$$

we prove the following:

Theorem 3.7. Assume that $f \in \mathcal{C}^2(\mathbb{R})$ satisfies (3.29). Then, for any constants $L, M, T > 0$, the following holds

$$\begin{aligned} \mathcal{H}_\varepsilon\left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R})\right) &\leq \log_2\left(\frac{16M(L + T \cdot f'_M)}{\varepsilon} + 1\right) \\ &\quad + 2[3 \log_2 5 + \log_2(5e)] \cdot \frac{\gamma_{[L,M]} \left(1 + \frac{1}{T}\right)}{\Psi\left(\frac{\varepsilon}{4L+4T \cdot f'_M}\right)} \end{aligned}$$

for every $\varepsilon > 0$ sufficiently small.

Proof. Let us define the following set

$$\begin{aligned} \tilde{S}_T(\mathcal{U}_{[L,M]}) &:= \left\{ v : [0, 2\ell_{[L,M,T]}] \rightarrow [-M, M] \mid \exists u_0 \in \mathcal{U}_{[L,M]} \text{ such that} \right. \\ &\quad \left. v(x) = S_T u_0(x - \ell_{[L,M,T]}) \text{ for all } x \in [0, 2\ell_{[L,M,T]}] \right\}. \end{aligned}$$

From Lemma 3.5 and Lemma 3.6, it holds that

$$\mathcal{H}_\varepsilon\left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R})\right) = \mathcal{H}_\varepsilon\left(\tilde{S}_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1([0, 2\ell_{[L,M,T]}], \mathbb{R})\right) \quad (3.31)$$

and

$$\tilde{S}_T(\mathcal{U}_{[L,M]}) \subseteq \mathcal{F}_{[2\ell_{[L,M,T]}, M, \gamma_{[L,M,T]}]}^\Psi,$$

where

$$\mathcal{F}_{[2\ell_{[L,M,T]}, M, \gamma_{[L,M,T]}]}^\Psi = \left\{ g \in BV^\Psi\left([0, 2\ell_{[L,M,T]}], [-M, M]\right) \mid TV^\Psi(g, [0, 2\ell_{[L,M,T]}]) \leq \gamma_{[L,M,T]} \right\}$$

is defined as in Corollary 3.3. By (3.30) and (3.31), we obtain

$$\begin{aligned} \mathcal{H}_\varepsilon\left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R})\right) &= \mathcal{H}_\varepsilon\left(\tilde{S}_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1([0, 2\ell_{[L,M,T]}], \mathbb{R})\right) \\ &\leq \mathcal{H}_\varepsilon\left(\mathcal{F}_{[2\ell_{[L,M,T]}, M, \gamma_{[L,M,T]}]}^\Psi \mid \mathbf{L}^1([0, 2\ell_{[L,M,T]}], \mathbb{R})\right) \\ &\leq [3 \log_2 5 + \log_2(5e)] \cdot \frac{2\gamma_{[L,M,T]}}{\Psi\left(\frac{\varepsilon}{4\ell_{[L,M,T]}}\right)} + \log_2\left(\frac{16M\ell_{[L,M,T]}}{\varepsilon} + 1\right). \end{aligned}$$

This completes the proof. \square

Remark 3.8. In general, the upper estimate of $\mathcal{H}_\varepsilon \left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R}) \right)$ in Theorem 3.7 is not optimal.

We complete this subsection by considering (3.28) 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 $p \geq 2$ such that

$$f^{(j)}(w) = 0 \quad \text{for all } j \in \overline{2, p} \quad \text{and} \quad f^{(p+1)}(w) \neq 0.$$

For every $w \in I_f$, let p_w be the minimal $p \geq 2$ 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 [34, Theorem 3], we have that $S_T u_0 \in BV^{\frac{1}{p_f}}(\mathbb{R}, \mathbb{R})$ and

$$TV^{\frac{1}{p_f}}(S_T u_0, \mathbb{R}) \leq \tilde{\gamma}_{[L,M]} \left(1 + \frac{1}{T} \right) = \tilde{\gamma}_{[L,M,T]}$$

for a constant $\tilde{\gamma}_{[L,M]}$ depending only on L, M and f . This yields

$$\tilde{S}_T(\mathcal{U}_{[L,M]}) \subseteq \mathcal{F}_{[2\ell_{[L,M,T]}, M, \tilde{\gamma}_{[L,M,T]}]}^{p_f},$$

where the set

$$\mathcal{F}_{[2\ell_{[L,M,T]}, M, \tilde{\gamma}_{[L,M,T]}]}^{p_f} = \left\{ g \in BV^{\frac{1}{p_f}} \left([0, 2\ell_{[L,M,T]}], [-M, M] \right) \mid TV^{\frac{1}{p_f}}(g, [0, L]) \leq \tilde{\gamma}_{[L,M,T]} \right\}$$

is defined as in (3.4). Using (3.30) one directly obtains an extended result on the upper estimate of the ε -entropy of solutions in [4, Theorem 1.5] for general fluxes having polynomial degeneracy.

