Generating Rectifiable Measures through Neural Networks

Erwin Riegler, Alex Bühler, Yang Pan, and Helmut Bölcskei

December 5, 2024

Abstract

We derive universal approximation results for the class of (countably) m-rectifiable measures. Specifically, we prove that m-rectifiable measures can be approximated as push-forwards of the one-dimensional Lebesgue measure on $[0,1]$ using ReLU neural networks with arbitrarily small approximation error in terms of Wasserstein distance. What is more, the weights in the networks under consideration are quantized and bounded and the number of ReLU neural networks required to achieve an approximation error of ε is no larger than $2^{b(\varepsilon)}$ with $b(\varepsilon) = \mathcal{O}(\varepsilon^{-m} \log^2(\varepsilon))$. This result improves Lemma IX.4 in Perekrestenko et al. as it shows that the rate at which $b(\varepsilon)$ tends to infinity as ε tends to zero equals the rectifiability parameter m, which can be much smaller than the ambient dimension. We extend this result to countably m-rectifiable measures and show that this rate still equals the rectifiability parameter m provided that, among other technical assumptions, the measure decays exponentially on the individual components of the countably m-rectifiable support set.

1 Introduction

Generative models [25] are used to approximate high-dimensional and very complex data measures by simplified model measures, which allow to generate new samples in a simpler fashion. These models can be applied to situations where it is difficult or expensive to gather a large amount of training data and synthetic data is generated to expand the dataset. In the last decade, the focus was on deep generative models, where deep neural networks are trained to generate the model measures [4, 10, 16]. Approximation results for deep generative models were obtained in [19, 27, 32, 33], which are briefly described next. First results along these lines appear in [19], where it was shown that deep nerural networks can be used to approximate measures arising from the composition of Barron functions [5]. In [32], it was shown that the uniform measure on $[0,1]^n$ can be approximated arbitrarily well as the push-forward of Lebesgue measure on $[0,1]$ under a ReLU neural network, where the error is measured in terms of Wasserstein distance. This result was generalized in [27] to arbitrary measures on $[0,1]^n$. Concretely, the measure under consideration is approximated by a uniform mixture, which in turn can then be approximated by the push-forward of Lebesgue measure on $[0,1]$ under a ReLU neural network. The number of parameters in the ReLU neural network that suffice to achieve an error no more than ε is of order $\mathcal{O}(\varepsilon^{-n})$ [27, Theorem VIII.1]. This result was improved [33], where it was shown that for arbitrary measures on $\mathbb{R}^n$, the number of parameters in the ReLU neural network that suffice to achieve an error no more than ε is of order $\mathcal{O}(\varepsilon^{-s})$, where s can be arbitrarily close to the upper dimension of the measure [33, Theorem 2.4], provided that the measure satisfies a certain moment condition. The proof idea is to construct a covering of an approximate support set of the measure, put mass points in the centers of the covering balls, and then connect these mass points by a piecewise linear curve, which can then be realized in terms of a ReLU neural network.

In the above summary, the results in [27] are of particular interest for two reasons. First, given a measure on $[0,1]^n$, the ReLU neural network that achieves the desired approximation error is explicit. And second, the ReLU neural networks have quanitzed and bounded weights, which in turn leads to a covering result for arbitrary measures on $[0,1]^n$. Specifically, it is shown in [27, Lemma IX.4] that, for general measures on $[1,0]^n$, the number of ReLU neural networks that suffice to achieve an approximation error of ε is no larger than $2^{b(\varepsilon)}$ with

$$\inf\{s \in (0,\infty) : \lim_{\varepsilon\to 0} \varepsilon^s b(\varepsilon) < \infty\} = n. \tag{1}$$

In many applications, the data measures are supported on low-dimensional objects in Euclidean space [13, 20–22, 30]. Although [27, Lemma IX.4] applies to such situations, the number of bits in (1) suffers from the curse of dimensionality as it depends on the ambient dimension n rather than the intrinsic dimension of the low-dimensional objects. In this paper, we consider the broad class of (countably) m-rectifiable measures that are supported on (countable) unions of Lipschitz images of compact sets in $\mathbb{R}^m$. We show that m-rectifiable measures can be approximated as push-forwards of the one-dimensional Lebesgue measure on $[0,1]$ using ReLU neural networks with arbitrarily small approximation error in terms of Wasserstein distance. What is more, the weights in the networks under consideration are quantized and bounded and the number of ReLU neural networks that suffice to achieve an approximation of ε is no larger than $2^{b(\varepsilon)}$ with

$$\inf\{s \in (0,\infty) : \lim_{\varepsilon\to 0} \varepsilon^s b(\varepsilon) < \infty\} = m. \tag{2}$$

Comparing (1) and (2), we see that the ambient dimension n is replaced by the intrinsic dimension of the objects, i.e., the rectifiability parameter m, which can be much smaller than n. We then extend this result to countably m-rectifiable measures and show that $b(\varepsilon)$ still satisfies (2) provided that, among other technical assumptions, the measure decays exponentially on the individual m-rectifiable components of the countably m-rectifiable set.

The main idea of our proof technique for m-rectifiable measures is depicted in Figure 1. The m-rectifiable measure ν is supported on the m-rectifiable set $f(\mathcal{A})$ and can be realized as push-forward of a Radom measure μ that is supported on $\mathcal{A}$ and the Lipschitz mapping f can be approximated by a ReLU neural network Φ. We then approximate μ by $\Sigma\#\lambda$, where λ denotes Lebesgue measure on $[0,1]$ and Σ is a ReLU neural network, in a space-filling fashion. The triangle inequality for Waserstein distance then yields

$$W_1(\nu, (\Phi\circ\Sigma)\#\lambda) \leq W_1(\nu, \Phi\#\mu) + W_1(\Phi\#\mu, (\Phi\circ\Sigma)\#\lambda)$$

$$\leq W_1(\nu, \Phi\#\mu) + \mathrm{Lip}(\Phi) W_1(\mu, \Sigma\#\lambda). \tag{3}$$

To upper-bound the individual terms in (3), it is crucial to control the Lipschitz constant $\mathrm{Lip}(\Phi)$ of the ReLU neural network approximating the Lipschitz function f. We emphasize that we can explicitly construct the ReLU neural network Φ approximating f and, if μ is known explicitly, which is, e.g., the case when f is injective (see Lemma 7), we can also explicitly construct the space-filling ReLU neural network Σ.

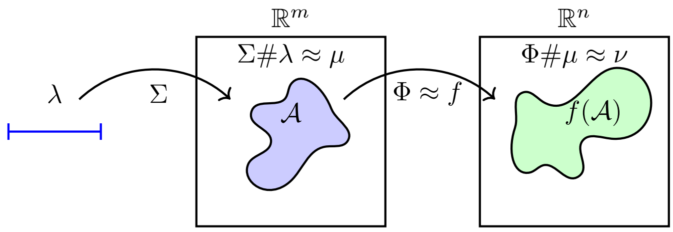

Figure 1: Illustration of proof technique for the m-rectifiable measure ν supported on the m-rectifiable set $f(\mathcal{A})$ on $\mathbb{R}^n$.

The paper is organized as follows. In Section 2, we introduce ReLU neural networks and state their main properties. Section 3 is devoted to the approximation of Lipschitz functions by ReLU neural networks. In Section 4, we introduce (countably) rectifiable measures, state their main properties, and derive our main approximation results. Finally, Appendix A contains approximation properties by uniform mixtures and quantitative statements on the space-filling approximation of uniform mixtures, which improve and simplify the results in [27]. In the appendices, we present the following material required in the current paper: Properties of Lipschitz functions in Appendix A, tools from general measure theory in Appendix B, properties of Wasserstein distance in Appendix C, auxiliary results in Appendix D, and properties of the sawtooth function in Appendix E.

1.1 Notation

We denote by e_i the i-th unit vector in $\mathbb{R}^m$ and write I_m for the $m \times m$ identity matrix. log refers to the logarithm of base 2 and $[\,\cdot\,]\colon \mathbb{R} \to \mathbb{Z}$ denotes rounding to the nearest integer. Sets are designated by calligraphic letters, e.g., $\mathcal{A}$, with $|\mathcal{A}|$ denoting cardinality and $\overline{\mathcal{A}}$ closure. The diameter of the set $\mathcal{A} \subseteq \mathbb{R}^d$ is $\mathrm{diam}(\mathcal{A}) = \sup_{x,y\in\mathcal{A}} \|x - y\|_\infty$. For every $p \in [1, \infty]$, we write $\mathrm{Lip}^{(p)}(f)$ for the Lipschitz constant with respect to the p-norm of the Lipschitz mapping f (see Definition 10). We write $\mathscr{L}^m$ and $\mathscr{H}^m$ for the m-dimensional Lebesgue and Hausdorff measure, respectively. $W^{1,\infty}(\mathcal{A})$ is the Sobolev space of order one equipped with the infinity norm on the open set $\mathcal{A} \subseteq \mathbb{R}^d$ [1, Definitions 3.1 and 3.2]. $W_1(\,\cdot\,,\,\cdot\,)$ denotes the Wasserstein distance of order one (see Definition 21). We use standard notation of general measure theory (see Appendix B). In particular, a measure μ will always refer to an outer measure, i.e., defined on all subsets, and we write $\mathscr{M}(\mu)$ for the sigma-algebra of μ-measurable sets.

2 Neural Networks

In this section, we give a brief overview of neural networks with ReLU activation function. We follow the expositions in [7, 12] and start with the definition of a ReLU activation function and a ReLU neural network.

Definition 1. For every $m \in \mathbb{N}$ and $x \in \mathbb{R}^m$, the ReLU activation function $\rho\colon \mathbb{R}^m \to \mathbb{R}^m$ is defined according to

$$\rho(x) = (\max\{0, x_1\}, \max\{0, x_2\}, \dots, \max\{0, x_m\})^\mathsf{T}. \tag{4}$$

Definition 2. Fix $L \in \mathbb{N}$. A ReLU neural network of depth L is a mapping $\Phi\colon \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ given by

$$\Phi = \begin{cases} W_1 & \text{if } L = 1 \\ W_L \circ \rho \circ W_{L-1} \circ \rho \dots \rho \circ W_1 & \text{if } L \geq 2, \end{cases} \tag{5}$$

where, for $i = 1, \dots, L$, $W_i\colon \mathbb{R}^{N_{i-1}} \to \mathbb{R}^{N_i}$ is an affine transformation defined according to $W_i(x) = A_i x + b_i$ with $A_i \in \mathbb{R}^{N_i \times N_{i-1}}$ and $b_i \in \mathbb{R}^{N_i}$. The numbers $N_1, \dots, N_L$ are called architecture of the network.

We denote by $\mathcal{N}_{m,n}$ the set of all ReLU neural networks of input dimension m and output dimension n of arbitrary depth and arbitrary architecture and introduce the following quantities.

Definition 3. For every $\Phi \in \mathcal{N}_{m,n}$, we designate the following quantities:

(i) Depth: $\mathcal{L}(\Phi) = L$;

(ii) Connectivity:

$$\mathcal{M}(\Phi) = \sum_{i=1}^{L} (\|A_i\|_0 + \|b_i\|_0); \tag{6}$$

(iii) Width: $\mathcal{W}(\Phi) = \max\{N_0, \dots, N_L\}$;

(iv) Set of weights: $\mathcal{K}(\Phi) = \mathcal{K}_1(\Phi) \cup \mathcal{K}_2(\Phi) \cup \dots \cup \mathcal{K}_L(\Phi)$ with

$$\mathcal{K}_i(\Phi) = \bigcup_{k_i=1}^{N_i} \Big\{ (b_i)_{k_i} \cup (A_i)_{k_i,1} \cup (A_i)_{k_i,2} \cup \dots \cup (A_i)_{k_i,N_{i-1}} \Big\} \tag{7}$$

for $i = 1, \dots, L$;

(v) Weight magnitude: $\mathcal{B}(\Phi) = \max\{\big|w\big| : w \in \mathcal{K}\}$.

We need the following two properties of ReLU neural networks.

Lemma 1. Set $W_1(x) = (I_m, -I_m)^\mathsf{T} x$ for all $x \in \mathbb{R}^m$ and $W_2(x) = (I_m, -I_m)x$ for all $x \in \mathbb{R}^{2m}$ and consider the ReLU neural network $i_m \in \mathcal{N}_{m,m}$ defined according to $i_m = W_2 \circ \rho \circ W_1$. Then, we have $i_m(x) = x$ for all $x \in \mathbb{R}^m$.

Proof. A direct calculation yields the result. □

Lemma 2. For $i = 1, \dots, \ell$, let $\Phi_i \in \mathcal{N}_{m_i,n_i}$ with $\mathcal{L}(\Phi_i) = L$. Further, set $m = m_1 + m_2 + \cdots + m_\ell$ and $n = n_1 + n_2 + \cdots + n_\ell$. Then, we can construct a $P(\Phi_1, \Phi_2, \dots, \Phi_\ell) \in \mathcal{N}_{m,n}$ with

$$P(\Phi_1, \Phi_2, \dots, \Phi_\ell)(x) = (\Phi_1, \Phi_2, \dots, \Phi_\ell)^\mathsf{T}(x) \quad \text{for all } x \in \mathbb{R}^m \tag{8}$$

satisfying the following properties:

$$\mathcal{L}(P(\Phi_1, \Phi_2, \dots, \Phi_\ell)) = L \tag{9}$$
$$\mathcal{M}(P(\Phi_1, \Phi_2, \dots, \Phi_\ell)) = \mathcal{M}(\Phi_1) + \mathcal{M}(\Phi_2) + \cdots + \mathcal{M}(\Phi_\ell) \tag{10}$$
$$\mathcal{W}(P(\Phi_1, \Phi_2, \dots, \Phi_\ell)) = \mathcal{W}(\Phi_1) + \mathcal{W}(\Phi_2) + \cdots + \mathcal{W}(\Phi_\ell) \tag{11}$$
$$\mathcal{K}(P(\Phi_1, \Phi_2, \dots, \Phi_\ell)) = \mathcal{K}(\Phi_1) \cup \mathcal{K}(\Phi_2) \cup \cdots \cup \mathcal{K}(\Phi_\ell) \cup \{0\} \tag{12}$$
$$\mathcal{B}(P(\Phi_1, \Phi_2, \dots, \Phi_\ell)) = \max\{\mathcal{B}(\Phi_1), \mathcal{B}(\Phi_2), \dots, \mathcal{B}(\Phi_\ell)\}. \tag{13}$$

Proof. By assumption, we have

$$\Phi_i = \begin{cases} W_1^{(i)} & \text{if } L = 1 \\ W_L^{(i)} \circ \rho \circ W_{L-1}^{(i)} \circ \rho \dots \rho \circ W_1^{(i)} & \text{if } L \geq 2, \end{cases} \tag{14}$$

where $W_j^{(i)}(x) = A_j^{(i)}x + b_j^{(i)}$ for $i = 1, \dots, \ell$ and $j = 1, \dots, L$. Now, set $A_j = A_j^{(1)} \bigoplus A_j^{(2)} \bigoplus \cdots \bigoplus A_j^{(\ell)}$, $b_j = (b_j^{(1)}, b_j^{(2)}, \dots, b_j^{(\ell)})^\mathsf{T}$, and $W_j(x) = A_j x + b_j$ for $j = 1, \dots, L$. It is then easily seen that

$$P(\Phi_1, \Phi_2, \dots, \Phi_\ell) = \begin{cases} W_1 & \text{if } L = 1 \\ W_L \circ \rho \circ W_{L-1} \circ \rho \dots \rho \circ W_1 & \text{if } L \geq 2, \end{cases} \tag{15}$$

has the desired properties. □

3 Approximation of Lipschitz Functions

We start with the definition of the spike function, introduced in [15, Appendix F], which will be the basic building block for approximating Lipschitz functions in terms of ReLU neural networks.

Definition 4. The spike function $\phi \colon \mathbb{R}^m \to \mathbb{R}$ is defined according to

$$\phi(x) = \max\{1 + \min\{x_1, \dots, x_m, 0\} - \max\{x_1, \dots, x_m, 0\}, 0\}. \tag{16}$$

We have the following representation of the spike function in terms of a ReLU neural network.

Lemma 3. [26, Lemma V.1] Let ϕ be as in Definition 4. Then, we can construct a $\Phi \in \mathcal{N}_{m,1}$ with $\Phi(x) = \phi(x)$ for all $x \in \mathbb{R}^m$ satisfying the following properties:

$$\mathcal{L}(\Phi) \leq \lceil \log(m+1) \rceil + 4 \tag{17}$$
$$\mathcal{M}(\Phi) \leq 60m - 28 \tag{18}$$
$$\mathcal{W}(\Phi) \leq 6m \tag{19}$$
$$\mathcal{K}(\Phi) \subseteq \{0, 1, -1\} \tag{20}$$
$$\mathcal{B}(\Phi) = 1. \tag{21}$$

Moreover, Φ has the structure

$$\Phi = W_2 \circ \rho \circ \Psi \circ \rho \circ W_1 \tag{22}$$

with $W_1(x) = (I_m, -I_m)^\mathsf{T} x$, $W_2(x) = x$, and $\Psi \in \mathcal{N}_{2m,1}$.

We will need the following properties of ϕ:

Lemma 4. The spike funktion ϕ in (16) enjoys the following properties:

(i) $\phi(x) \in [0, 1]$ for all $x \in (-1, 1)^m$;

(ii) $\phi(x) = 0$ for all $x \in \mathbb{R}^m \setminus (-1, 1)^m$;

(iii) $\mathrm{Lip}(\phi) \leq 2$;

(iv) Let

$$\mathcal{S} = \{x \in [0, 1]^m : x_1 \geq x_2 \geq \cdots \geq x_m\} \tag{23}$$

and define $\mathcal{M} \subseteq \mathbb{N}_0^m$ according to

$$\mathcal{M} = \{0, e_1, e_1 + e_2, \dots, e_1 + \cdots + e_m\}. \tag{24}$$

Then,

$$\phi(x - n) = 0 \quad \text{for all } x \in \mathcal{S} \text{ and } n \in \mathbb{N}_0^m \setminus \mathcal{M}. \tag{25}$$

Moreover, we have

$$\phi(x) = 1 - x_1 \tag{26}$$
$$\phi(x - e_1 - e_2 \cdots - e_k) = x_k - x_{k+1} \quad \text{for } k = 1, \dots, m-1 \tag{27}$$
$$\phi(x - e_1 - e_2 \cdots - e_m) = x_m \tag{28}$$

for all $x \in \mathcal{S}$.

(v) We have

$$\sum_{n \in \mathbb{Z}^m} \phi(x - n) = 1 \quad \text{for all } x \in \mathbb{R}^m. \tag{29}$$

Proof. (i), (ii), and (v) follow from [26, Lemma III.6] and (iv) is by [26, (24) and (26)–(28)]. It remains to prove (iii). To this end, we write

$$\phi(x) = f_1 \circ (1 + f_2 - f_3) \tag{30}$$

with $f_1(x) = \max(x, 0)$, $f_2(x_1, \dots, x_d) = \min\{x_1, \dots, x_m, 0\}$, and $f_3(x_1, \dots, x_d) = \max\{x_1, \dots, x_m, 0\}$. Since $\mathrm{Lip}(f_1) = 1$ and $\mathrm{Lip}(f_2) = \mathrm{Lip}(f_3)$, we have

$$\mathrm{Lip}(\phi) \leq \mathrm{Lip}(f_1)(\mathrm{Lip}(f_2) + \mathrm{Lip}(f_3)) \leq 2\,\mathrm{Lip}(f_3). \tag{31}$$

Now, set $g(x) = \max\{x_1, \dots, x_d\}$. It can be easily shown by induction that $\mathrm{Lip}(g) \leq 1$ so that

$$\mathrm{Lip}(f_3) = \mathrm{Lip}(f_1 \circ g) \leq \mathrm{Lip}(f_1)\,\mathrm{Lip}(g) \leq 1. \tag{32}$$

Using (32) in (31) yields $\mathrm{Lip}(\phi) \leq 2$. □

As an immediate consequence, for every $N \in \mathbb{N}$, the functions $\{\phi(Nx - n)\}_{n\in\mathbb{N}}$ form a partition of unity on $\mathbb{R}^m$.

Corollary 1. Let $N \in \mathbb{N}$. Then, we have

$$\psi_N(x) := \sum_{n\in\mathbb{Z}^m} \phi(Nx - n) = 1 \quad \text{for all } x \in \mathbb{R}^m. \tag{33}$$

Proof. Since $\psi_N(x) = 1$ for all $x \in \mathbb{R}^m$ if and only if $\psi_N(x/N) = 1$ for all $x \in \mathbb{R}^m$, we can assume, without loss of generality, that $N = 1$. But $\psi_1(x) = 1$ for all $x \in \mathbb{R}^m$ owing to (v) in Lemma 4. □

We are now in a position to state our main approximation result for Lipschitz functions.