Proposition 3.2. 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

$$\mathcal{H}_\varepsilon \left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R}) \right) \leq \frac{\Gamma_{[T,L,M,f]}}{\varepsilon^{p_f}} + \log_2 \left(\frac{16(L + T f'_M)M}{\varepsilon} + 1 \right),$$

where

$$\Gamma_{[T,L,M,f]} = 2^{2p_f+1} [3 \log_2 5 + \log_2(5e)] \tilde{\gamma}_{[L,M]} (L + T \cdot f'_M)^{p_f} \left(1 + \frac{1}{T} \right).$$

Remark 3.9. The above estimate is sharp in this special case. Indeed, we may exactly follow the same argument as in the proof of [4, Theorem 1.5] to show that

$$\mathcal{H}_\varepsilon \left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R}) \right) \geq \Lambda_{T,L,M,f} \cdot \frac{1}{\varepsilon^{p_f}},$$

where $\Lambda_{T,L,M,f} > 0$ is a constant depending on L, M, T and f . Hence, $\mathcal{H}_\varepsilon \left(S_T(\mathcal{U}_{[L,M]}) \mid \mathbf{L}^1(\mathbb{R}) \right)$ is of the order $\frac{1}{\varepsilon^{p_f}}$.

Acknowledgments. This research by K. T. Nguyen was partially supported by a grant from the Simons Foundation/SFARI (521811, NTK). The authors would like to warmly thank the anonymous referees for carefully reading the manuscript and for their suggestions, which greatly helped in improving the paper overall.

References

- [1] F. Ancona, P. Cannarsa and Khai T. Nguyen, Quantitative compactness estimates for Hamilton-Jacobi equations, *Arch. Rat. Mech. Anal.*, **219**, no. 2, 793–828, 2016.
- [2] F. Ancona, P. Cannarsa and Khai T. Nguyen, The compactness estimates for Hamilton Jacobi Equations depending on space, *Bulletin of the Institute of Mathematics, Academia Sinica* **11**, no. 1, 63–113, 2016.
- [3] F. Ancona, O. Glass and K. T. Nguyen, Lower compactness estimates for scalar balance laws, *Comm. Pure Appl. Math* **65**, no. 9, 1303–1329, 2012.
- [4] F. Ancona, O. Glass and K. T. Nguyen, On Kolmogorov entropy compactness estimates for scalar conservation laws without uniform convexity, *SIAM J. Math. Anal.* **51**, no. 4, 3020–3051, 2019.
- [5] F. Ancona, O. Glass and Khai T. Nguyen, On lower compactness estimates for general nonlinear hyperbolic systems, *Ann. Inst. H. Poincaré Anal. Non Linéaire*, **32**, no. 6, 1229–1257, 2015.
- [6] F. Ancona, O. Glass and K. T. Nguyen, On quantitative compactness estimates for hyperbolic conservation laws, to appear on Hyperbolic problems: theory, numerics and applications; proceedings of the 14th International Conference on Hyperbolic Problems (HYP2012), AIMS, Springfield, MO, 2014.
- [7] P. Assouad, Plongements lipschitziens dans \mathbb{R}^d , *Bull. Soc. Math. France*, **111**, 429–448, 1983.
- [8] P. L. Bartlett, S. R. Kulkarni and S.E. Posner, Covering numbers for real-valued function classes, *IEEE Trans. Inform. Theory* **43**, no. 5, 1721–1724, 1997.
- [9] S. Bianchini and E. Marconi, On the structure of \mathbf{L}^∞ - entropy solutions to scalar conservation laws in one space dimension, *Arch. Rat. Mech. Anal.*, **226**, 441–493, 2017.
- [10] L. Birgé, Approximation dans les espaces metriques et theorie de l’estimation, *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **65**, 181–237, 1983.
- [11] L. Birgé, Estimating a density under order restrictions: nonasymptotic minimal risk, *Ann. Stat.* **15**, 995–1012, 1987.
- [12] L. Birgé and P. Marssart, Rates of convergence for minimum contrast estimators, *Probab. Theory Related Fields* **97**, 113–150, 1993.
- [13] A. Bressan, Hyperbolic systems of conservation laws, Oxford Lecture Series in Mathematics and its applications 20, Oxford University Press, Oxford, 2000.
- [14] E. M. Bronshtein, ε -entropy of convex sets and functions, *Siberian Math J.* **17**, 393–398, 1976.
- [15] L. Le. Cam, Convergence of estimates under dimensionality restrictions, *Ann. Statist.* **1**, 38–53, 1973.
- [16] K.-S. Cheng, A regularity theorem for a nonconvex scalar conservation law, *J. Differential Equations* **61**, no. 1, 79–127, 1986.