Theorem 1. Let $m, N \in \mathbb{N}$ and set $\mathcal{J} = \{0, 1, \ldots, N\}$. Further, consider a Lipschitz function $f\colon [0,1]^m \to \mathbb{R}$ and set $\mathcal{D}_f = \{x \in \mathbb{R} : \lvert x\rvert \leq \|f\|_{L_\infty([0,1]^m)}\}$. Suppose that $\delta \in [0,\infty)$ and let $h\colon \mathcal{D}_f \to \mathbb{R}$ be an arbitrary function satisfying $\lvert h(x) - x\rvert \leq \delta$ for all $x \in \mathcal{D}_f$. Finally, define $f_N\colon [0,1]^m \to \mathbb{R}$ according to

$$f_N(x) = \sum_{n\in\mathcal{J}^m} (h \circ f)\Big(\frac{n}{N}\Big)\phi(Nx - n) \tag{34}$$

with ϕ the spike function from Definition 4. Then, the following properties hold:

(i) $\|f - f_N\|_{L_\infty(\mathcal{A})} \leq \mathrm{Lip}(f)/N + \delta$;

(ii) We have

$$\mathrm{Lip}(f_N)/m \leq \mathrm{Lip}^{(1)}(f_N) \leq \mathrm{Lip}^{(1)}(f) + 2N\delta \leq \mathrm{Lip}(f) + 2N\delta. \tag{35}$$

Finally, consider the following ReLU neural networks:

(a) Let Ψ be as in Lemma 3, and set, for every $n \in \mathcal{J}^m$,

$$\Phi_{n,N} = \Psi \circ \rho \circ W_{n,N} \tag{36}$$

with $W_{n,N}(x) = N(I_m, -I_m)^\mathsf{T} x - (n, -n)^\mathsf{T}$.

(b) Set $F = (h(f(n/N)) : n \in \mathcal{J}^m)$, define $W\colon \mathbb{R}^{|\mathcal{J}^m|} \to \mathbb{R}$ according to $W(x) = Fx$, and define the ReLU neural network Φ as follows:

$$\Phi = W \circ \rho \circ P(\Phi_{n,N} : n \in \mathcal{J}^m). \tag{37}$$

Then, we have

$$\Phi(x) = f_N(x) \qquad \text{for all } x \in [0,1]^m \tag{38}$$

with

$$\mathcal{L}(\Phi) \leq \lceil \log(m+1)\rceil + 4 \tag{39}$$
$$\mathcal{M}(\Phi) \leq (N+1)^m(62m - 28) \tag{40}$$
$$\mathcal{W}(\Phi) \leq (N+1)^m 6m \tag{41}$$
$$\mathcal{B}(\Phi) \leq \max\{N, \|f\|_{L_\infty([0,1]^m)} + \delta\}. \tag{42}$$

Proof. First, define $\hat{f}_N : [0,1]^m \to \mathbb{R}$ according to

$$\hat{f}_N(x) = \sum_{n \in \mathcal{J}^m} f\Big(\frac{n}{N}\Big)\phi(Nx - n). \tag{43}$$

Next, note that

$$\begin{aligned} f(x) &= \sum_{n \in \mathbb{Z}^m} \phi(Nx-n) f(x) &(44)\\ &= \sum_{n \in \mathcal{J}^m} \phi(Nx-n) f(x) \quad \text{for all } x \in [0,1]^m, &(45)\end{aligned}$$

where (44) follows from Corollary 1 and in (45) we applied (ii) in Lemma 4. We therefore have

$$\begin{aligned} &|f(x) - f_N(x)| &(46)\\ &\le \Big|f(x) - \hat{f}_N(x)\Big| + \Big|\hat{f}_N(x) - f_N(x)\Big| &(47)\\ &\le \sum_{n \in \mathcal{J}^m} \Big(|f(x) - f(n/N)| + |(h \circ f)(n/N) - f(n/N)|\Big)\phi(Nx-n) &(48)\\ &\le \Big(\frac{\mathrm{Lip}(f)}{N} + \delta\Big) \sum_{n \in \mathcal{J}^m} \phi(Nx-n) &(49)\\ &\le \frac{\mathrm{Lip}(f)}{N} + \delta \quad \text{for all } x \in [0,1]^m, &(50)\end{aligned}$$

where (48) is by (44)–(45), in (49) we applied the Lipschitz property of f and $\big|h(x) - x\big| \le \delta$ for all $\mathcal{D}_f$, and (50) is again by Corollary 1. This establishes (i). We next prove (ii). The first and the third inequality in (35) follow from the norm inequalities $\|\cdot\|_1 \le m\|\cdot\|_\infty \le m\|\cdot\|_1$ in $\mathbb{R}^m$. It remains to establish the second inequality in (35). To this end, fix $k \in \{0, \dots, N-1\}^m$ arbitrarily and set $\mathcal{Q}_k = (0, 1/N)^m + \{k/N\}$ and

$$\mathcal{S}_k := \{x \in \mathcal{Q}_k : x_1 > x_2 > \cdots > x_m\}. \tag{51}$$

Now, let $\mathcal{M}$ be as defined in (24). Then, we have

$$\begin{aligned} f_N(x) &= \sum_{n \in \{k\}+M} (h \circ f)\Big(\frac{n}{N}\Big)\phi(Nx-n) &(52)\\ &= (h \circ f)(k/N) + N \sum_{i=1}^{m} x_i \Big((h \circ f)\Big(\frac{z_i}{N}\Big) - (h \circ f)\Big(\frac{z_{i-1}}{N}\Big)\Big) &(53)\end{aligned}$$

for all $x \in \mathcal{S}_k$, where in (52) we applied (25) and (53) follows from (26)–(28) with $z_0 = k$ and $z_i = k + e_1 + e_2 + \cdots + e_i$ for $i = 1, \dots, m$. We thus have

$$\begin{aligned} |(\partial_{x_i} f_N)(x)| &= N\Big|(h \circ f)\Big(\frac{z_i}{N}\Big) - (h \circ f)\Big(\frac{z_{i-1}}{N}\Big)\Big| &(54)\\ &\le N\Big|(h \circ f)\Big(\frac{z_i}{N}\Big) - f\Big(\frac{z_i}{N}\Big)\Big| + N\Big|f\Big(\frac{z_i}{N}\Big) - f\Big(\frac{z_{i-1}}{N}\Big)\Big| &(55)\\ &\quad + N\Big|f\Big(\frac{z_{i-1}}{N}\Big) - (h \circ f)f\Big(\frac{z_{i-1}}{N}\Big)\Big| &(56)\\ &\le \mathrm{Lip}^{(1)}(f) + 2N\delta \quad \text{for all } x \in \mathcal{S}_k \text{ and } i = 1, \dots, m, &(57)\end{aligned}$$

where (54) follows from (52)–(53) and in (57) we applied the Lipschitz property of f and $\big|h(x) - x\big| \leq \delta$ for all $x \in \mathcal{D}_f$. We can therefore conclude that

$$\begin{aligned} \mathrm{Lip}^{(1)}(f_N|_{\mathcal{S}_k}) &\leq \big\| \|\nabla f_N\|_\infty \big\|_{L_\infty(\mathcal{S}_k)} && (58)\\ &\leq \mathrm{Lip}^{(1)}(f) + 2N\delta, && (59) \end{aligned}$$

where (58) follows from the fact that f_N is an affine function on $\mathcal{S}_k$ owing to (52)–(53) and (59) is by (54)–(57). Since f_N as the finite sum of Lipschitz functions is Lipschitz on $[0,1]^m$, (58)–(59) yields

$$\mathrm{Lip}^{(1)}(f_N|_{\overline{\mathcal{S}_k}}) \leq \mathrm{Lip}^{(1)}(f) + 2N\delta. \tag{60}$$

Now, for every permutation $\pi\colon \{1,\dots,m\} \to \{1,\dots,m\}$, set

$$\mathcal{S}_k^{(\pi)} = \{x \in \mathcal{Q}_k : x_{\pi(1)} > x_{\pi(2)} > \dots > x_{\pi(m)}\}. \tag{61}$$

As f_N is a symmetric function and k was assumed to be arbitrary, (62) yields

$$\mathrm{Lip}^{(1)}(f_N|_{\overline{\mathcal{S}_k^{(\pi)}}}) \leq \mathrm{Lip}^{(1)}(f) + 2N\delta \tag{62}$$

for all such permutations π and $k \in \{0,\dots,N-1\}^m$. Finally, since the sets $\overline{\mathcal{S}_k^{(\pi)}}$ are all convex and closed and cover $[0,1]^m$, Lemma 12 in combination with (62) establishes

$$\mathrm{Lip}^{(1)}(f_N) \leq \mathrm{Lip}^{(1)}(f) + 2N\delta. \tag{63}$$

Alternatively, (63) can also be proved using the correspondence between Lipschitz functions and Sobolev spaces stated in Theorem 7. Concretely, as f_N is a symmetric function, (54)–(57) implies

$$\big\| \|\nabla f_N\|_\infty \big\|_{L_\infty(\mathcal{Q}_k)} \leq \mathrm{Lip}^{(1)}(f) + 2N\delta \quad \text{for } i = 1,\dots,m. \tag{64}$$

Since $\mathscr{L}^m((0,1)^m \backslash \bigcup_{k\in\{0,\dots,N-1\}^m} \mathcal{Q}_k) = 0$, by the arbitrariness of k, (64) yields

$$\big\| \|\nabla f_N\|_\infty \big\|_{L_\infty((0,1)^m)} \leq \mathrm{Lip}^{(1)}(f) + 2N\delta. \tag{65}$$

We therefore have

$$\begin{aligned} \mathrm{Lip}^{(1)}(f_N) &= \mathrm{Lip}^{(1)}(f_N|_{(0,1)^m}) && (66)\\ &\leq \big\| \|\nabla f_N\|_\infty \big\|_{L_\infty((0,1)^m)} && (67)\\ &\leq \mathrm{Lip}^{(1)}(f) + 2N\delta, && (68) \end{aligned}$$

where (66) follows from the fact that f_N as the finite sum of Lipschitz functions is Lipschitz on $[0,1]^m$ and $[0,1]^m = \overline{(0,1)^m}$, in (67) we applied Theorem 7, and (68) follows from (65), which again establishes (63). Finally, (38) follows from the fact that

$$\rho \circ \Phi_{n,N}(x) = \phi(Nx - n) \quad \text{for all } x \in [0,1]^m \text{ and } n \in \mathcal{J}^m \tag{69}$$

and (39)–(42) are by construction. □

Some comments are in order. First, note that only the last layer W of the ReLU neural network Φ in (37) depends on the specific choice of f. For the particular choice $h(x) = x$, i.e., $\delta = 0$, Theorem 1 recovers [14, Theorem 1.1]. Following the ideas in the proof of [14, Corollary 5.1], by composing $\Phi \circ \kappa$ with $\kappa\colon \mathbb{R}^m \to \mathbb{R}^m$ defined according to

$$\kappa(x) = \begin{pmatrix} \rho(x_1) - \rho(x_1 - 1) \\ \vdots \\ \rho(x_m) - \rho(x_m - 1) \end{pmatrix}, \tag{70}$$

we obtain a ReLU neural network satisfying $(\Phi \circ \kappa)(x) = \Phi(x)$ for all $x \in [0,1]^m$ with

$$\mathrm{Lip}^{(1)}(\Phi \circ \kappa) \leq \mathrm{Lip}^{(1)}(\Phi|_{[0,1]^m}). \tag{71}$$

Finally, and most importantly, we want to emphasize that the function h in Theorem 1 need not be Lipschitz. This property allows us now to particularize Theorem 1 to the case where h is a quantization function, which establishes that Lipschitz functions can be approximated by ReLU neural networks of bounded Lipschitz constants with quantized weights and solves the open problem pointed out in [14, Section 6.3]:

Corollary 2. Let $m, N \in \mathbb{N}$ and set

$$\mathcal{F}_N = \{k/N : k \in \mathbb{Z}\} \cap [-N, N]. \tag{72}$$

Then, we can construct a collection $\mathscr{K}_N^{(m)} \subseteq \mathcal{N}_{m,1}$ of ReLU neural networks with

$$\log\left(\left|\mathscr{K}_N^{(m)}\right|\right) = (N+1)^m \log(2N^2 + 1) \tag{73}$$

with the property that for every $\mathcal{A} \subseteq [0,1]^m$ and every Lipschitz function $f\colon \mathcal{A} \to \mathbb{R}$ satisfying

$$\|f\|_{L_\infty(\mathcal{A})} + \mathrm{Lip}(f) \leq N, \tag{74}$$

we can pick a $\Phi \in \mathscr{K}_N^{(m)}$ satisfying

(i) $\|f - \Phi\|_{L_\infty(\mathcal{A})} \leq (\mathrm{Lip}(f) + 1/2)/N$;

(ii) $\mathrm{Lip}(\Phi|_{[0,1]^m})/m \leq \mathrm{Lip}^{(1)}(\Phi|_{[0,1]^m}) \leq \mathrm{Lip}(f) + 1$.

Moreover, every $\Phi \in \mathscr{K}_N^{(m)}$ has the same architecture and satisfies:

$$\mathcal{L}(\Phi) \leq \lceil \log(m+1) \rceil + 4 \tag{75}$$

$$\mathcal{M}(\Phi) \leq (N+1)^m (62m - 28) \tag{76}$$

$$\mathcal{W}(\Phi) \leq (N+1)^m 6m \tag{77}$$

$$\mathcal{K}(\Phi) \subseteq \mathcal{F}_N \tag{78}$$

$$\mathcal{B}(\Phi) \leq N. \tag{79}$$

Proof. Fix a set $\mathcal{A} \subseteq [0,1]^m$ and a Lipschitz function $f\colon \mathcal{A} \to \mathbb{R}$ satisfying (74) arbitrarily. Further, set $\mathcal{J} = \{0, 1, \ldots, N\}$, Next, define $\tilde{f}\colon [0,1]^m \to \mathbb{R}$ according to

$$\tilde{f}(x) = \inf_{y \in \mathcal{A}} (f(y) + \mathrm{Lip}(f)\|x - y\|_\infty). \tag{80}$$

Then, $\tilde{f}(x) = f(x)$ for all $x \in \mathcal{A}$ and $\mathrm{Lip}(\tilde{f}) = \mathrm{Lip}(f)$ owing to [24, Theorem 1]. Moreover, by construction, we have $\|\tilde{f}\|_{L_\infty([0,1]^m)} \leq \|f\|_{L_\infty(\mathcal{A})} + \mathrm{Lip}(f)$. Now, consider the quantization function $g\colon [-\|\tilde{f}\|_{L_\infty([0,1]^m)}, \|\tilde{f}\|_{L_\infty([0,1]^m)}] \to \mathcal{F}_N$, $y \mapsto [Ny]/N$ and note that

$$|g(y) - y| \leq 1/(2N) \quad \text{for all } y \in [-\|\tilde{f}\|_{L_\infty([0,1]^m)}, \|\tilde{f}\|_{L_\infty([0,1]^m)}]. \tag{81}$$

Further, note that g is well-defined since $[N\tilde{f}(x)]/N \in \mathbb{Z}/N$ and

$$\begin{aligned} \left|[N\tilde{f}(x)]\right|/N &\leq [N\|\tilde{f}\|_{L_\infty([0,1]^m)}]/N &(82)\\ &\leq [N(\|f\|_{L_\infty(\mathcal{A})} + \mathrm{Lip}(f))]/N &(83)\\ &\leq [N^2]/N &(84)\\ &= N \quad \text{for all } x \in [0,1]^m. &(85)\end{aligned}$$

Now, let Φ be the ReLU neural network obtained by applying Theorem 1 to the Lipschitz function $\tilde{f}$ with $h = g$ and $\delta = 1/(2N)$. Then, (i) and (ii) follow from (i) and (ii) in Theorem 1, respectively, upon noting that $\Phi = f_N$ thanks to (b) in Theorem 1. Moreover, (75)–(77) follow from (39)–(41) and (78)–(79) are valid by construction. Finally, since Φ in (37) is completely determined by

$$F = (g(\tilde{f}(n/N)) : n \in \mathcal{J}^m) \in \mathcal{F}_N^{|\mathcal{J}^m|}, \tag{86}$$

we conclude that

$$\log\left(\left|\mathscr{K}_N^{(m)}\right|\right) = (N+1)^m \log(|\mathcal{F}_N|) = (N+1)^m \log(2N^2 + 1). \tag{87}$$

□

4 Rectifiable Measures

In this section, we show how to generate measures on low-dimensional objects using ReLU neural networks. Specifically, we are interested in the class of measures on $\mathbb{R}^n$ that are supported on Lipschitz images of compact sets in $\mathbb{R}^m$. To this end, we start with the definition of rectifiable sets.

Definition 5. Let $m, n \in \mathbb{N}$ with $m \leq n$. Then, $\mathcal{E} \subseteq \mathbb{R}^n$ is called

(i) m-rectifiable if there exists a compact and nonempty set $\mathcal{A} \subseteq \mathbb{R}^m$ and a Lipschitz mapping $f\colon \mathcal{A} \to \mathbb{R}^n$ such that $\mathcal{E} = f(\mathcal{A})$;

(ii) countably m-rectifiable if

$$\mathcal{E} = \bigcup_{i=1}^{\infty} \mathcal{E}_i \tag{88}$$

with $\mathcal{E}_i$ m-rectifiable for all $i \in \mathbb{N}$.

We need the following quantitative improvement of [29, Lemma 3.1], which states that finite unions of m-rectifiable sets are again m-rectifiable.

Lemma 5. Let $\ell \in \mathbb{N}$. For $k = 1, \ldots, \ell$, let $\mathcal{A}_k \subseteq \mathbb{R}^m$ be compact and let $f_k \colon \mathcal{A}_k \to \mathbb{R}^n$ be Lipschitz. Further, for $k = 1, \ldots, \ell$, fix a closed cube $\mathcal{Q}_k \subseteq \mathbb{R}^m$ of sidelength s_k such that $\mathcal{A}_k \subseteq \mathcal{Q}_k$ and consider the m-rectifiable set $\mathcal{E} = \bigcup_{k=1}^{\ell} f_k(\mathcal{A}_k)$. Then, we can construct a compact set $\mathcal{B} \subseteq [0,1]^m$ and a Lipschitz mapping $g \colon \mathcal{B} \to \mathbb{R}^n$ with $\|g\|_{L_\infty(\mathcal{B})} = \max_{k=1,\ldots,\ell} \|f_k\|_{L_\infty(\mathcal{A}_k)}$ and

$$\operatorname{Lip}(g) \leq 2\ell \max\left\{ \operatorname{diam}(\mathcal{E}), s_1 \operatorname{Lip}(f_1), s_2 \operatorname{Lip}(f_2), \ldots, s_\ell \operatorname{Lip}(f_\ell) \right\} \tag{89}$$

satisfying $\mathcal{E} = g(\mathcal{B})$.

Proof. For $k = 1, \ldots, m$, let c_k denote the center of $\mathcal{Q}_k$ and define the affine bijections $\phi_k \colon [0,1]^m \to \mathcal{Q}_k$ according to

$$\phi_k(x) = s_k x + c_k - (s_k/2)(e_1 + e_2 + \cdots + e_m). \tag{90}$$

Further, for $k = 1, \ldots, \ell$, define the affine bijections $\psi_k \colon \mathbb{R}^m \to \mathbb{R}^m$ according to

$$\psi_k(x) = (2\ell x_1 - 2(k-1), x_2, x_3, \ldots, x_m)^\mathsf{T} \tag{91}$$

and set $\mathcal{B}_k = \psi^{-1}(\phi^{-1}(\mathcal{A}_k))$. Since

$$\psi_k^{-1}(x) = (x_1/(2\ell) + (k-1)/\ell, x_2, x_3, \ldots, x_m)^\mathsf{T} \tag{92}$$

and $\phi^{-1}(\mathcal{A}_k) \subseteq [0,1]^m$, we have

$$\mathcal{B}_k \subseteq [(k-1)/\ell, (2k-1)/2\ell] \times [0,1]^{m-1} \quad \text{for } k = 1, \ldots, m \tag{93}$$

by construction. What is more, the sets $\mathcal{B}_k$ are compact and, by (93), pairwise disjoint subsets of $[0,1]^m$. Now, set $\mathcal{B} = \bigcup_{k=1}^{\ell} \mathcal{B}_k$ and define $g \colon \mathcal{B} \to \mathbb{R}^n$ according to

$$g(x) = f_k(\phi_k(\psi_k(x))) \quad \text{for all } x \in \mathcal{B}_k \text{ and } k = 1, \ldots, \ell \tag{94}$$

so that $\mathcal{E} = g(\mathcal{B})$. If $x, y \in \mathcal{B}_k$, then

$$\|g(x) - g(y)\|_\infty \leq \operatorname{Lip}(\phi_k) \operatorname{Lip}(\psi_k) \operatorname{Lip}(f_k) \|x - y\|_\infty = 2\ell s_k \operatorname{Lip}(f_k) \|x - y\|_\infty. \tag{95}$$

If $x \in \mathcal{B}_{k_1}$ and $y \in \mathcal{B}_{k_2}$ with $k_1 \neq k_2$, then $\|x - y\|_\infty \geq 1/(2\ell)$ thanks to (93), which implies

$$\frac{\|g(x) - g(y)\|_\infty}{\|x - y\|_\infty} \leq 2\ell \operatorname{diam}(\mathcal{E}). \tag{96}$$

Combining (95) and (96) yields (89). ☐

We next introduce measures on (countably) m-rectifiable sets.