- [17] C. M. Dafermos, Hyperbolic conservation laws in continuum physics, Grundlehren Math. Wissenschaften Series, Vol. 325. Fourth ed. Berlin: Springer-Verlag, 2016.
- [18] C. De Lellis and F. Golse, A quantitative compactness estimate for scalar conservation laws, *Comm. Pure Appl. Math.* **58**, no. 7, 989–998, 2005.
- [19] D. Dryanov, A. N. Kolmogorov entropy for classes of convex functions, *Constructive Approx* **30**, 137–153, 2009.
- [20] R. M. Duley, Central limits theorems for empirical measure, *Ann. Probability* **6**, 899–929, 1978.
- [21] P. Dutta and K. T. Nguyen, Covering numbers for bounded variation functions, *J. Math. Anal. Appl.* **468**, no. 2, 1131–1143, 2018.
- [22] P. Groeneboom, Some current developments in density of estimation, *CWI Monographs*, North Holland, 1986.
- [23] A. Guntuboyina and B. Sen, Covering Numbers for Convex Functions, *IEEE Transactions On Information Theory* **59**, no. 4, 1957–1965, 2013.
- [24] D. Haussler, Decision theoretic generalizations of the PAC model for neural net and other learning applications, *Information and Computation* **100**, 78–150, 1992.
- [25] D. Haussler, Sphere packing numbers for subsets of the Boolean n -cube with bounded Vapnik-Chervonenkis, *Journal of Combinatorial Theorem, Series A* **69**, 1995.
- [26] W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30, 1963.
- [27] A.N. Kolmogorov and V.M Tikhomirov, ε -entropy and ε -capacity of sets in functional spaces, *Uspekhi Mat. Nauk* **14**, 3-86, 1959.
- [28] S. N. Kravtsov, First order quasilinear equations with several independent variables, *Mat. Sb. (N.S.)* **81** (123) 1970, 228–255. (Russian) English translation in *Math. USSR Sbornik* Vol. 10, No. 2, 217–243, 1970.
- [29] S.R. Kulkarni, S.K. Mitter, and J.N. Tsitsiklis, Active learning using arbitrary binary-valued queries, *Machine Learning* **11**, 23–35 1993.
- [30] P. D. Lax, Accuracy and resolution in the computation of solutions of linear and nonlinear equations, in : *Recent Advances in Numerical Analysis, Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1978*, Publ. Math. Res. Center, Univ. Wisconsin, Academic Press, New York, 107–117, 1978.
- [31] P.D. Lax, *Course on Hyperbolic Systems of Conservation Laws*, XXVII Scuola Estiva di Fis. Mat., Ravello, 2002.
- [32] P. D. Lax, Hyperbolic systems of conservation laws II, *Comm. on Pure and Applied Math.*, **10**, 537-566, 1957.
- [33] W. S. Lee, P. L. Bartlett, and R. C. Williamson, On efficient learning of linear combinations of basic function, *Proceedings of the Eight Annual Conference on Computational learning theory*, ACM Press, 369–376, 1995.

- [34] E. Marconi, Regularity estimates for scalar conservation laws in one space dimension, *J. Hyperbolic Differential Equations*, **15**, no. 4, 623–691, 2018.
- [35] J. Musielak and W. Orlicz, On generalized variations, *I. Studia Math.*, 18,11–41, 1959.
- [36] O. A. Oleinik, Discontinuous solutions of non-linear differential equations, *Uspehi Mat. Nauk (N.S.)* **12** (1957) no.3 (75), 3–73. (Russian) English translation in *Ann. Math. Soc. Trans. Ser. 2* **26**, 95–172.
- [37] O. A. Oleinik, Uniqueness and stability of the generalized solution of the Cauchy problem for a quasi-linear equation, *Uspehi Mat. Nauk* **14**, no. 2(86), 165–170 (Russian), 1959.
- [38] D. Pollard, *Convergence of Stochastic Processes*, Springer, New York, 1984.
- [39] L. Tartar, Compensated compactness and applications to partial differential equations, In: *Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV*, volume 39 of *Res. Notes in Math.*, Pitman, Boston, Mass.-London, 136–212, 1979.
- [40] S. Van de Geer, *Applications of empirical process theory*, Cambridge Univ. Press, Cambridge, U.K., 2000.
- [41] Y. Yang and A. Barron, Information-theoretic determination of minimax rates of convergence, *Ann. Statist.* **27**, 1564–1599, 1999.