Definition 6. Let ν be a finite measure on $\mathbb{R}^n$. Then, we call ν

(i) m-rectifiable subordinary to $\mathcal{E}$ if it is Borel regular and supported on the m-rectifiable set $\mathcal{E} \subseteq \mathbb{R}^n$;

(ii) $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$ if $\nu = \phi(\mathscr{H}^m|_{\mathcal{E}})$, where $\mathcal{E} \subseteq \mathbb{R}^n$ is m-rectifiable and $\phi\colon \mathbb{R}^n \to [0, \infty)$ is a $\mathscr{H}^m$-measurable mapping;

(iii) countably m-rectifiable subordinary to $\mathcal{E}$ if it is Borel regular and supported on the countably m-rectifiable set $\mathcal{E} \subseteq \mathbb{R}^n$;

(iv) countably $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$ if $\nu = \phi(\mathscr{H}^m|_{\mathcal{E}})$, where $\mathcal{E} \subseteq \mathbb{R}^n$ is countably m-rectifiable and $\phi\colon \mathbb{R}^n \to [0, \infty)$ is a Borel measurable mapping.

One could ask why the assumption of Borel-regularity is missing for the measures in (ii) and (iv) of Definition 6. It turns out that these measures are inherently Borel-regular, which is established next.

Lemma 6. (Countably) $\mathscr{H}^m$-rectifiable measures are Borel-regular. In particular, the measures in (i)–(iv) of Definition 6 are all Radon measures, and we have the implications (ii) $\Rightarrow$ (i) and (iv) $\Rightarrow$ (iii) in Definition 6.

Proof. Let ν be (countably) $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$. Since $\mathscr{H}^m|_{\mathcal{E}}$ is Borel regular owing to (ii) in Theorem 10 and ϕ is $\mathscr{H}^m$-measurable, by Lemma 18, $\nu = \phi(\mathscr{H}^m|_{\mathcal{E}})$ must be a Borel regular measure. The measures in (i)–(iv) of Definition 6 are therefore all Borel-regular and, in turn, Radon measures as they are finite by assumption. The implications (ii) $\Rightarrow$ (i) and (iv) $\Rightarrow$ (iii) in Definition 6 are trivial. $\square$

The following result states that m-rectifiable and $\mathscr{H}^m$-rectifiable measures can be generated as push-forwards of Radon measures.

Theorem 2. Let $\mathcal{A} \subseteq \mathbb{R}^m$ be compact, suppose that $f\colon \mathcal{A} \to \mathbb{R}^n$ is a Lipschitz mapping, and consider the m-rectifiable set $\mathcal{E} = f(\mathcal{A})$. Then, the following properties hold.

(i) If ν is m-rectifiable subordinary to $\mathcal{E}$, then, there exists a Radon measure μ on $\mathcal{A}$ such that $\nu = f\#\mu$.

(ii) Suppose that $\phi\colon \mathbb{R}^n \to [0, \infty)$ is a $\mathscr{H}^m$-measurable mapping and ν is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$. Then, there exists a Radon measure μ on $\mathcal{A}$ such than $\nu = f\#\mu$ with $\mu \ll \mathscr{L}^m|_{\mathcal{A}}$. Moreover, we have $\dim_{\mathrm{H}}(\mu) = m$ provided that ν is nontrivial.

Proof. In (i) and (ii), ν is supported on the m-rectifiable set $\mathcal{E} \subseteq \mathbb{R}^n$. The existence of a Radon measure μ on $\mathcal{A}$ satisfying $\nu = f\#\mu$ is established in Corollary 9. We next prove that $\mu \ll \mathscr{L}^m|_{\mathcal{A}}$ if ν is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$. Toward a contradiction, suppose that there exists a set $\mathcal{C} \subseteq \mathcal{A}$ such that $\mu(\mathcal{C}) > 0$ and $\mathscr{L}^m(\mathcal{C}) = 0$. Then, we have

$$\nu(f(\mathcal{C})) = \mu(f^{-1}(f(\mathcal{C}))) \geq \mu(\mathcal{C}) > 0. \tag{97}$$

Now, Borel regularity of $\mathscr{L}^m$ implies that there exists a Borel set $\mathcal{B} \subseteq \mathbb{R}^m$ with $\mathcal{C} \subseteq \mathcal{B}$ and $\mathscr{L}^m(\mathcal{B}) = \mathscr{L}^m(\mathcal{C}) = 0$. Using the properties of Hausdorff measures in Lemma 21, we obtain We thus have

$$\mathscr{H}^m(f(\mathcal{C})) \leq \mathrm{Lip}^m(f)\mathscr{H}^m(\mathcal{C}) \leq \mathrm{Lip}^m(f)\mathscr{H}^m(\mathcal{B}) = \mathrm{Lip}^m(f)\mathscr{L}^m(\mathcal{B}) = 0, \tag{98}$$

which is in contradiction to (97) since $\nu \ll \mathscr{H}^m$ by Lemma 18. Hence, $\mu(\mathcal{C}) > 0$ implies $\mathscr{L}^m(\mathcal{C}) > 0$ so that $\mu \ll \mathscr{L}^m|_{\mathcal{A}}$.

It remains to show that $\dim_{\mathrm{H}}(\mu) = m$ if ν is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$ and nontrivial. The elementary properties of Hausdorff dimension [9, Section 3.2] imply that Hausdorff dimension can not exceed the ambient dimension so that $\dim_{\mathrm{H}}(\mu) \leq m$. Next, note that $\nu(f(\mathrm{spt}(\mu)) > 0$ since $\nu = f\#\mu$ and ν was supposed to be nontrivial. This implies $\mathscr{H}^m(f(\mathrm{spt}(\mu))) > 0$ as $\nu \ll \mathscr{H}^m$. We thus have $\mathscr{H}^m(\mathrm{spt}(\mu)) > 0$ by (ii) of Lemma 21. Therefore, $\dim_{\mathrm{H}}(\mu) \geq m$ owing to the definition of Hausdorff dimension Definition 20. $\square$

Note that, in Theorem 2, μ in (i) can have arbitrary small $\dim_{\mathrm{H}}(\mu)$ as we did not impose any regularity conditions on ν, whereas μ in (ii) satisfies $\dim_{\mathrm{H}}(\mu) = m$, which is a consequence of $\nu \ll \mathscr{H}^m$. In (ii) of Theorem 2, if the functon ϕ is Borel regular and f is injective, we have an explicit form of μ:

Lemma 7. Let $\mathcal{A} \subseteq \mathbb{R}^m$ be compact, suppose that $f\colon \mathcal{A} \to \mathbb{R}^n$ is an injective Lipschitz mapping, and consider the m-rectifiable set $\mathcal{E} = f(\mathcal{A})$. Suppose that $\phi\colon \mathbb{R}^n \to [0, \infty)$ is a Borel measurable mapping and ν is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$. Then, $\nu = f\#\mu$ with $\psi = (Jf)(\phi \circ f)$.

Proof. We first establish that ψ is Lebesgue measurable my inspecting the individual components: The function $\phi \circ f\colon \mathcal{A} \to \mathbb{R}$ as the composition of a Borel measurable mappings is Borel measurable. The function $Jf : \mathcal{A} \to \mathbb{R}$ is Lebesgue measurable thanks to Rademacher's theorem [18, Theorem 5.1.11]. Next, we prove that μ is a finite measure. We have

$$\mu(\mathcal{A}) = \int_{\mathcal{A}} Jf\,(\phi \circ f)\,\mathrm{d}\mathscr{L}^m \tag{99}$$

$$= \int_{f(\mathcal{A})} \phi\,\mathrm{d}\mathscr{H}^m \tag{100}$$

$$= \nu(f(\mathcal{A})) \tag{101}$$

$$\leq \nu(\mathcal{E}) \tag{102}$$

$$< \infty, \tag{103}$$

where in (100) we used Theorem 8 upon noting that

$$\sum_{x \in \mathcal{A} \cap f^{-1}(y)} \phi(f(x)) = \phi(y)\chi_{f(\mathcal{A})}(y) \quad \text{for all } y \in \mathbb{R}^n, \tag{104}$$

which implies that μ is a finite measure. By Lemma 18, $\mu = \psi\mathscr{L}^m$ is also Borel regular since ψ is Lebesgue measurable and $\mathscr{L}^m$ is Borel regular. Therefore, by Lemma 14, μ is a Radon measure. Now, fix a Borel set $\mathcal{C} \subseteq \mathbb{R}^n$ arbitrarily. Then, we have

$$(f\#\mu)(\mathcal{C}) = \int_{f^{-1}(\mathcal{C})} \psi\,\mathrm{d}\mathscr{L}^m \tag{105}$$

$$= \int h\,\mathrm{d}\mathscr{H}^m \tag{106}$$

$$= \int_{\mathcal{C}} \phi\,\mathrm{d}\mathscr{H}^m \tag{107}$$

$$= (\phi\,\mathscr{H}^m)(\mathcal{C}) \tag{108}$$

$$= \nu(\mathcal{C}), \tag{109}$$

where (106) follows from Theorem 8 with

$$h(y) = \sum_{x \in f^{-1}(\mathcal{C}) \cap f^{-1}(y)} \frac{\psi(x)}{(Jf)(x)} \tag{110}$$

$$= \phi(y)\chi_{\mathcal{C}}(y) \quad \text{for all } y \in \mathbb{R}^n, \tag{111}$$

upon noting that $f^{-1}(\mathcal{C})$ is Borel measurable since f is continuous and $\mathcal{C}$ is Borel. Now, ν is a Radon measure owing to Lemma 6. Since μ is a Radon measure, so is $f\#\mu$ thanks to Theorem 11. Summarizing, ν and $f\#\mu$ are both Radon measure so that (105)–(109) in combination with (105)–(109) yields $\nu = f\#\mu$ by Theorem 11. ☐

We next show that the class of m-rectifiable and $\mathscr{H}^m$-rectifiable measures is closed under addition.

Lemma 8. Let $\mathcal{E}_1, \mathcal{E}_2 \subseteq \mathbb{R}^n$ be m-rectifiable. Then, the following properties hold:

(i) If ν_k is m-rectifiable subordinary to $\mathcal{E}_k$ for $k = 1, 2$, then $\nu_1 + \nu_2$ is m-rectifiable subordinary to $\mathcal{E}_1 \cup \mathcal{E}_2$.

(ii) Suppose that $\phi_k \colon \mathbb{R}^n \to [0, \infty)$ is Borel-regular and ν_k is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}_k, \phi_k)$ for $k = 1, 2$. Then, $\nu_1 + \nu_2$ is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}_1 \cup \mathcal{E}_2, \chi_{\mathcal{E}_1} \phi_1 + \chi_{\mathcal{E}_2} \phi_2)$.

Proof. We first prove (i). The set $\mathcal{E}_1 \cup \mathcal{E}_2$ is m-rectifiable owing to Lemma 5. Since $\nu_1 + \nu_2$ is supported on $\mathcal{E}_1 \cup \mathcal{E}_2$ and sums of Borel-regular measures are again Borel regular thanks to Lemma 15, we conclude that $\nu_1 + \nu_2$ is m-rectifiable. Next, we establish (ii). Set $\phi = \chi_{\mathcal{E}_1} \phi_1 + \chi_{\mathcal{E}_2} \phi_2$ and consider the measure $\nu = \phi\, \mathscr{H}^m|_{\mathcal{E}_1 \cup \mathcal{E}_2}$. Since $\mathcal{E}_1 \cup \mathcal{E}_2$ is m-rectifiable owing to Lemma 5, ν is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}_1 \cup \mathcal{E}_2, \phi)$. It remains to show that $\nu = \nu_1 + \nu_2$. Since ν and $\nu_1 + \nu_2$ are Borel regular measures thanks to Lemma 6 and Lemma 15 and

$$\nu(\mathcal{B}) = \int_{\mathcal{B}} \phi\, \mathscr{H}^m|_{\mathcal{E}_1 \cup \mathcal{E}_2} \tag{112}$$

$$= \int_{\mathcal{B}} \phi_1 \, \mathrm{d}\mathscr{H}^m|_{\mathcal{E}_1} + \int_{\mathcal{B}} \phi_2 \, \mathrm{d}\mathscr{H}^m|_{\mathcal{E}_2} \tag{113}$$

$$= \nu_1(\mathcal{B}) + \nu_2(\mathcal{B}) \quad \text{for all Borel sets } \mathcal{B} \subseteq \mathbb{R}^n, \tag{114}$$

we have $\nu = \nu_1 + \nu_2$ by Lemma 16. Thus, $\nu_1 + \nu_2$ is $\mathscr{H}^m$-rectifiable. ☐

The following result states that countably m-rectifiable measures can be approximated by m-rectifiable measures. The same holds for countably $\mathscr{H}^m$-rectifiable measures.

Lemma 9. Suppose that $\mathcal{A}_k \subseteq \mathbb{R}^m$ is compact and $f_k \colon \mathcal{A}_k \to \mathbb{R}^n$ is Lipschitz for all $k \in \mathbb{N}$ and consider the m-rectifialbe set $\mathcal{E} = \bigcup_{k=1}^{\infty} f_k(\mathcal{A}_k)$. Suppose that $\mathcal{E}$ is bounded and set $\mathcal{E}_\ell = \bigcup_{k=1}^{\ell} f_k(\mathcal{A}_k)$ for every $\ell \in \mathbb{N}$. Then, the following properties hold:

(i) Suppose that ν is countably m-rectifiable subordinary to $\mathcal{E}$. Then, the measure $\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell)$ is m-rectifiable subordinary to $\mathcal{E}_\ell$ and satisfies

$$W_1\big(\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell), \nu\big) \leq (1 - \nu(\mathcal{E}_\ell)) \operatorname{diam}(\mathcal{E}) \quad \text{for all } \ell \in \mathbb{N}. \tag{115}$$

(ii) Suppose that $\phi\colon \mathbb{R}^n \to [0, \infty)$ is Borel-regular and ν is countably $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$. Then, the measure $\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell)$ is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}_\ell, \phi/\nu(\mathcal{E}_\ell))$ and satisfies (115) for all $\ell \in \mathbb{N}$.

Proof. Fix $\ell \in \mathbb{N}$ arbitrarily. Suppose that ν is countably m-rectifiable subordinary to $\mathcal{E}$. In particular, this implies that ν is Borel-regular. Moreover, $\mathcal{E}_\ell$ is m-rectifiable owing to Lemma 5. Since $\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell)$ is supported on $\mathcal{E}_\ell$ and Borel-regular thanks to (ii) of Theorem 10, we conclude that $\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell)$ is m-rectifiable subordinary to $\mathcal{E}_\ell$. Next, suppose that ν is countably $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}, \phi)$. Then, we have

$$\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell) = (\phi(\mathscr{H}^m|_{\mathcal{E}}))|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell) \tag{116}$$

$$= \phi/\nu(\mathcal{E}_\ell)(\mathscr{H}^m|_{\mathcal{E}_\ell}), \tag{117}$$

where (117) follows from Lemma 18. Since $\mathcal{E}_\ell$ as the finite union of m-rectifiable sets is m-rectifiable owing to Lemma 5, we conclude that $\nu|_{\mathcal{E}_\ell}/\nu(\mathcal{E}_\ell)$ is $\mathscr{H}^m$-rectifiable subordinary to $(\mathcal{E}_\ell, \phi/\nu(\mathcal{E}_\ell))$. It remains to prove (115) for (i) and (ii). Set $\delta_\ell = 1 - \nu(\mathcal{E}_\ell)$. Then, we have

$$W_1\big(\nu|_{\mathcal{E}_\ell}/(1-\delta_\ell), \nu\big) = W_1\Big((1-\delta_\ell)(\nu|_{\mathcal{E}_\ell}/(1-\delta_\ell)) + \delta_\ell(\nu|_{\mathcal{E}_\ell}/(1-\delta_\ell)), \tag{118}$$

$$(1-\delta_\ell)(\nu|_{\mathcal{E}_\ell}/(1-\delta_\ell)) + \delta_\ell(\nu|_{\mathcal{E}\setminus\mathcal{E}_\ell}/\delta_\ell)\Big) \tag{119}$$

$$\leq W_1(\nu|_{\mathcal{E}_\ell}, \nu|_{\mathcal{E}_\ell}) + \delta_\ell W_1\big(\nu|_{\mathcal{E}_\ell}/(1-\delta_\ell), \nu|_{\mathcal{E}\setminus\mathcal{E}_\ell}/\delta_\ell\big) \tag{120}$$

$$\leq \delta_\ell \operatorname{diam}(\mathcal{E}), \tag{121}$$

where (120) follows from Lemma 28 and in (121) we applied Lemma 23. □

4.1 Generation of Rectifiable Measures

We are now in a position to state our results on the generation of rectifiable measures through ReLU neural networks. We start with m-rectifiable measures.

Theorem 3. Let $m, n, N \in \mathbb{N}$ and set

$$\mathcal{F}_N = \{k/N : k \in \mathbb{Z}\} \cap [-N, N]. \tag{122}$$

Then, we can construct a collection $\mathscr{K}_N^{(m,n)} \subseteq \mathcal{N}_{m,n}$ of ReLU neural networks with

$$\log\left(\left|\mathscr{K}_N^{(m,n)}\right|\right) = n(N+1)^m \log(2N^2+1) \tag{123}$$

with the property that for every compact set $\mathcal{A} \subseteq \mathbb{R}^m$, every Lipschitz mapping $f\colon \mathcal{A} \to \mathbb{R}^n$ satisfying

$$\|f\|_{L_\infty(\mathcal{A})} + \operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) \leq N, \tag{124}$$

and every measure ν that is m-rectifiable subordinary to $f(\mathcal{A})$, there is a $\Phi \in \mathscr{K}_N^{(m,n)}$, a compact set $\mathcal{B} \subseteq [0,1]^m$, and a Radon measure μ on $\mathcal{B}$ such that $\Phi|_{\mathcal{B}}\#\mu$ is m-rectifiable subordinary to $\Phi(\mathcal{B})$ and satisfies

$$\mathrm{Lip}(\Phi|_{[0,1]^m}) \leq m(\mathrm{diam}(\mathcal{A})\,\mathrm{Lip}(f) + 1) \tag{125}$$

and

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu) \leq \frac{\nu(\mathcal{E})(\mathrm{diam}(\mathcal{A})\,\mathrm{Lip}(f) + 1/2)}{N}. \tag{126}$$

Moreover, every $\Phi \in \mathscr{K}_N^{(m,n)}$ has the same architecture and obeys

$$\mathcal{L}(\Phi) \leq \lceil \log(m+1) \rceil + 4 \tag{127}$$
$$\mathcal{M}(\Phi) \leq n(N+1)^m(62m - 28) \tag{128}$$
$$\mathcal{W}(\Phi) \leq n(N+1)^m 6m \tag{129}$$
$$\mathcal{K}(\Phi) \subseteq \mathcal{F}_N \tag{130}$$
$$\mathcal{B}(\Phi) \leq N. \tag{131}$$

Proof. Fix a compact set $\mathcal{A} \subseteq \mathbb{R}^m$ and a Lipschitz function $f\colon \mathcal{A} \to \mathbb{R}$ satisfying (124) arbitrary and let $\mathcal{Q}$ be a closed cube of side-length $\mathrm{diam}(\mathcal{A})$ such that $\mathcal{A} \subseteq \mathcal{Q}$, denote by $\phi\colon [0,1]^m \to \mathcal{Q}$ the affine mapping satisfying $\mathcal{Q} = \phi([0,1]^m)$, and set $\mathcal{B} = \phi^{-1}(\mathcal{A})$. The function $g = (g_1, g_2, \ldots, g_n)^\mathsf{T}\colon \mathcal{B} \to \mathbb{R}^n$ defined according to $g = f \circ \phi$ is Lipschitz with $\|g\|_{L^\infty(\mathcal{B})} = \|f\|_{L^\infty(\mathcal{A})}$ and $\mathrm{Lip}(g) \leq \mathrm{diam}(\mathcal{A})\,\mathrm{Lip}(f)$ and satisfies $\mathcal{E} = g(\mathcal{B})$. Moreover, by (i) in Theorem 2, there exists a Radon measure μ on $\mathcal{B}$ such that $\nu = g\#\mu$. Now, let Φ_j denote the ReLU neural network resulting from applying Corollary 2 to the Lipschitz mappings g_j satisfying

$$\|g_j - \Phi_j\|_{L^\infty(\mathcal{B})} \leq \frac{\mathrm{diam}(\mathcal{A})\,\mathrm{Lip(f)} + 1/2}{N} \tag{132}$$

and

$$\mathrm{Lip}(\Phi_j|_{[0,1]^m}) \leq m(\mathrm{diam}(\mathcal{A})\,\mathrm{Lip}(f) + 1) \quad \text{for } j = 1, \ldots, n \tag{133}$$

and set $\Phi = P(\Phi_1, \ldots, \Phi_n)$. Then, (133) implies that Φ satisfies (125). Moreover, properties (75)–(79) for Φ_j listed in Corollary 2 imply that Φ satisfies (127)–(131). Next, note that

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu) = W_1(g\#\mu, \Phi\#\mu) \tag{134}$$
$$\leq \int \|g - \Phi|_{\mathcal{B}}\|_{L_\infty(\mathcal{B})} \mathrm{d}\mu \tag{135}$$
$$= \int \max_{j=1,\ldots,n} |g_j - \Phi_j| \mathrm{d}\mu \tag{136}$$
$$\leq \frac{\mu(\mathcal{B})(\mathrm{diam}(\mathcal{A})\,\mathrm{Lip(f)} + 1/2)}{N} \tag{137}$$
$$= \frac{\nu(\mathcal{E})(\mathrm{diam}(\mathcal{A})\,\mathrm{Lip(f)} + 1/2)}{N}, \tag{138}$$

where (135) follows form Lemma 26 and (137) is by (132), which proves (126). Now, $\Phi|_{\mathcal{B}}\#\mu$ is a Radon measure by Theorem 11 and supported on the m-rectifiable set $\Phi(\mathcal{B})$. Therefore, $\Phi|_{\mathcal{B}}\#\mu$ is m-rectifiable subordinary to $\Phi(\mathcal{B})$. $\square$

We can combine Theorem 3 with Corollary 8 to obtain the following space-filling result for m-rectifiable probability measures.

Corollary 3. Let $n, N \in \mathbb{N}$ and set

$$\mathcal{D}_{3N} = \{a/b : a \in \mathbb{Z}, b \in \mathbb{N}, |a| \leq 4(3N)^{m+1}, \text{ and } b \leq 4(3N)^{m+2}\}. \tag{139}$$

Then, we can construct a collection $\mathscr{I}_N^{(m,n)} \subseteq \mathcal{N}_{1,n}$ of ReLU neural networks with

$$\log\left(\left|\mathscr{I}_N^{(m,n)}\right|\right) \leq 3n(3N)^m \log(6N) \tag{140}$$

with the property that for every compact set $\mathcal{A} \subseteq \mathbb{R}^m$, every Lipschitz mapping $f : \mathcal{A} \to \mathbb{R}^n$ satisfying

$$\|f\|_{L_\infty(\mathcal{A})} + \operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) \leq N, \tag{141}$$

and every measure ν that is m-rectifiable subordinary to $f(\mathcal{A})$, there is a $\Psi \in \mathscr{I}_N^{(m,n)}$ satisfying

$$\operatorname{Lip}(\Psi|_{[0,1]}) \leq m(6(N+1))^{m+1} \tag{142}$$

and

$$W_1\left(\nu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\right) \leq \frac{(m+1)(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1)}{N}. \tag{143}$$

Moreover, every $\Psi \in \mathscr{I}_N^{(m,n)}$ has the same architecture and obeys

$$\mathcal{L}(\Psi_N) \leq \lceil \log(m+1) \rceil + 6(m-1) + 7 \tag{144}$$
$$\mathcal{M}(\Psi_N) \leq 78mn(3N)^m \tag{145}$$
$$\mathcal{W}(\Psi_N) \leq 6mn(3N)^m \tag{146}$$
$$\mathcal{K}(\Psi_N) \subseteq \mathcal{D}_{3N} \tag{147}$$
$$\mathcal{B}(\Psi_N) \leq 4(3N)^{m+1}. \tag{148}$$

Proof. Consider the collections $\mathscr{K}_N^{(m,n)}$ and $\mathscr{G}_{3N}^{(m)}$ from Theorem 3 and Corollary 8 (with $d = m$ and $K = 3N$), respectively and set

$$\mathscr{I}_N^{(m,n)} = \{\Phi \circ \rho \circ \Sigma : \Phi \in \mathscr{K}_N^{m,n} \text{ and } \Sigma \in \mathscr{G}_{3N}^{(m)}\}. \tag{149}$$

Then, we have

$$\left|\mathscr{I}_N^{(m,n)}\right| \leq \left|\mathscr{G}_{3N}^{(m)}\right|\left|\mathscr{K}_N^{(m,n)}\right| \tag{150}$$
$$\leq (2N^2+1)^{n(N+1)^m}(36N)^{(3N)^m} \tag{151}$$
$$\leq (6N)^{3n(3N)^m}, \tag{152}$$

which establishes (140). Since every $\Phi \in \mathscr{K}_N^{m,n}$ satisfies (127)–(131) and every $\Sigma \in \mathscr{G}_{3N}$ obeys (395)–(400) for $d = m$ and $K = 3N$, we conclude that every $\Psi \in \mathscr{I}_N^{(m,n)}$ satisfies (144)–(148) upon noting that

$$\mathcal{F}_N = \{k/N : k \in \mathbb{Z}\} \cap [-N, N] \subseteq \mathcal{D}_{3N}. \tag{153}$$

Now, fix a compact set $\mathcal{A} \subseteq \mathbb{R}^m$, a Lipschitz function $f\colon \mathcal{A} \to \mathbb{R}$ satisfying (141), and a measure ν that is m-rectifiable subordinary to $f(\mathcal{A})$ arbitrarily. Then, Theorem 3 implies that there exists a compact set $\mathcal{B} \subseteq [0,1]^m$, a Radon measure μ on $\mathcal{B}$, and a $\Phi \in \mathscr{K}_N^{m,n}$ satisfying

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu) \leq \frac{(m+1)(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1)}{N} \tag{154}$$

and

$$\operatorname{Lip}(\Phi|_{[0,1]^m}) \leq m(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1). \tag{155}$$

Moreover, by Corollary 8, there is a $\Sigma \in \mathscr{G}_{3N}^{(m)}$ satisfying

$$\operatorname{Lip}(\Sigma) \leq 2^{m+1}(3N)^m \tag{156}$$

and

$$W_1\Big(\mu, \Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \leq 1/N. \tag{157}$$

We set $\Psi = \Phi \circ \rho \circ \Sigma$ and can therefore conclude that

$$\operatorname{Lip}(\Psi|_{[0,1]}) \leq m(6(N+1))^{m+1} \tag{158}$$

and

$$
\begin{aligned}
&W_1\Big(\nu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) && (159)\\
&\leq W_1\Big(\nu, \Phi|_{\mathcal{B}}\#\mu\Big) + W_1\Big(\Phi|_{\mathcal{B}}\#\mu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) && (160)\\
&\leq \frac{\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1/2}{N} + \operatorname{Lip}(\Phi|_{[0,1]^m}) W_1\Big(\mu, \Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) && (161)\\
&\leq \frac{\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1/2}{N} + \frac{m(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1)}{N} && (162)\\
&\leq \frac{(m+1)(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1)}{N}, && (163)
\end{aligned}
$$

where (161) follows from (154) and Lemma 27 and in (162) we applied (155) and (157). ☐

We next introduce classes of m-rectifiable measures, which will allow us to derive universality result on the generation of the measures in the individual classes.

Definition 7. Let $m, n \in \mathbb{N}$ and $C \in (0, \infty)$. The class $\mathscr{G}(n, m, C)$ consists of all m-rectifiable probability measures ν on $\mathbb{R}^n$ with the property that there exists an $\mathcal{A} \subseteq \mathbb{R}^m$ and a Lipschitz mapping $f\colon \mathcal{A} \to \mathbb{R}^n$ satisfying

$$\max\Big\{(m+1)(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1),\ \|f\|_{L_\infty(\mathcal{A})} + \operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f)\Big\} \leq C \tag{164}$$

such that ν is m-rectifiable subordinary to $f(\mathcal{A})$.

The next result is a universality result on the approximation of m-rectifiable probability measures.

Corollary 4. Let $m, n \in \mathbb{N}$ and $C \in (0, \infty)$ and consider the class $\mathscr{G}(n, m, C)$ as defined in Definition 7. Then, for every $\varepsilon \in (0, 1]$, there exists a collection $\mathscr{R}(n, m, C, \varepsilon)$ of ReLU neural networks of cardinality

$$|\mathscr{R}(n, m, C, \varepsilon)| \leq 2^{b(\varepsilon)} \tag{165}$$

with

$$b(\varepsilon) = 3n(3\lceil C/\varepsilon \rceil)^m \log(6\lceil C/\varepsilon \rceil) \tag{166}$$

such that for every $\nu \in \mathscr{G}(n, m, C)$, there is a $\Psi \in \mathscr{R}(n, m, C, \varepsilon)$ satisfying

$$\operatorname{Lip}(\Psi|_{[0,1]}) \leq m(6(\lceil C/\varepsilon \rceil + 1))^{m+1} \tag{167}$$

and

$$W_1\Big(\nu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \leq \varepsilon. \tag{168}$$

Moreover, every $\Psi \in \mathscr{R}(n, m, C, \varepsilon)$ has the same architecture and satisfies (144)–(148) with $N = \lceil C/\varepsilon \rceil$.

Proof. Follows from Corollary 3 with $N = \lceil C/\varepsilon \rceil$ and $\mathscr{R}(n, m, C, \varepsilon) = \mathscr{I}_N^{(m,n)}$ upon noting that (141) is satisfied since (164) implies

$$\|f\|_{L_\infty(\mathcal{A})} + \operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) \leq C \leq N \tag{169}$$

and (154) in combination with (164) yields

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu) \leq \frac{(m+1)(\operatorname{diam}(\mathcal{A}) \operatorname{Lip}(f) + 1)}{N} \leq C/N \leq \varepsilon. \tag{170}$$

□

Qualitatively, Corollary 4 states that the rate at which $b(\varepsilon)$ in (165) tends to ∞ as $\varepsilon \to 0$ equals the rectifiability parameter, i.e., we have

$$\inf\{s \in (0, \infty) : \lim_{\varepsilon \to 0} \varepsilon^s b(\varepsilon) < \infty\} = m. \tag{171}$$

4.2 Generation of Countably m-rectifiable Measures

The result for countably m-rectifiable measures, stated next, follows from combining Lemmata 5 and 9 and Theorem 3. The main idea is to approximate countably m-rectifiable measures by m-rectifiable measures that are supported on a finite number of components of the underlying countably m-rectifiable sets.

Theorem 4. Let $m, n, \ell, N \in \mathbb{N}$ and set

$$\mathcal{F}_N = \{k/N : k \in \mathbb{Z}\} \cap [-N, N]. \tag{172}$$

Then, we can construct a collection $\mathscr{K}_N^{(m,n)} \subseteq \mathcal{N}_{m,n}$ of ReLU neural networks with

$$\left|\mathscr{K}_N^{(m,n)}\right| = (2N^2 + 1)^{n(N+1)^m} \tag{173}$$

with the property that for every sequence $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ of compact sets $\mathcal{A}_i \subseteq \mathbb{R}^m$, every sequence $\{f_i\}_{i\in\mathbb{N}}$ of Lipschitz mappings $f_i\colon \mathcal{A}_i \to \mathbb{R}^n$ satisfying

$$\max_{k=1,\dots,\ell} \|f_k\|_{L_\infty(\mathcal{A}_k)} + \lambda_\ell \leq N, \tag{174}$$

where we set

$$\lambda_\ell = 2\ell \max\big\{\operatorname{diam}(\mathcal{E}), \operatorname{diam}(\mathcal{A}_1)\operatorname{Lip}(f_1), \dots, \operatorname{diam}(\mathcal{A}_\ell)\operatorname{Lip}(f_\ell)\big\}, \tag{175}$$

and every measure ν that is countably m-rectifiable subordinary to $\mathcal{E} = \bigcup_{i\in\mathbb{N}} f_i(\mathcal{A}_i)$, there is a $\Phi \in \mathscr{K}_N^{(m,n)}$, a compact set $\mathcal{B} \subseteq [0,1]^m$, and a Radon measure μ on $\mathcal{B}$ such that $\Phi|_{\mathcal{B}}\#\mu$ is m-rectifiable subordinary to $\Phi(\mathcal{B})$ and satisfies

$$\operatorname{Lip}(\Phi|_{[0,1]^m}) \leq m(\lambda_\ell + 1) \tag{176}$$

and

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu_\ell) \leq \frac{\nu(\mathcal{E})(\lambda_\ell + 1/2)}{N} + (1 - \nu(\mathcal{E}_\ell))\operatorname{diam}(\mathcal{E}). \tag{177}$$

Moreover, every $\Phi \in \mathscr{K}_N^{(m,n)}$ has the same architecture and satisfies

$$\mathcal{L}(\Phi) \leq \lceil \log(m+1) \rceil + 4 \tag{178}$$
$$\mathcal{M}(\Phi) \leq n(N+1)^m(62m - 28) \tag{179}$$
$$\mathcal{W}(\Phi) \leq n(N+1)^m 6m \tag{180}$$
$$\mathcal{K}(\Phi) \subseteq \mathcal{F}_N \tag{181}$$
$$\mathcal{B}(\Phi) \leq N. \tag{182}$$

Proof. Fix a sequence $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ of compact sets $\mathcal{A}_i \subseteq \mathbb{R}^m$, a sequence $\{f_i\}_{i\in\mathbb{N}}$ of Lipschitz mappings $f_i\colon \mathcal{A}_i \to \mathbb{R}^n$ satisfying (174), and a measure ν that is countably m-rectifiable subordinary to $\bigcup_{i\in\mathbb{N}} f_i(\mathcal{A}_i)$ arbitrarily and set $\mathcal{E}_\ell = \bigcup_{i=1}^{\ell} f_i(\mathcal{A}_i)$ and $\nu_\ell = \nu|\mathcal{E}_\ell/\nu(\mathcal{E}_\ell)$. Now, ν_ℓ is m-rectifiable subordinary to $\mathcal{E}_\ell$ and satisfies

$$W_1(\nu, \nu_\ell) \leq (1 - \nu(\mathcal{E}_\ell))\operatorname{diam}(\mathcal{E}) \tag{183}$$

owing to Lemma 9. By Lemma 5, there exist a compact set $\mathcal{B} \subseteq [0,1]^m$ and a Lipschitz mapping $g\colon \mathcal{B} \to \mathbb{R}^n$ with

$$\|g\|_{L_\infty(\mathcal{B})} \leq \max_{k=1,\dots,\ell} \|f_k\|_{L_\infty(\mathcal{A}_k)} \tag{184}$$

and $\operatorname{Lip}(g) \leq \lambda_\ell$ satisfying $\mathcal{E}_\ell = g(\mathcal{B})$. Now, Theorem 3 implies that there exists a ReLU neural network $\Phi \in \mathscr{K}_N^{(m,n)}$ and a Radon measure μ on $\mathcal{B}$ satisfying

$$\operatorname{Lip}(\Phi|_{[0,1]^m}) \leq m(\operatorname{diam}(\mathcal{B})\operatorname{Lip}(g) + 1) \leq m(\lambda_\ell + 1) \tag{185}$$

and

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu) \leq \frac{\nu(\mathcal{E})(\operatorname{diam}(\mathcal{B})\operatorname{Lip}(g) + 1/2)}{N} \leq \frac{\nu(\mathcal{E})(\lambda_\ell + 1/2)}{N}. \tag{186}$$

Combining (183) with (186) yields (177). Finally, Theorem 3 implies that every $\Phi \in \mathscr{K}_N^{(m,n)}$ obeys (178)–(182). ☐

We can combine Theorem 4 with Corollary 8 to obtain the following space-filling result for countably m-rectifiable probability measures.

Corollary 5. Let $m, n, \ell, N \in \mathbb{N}$ and set

$$\mathcal{D}_{3N} = \{a/b : a \in \mathbb{Z}, b \in \mathbb{N}, |a| \leq 4(3N)^{m+1}, \text{ and } b \leq 4(3N)^{m+2}\}. \tag{187}$$

Then, we can construct a collection $\mathscr{J}_N^{(m,n)} \subseteq \mathcal{N}_{1,n}$ of ReLU neural networks with

$$\log\left(\left|\mathscr{J}_N^{(m,n)}\right|\right) \leq 3n(3N)^m \log(6N) \tag{188}$$

with the property that for every sequence $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ of compact sets $\mathcal{A}_i \subseteq \mathbb{R}^m$, every sequence $\{f_i\}_{i\in\mathbb{N}}$ of Lipschitz mappings $f_i \colon \mathcal{A}_i \to \mathbb{R}^n$ satisfying

$$\max_{k=1,\dots,\ell} \|f_k\|_{L_\infty(\mathcal{A}_k)} + \lambda_\ell \leq N, \tag{189}$$

where we set

$$\lambda_\ell = 2\ell \max\left\{ \operatorname{diam}(\mathcal{E}), \operatorname{diam}(\mathcal{A}_1)\operatorname{Lip}(f_1), \dots, \operatorname{diam}(\mathcal{A}_\ell)\operatorname{Lip}(f_\ell)\right\}, \tag{190}$$

and every probability measure ν that is countably m-rectifiable subordinary to $\mathcal{E} = \bigcup_{i\in\mathbb{N}} f_i(\mathcal{A}_i)$, there is a $\Psi \in \mathscr{J}_N^{(m,n)}$ satisfying

$$\operatorname{Lip}(\Psi|_{[0,1]}) \leq m(6(N+1))^{m+1} \tag{191}$$

and

$$W_1\left(\nu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\right) \leq \frac{(m+1)(\lambda_\ell+1)}{N} + (1-\nu(\mathcal{E}_\ell))\operatorname{diam}(\mathcal{E}). \tag{192}$$

Moreover, every $\Psi \in \mathscr{J}_N^{(m,n)}$ has the same architecture and satisfies

$$\mathcal{L}(\Psi) \leq \lceil \log(m+1) \rceil + 6(m-1) + 7 \tag{193}$$
$$\mathcal{M}(\Psi) \leq 78mn(3N)^m \tag{194}$$
$$\mathcal{W}(\Psi) \leq 6mn(3N)^m \tag{195}$$
$$\mathcal{K}(\Psi) \subseteq \mathcal{D}_{3N} \tag{196}$$
$$\mathcal{B}(\Psi) \leq 4(3N)^{m+1}. \tag{197}$$

Proof. Consider the collections $\mathscr{K}_N^{(m,n)}$ and $\mathscr{G}_{3N}^{(m)}$ from Theorem 3 and Corollary 8 (with $d = m$ and $K = 3N$), respectively and set

$$\mathscr{J}_N^{(m,n)} = \{\Phi \circ \rho \circ \Sigma : \Phi \in \mathscr{K}_N^{m,n} \text{ and } \Sigma \in \mathscr{G}_{3N}^{(m)}\}. \tag{198}$$

Then,(144)–(153) implies (188) and (193)–(197). Now, fix a sequence $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ of compact sets $\mathcal{A}_i \subseteq \mathbb{R}^m$, a sequence $\{f_i\}_{i\in\mathbb{N}}$ of Lipschitz mappings $f_i \colon \mathcal{A}_i \to \mathbb{R}^n$ satisfying (189), and a measure ν that is countably m-rectifiable subordinary to $\bigcup_{i\in\mathbb{N}} f_i(\mathcal{A}_i)$ arbitrarily and set $\mathcal{E}_\ell = \bigcup_{i=1}^{\ell} f_i(\mathcal{A}_i)$ and $\nu_\ell = \nu|\mathcal{E}_\ell/\nu(\mathcal{E}_\ell)$. By Theorem 4, there is a $\Phi \in \mathscr{K}_N^{(m,n)}$, a compact set $\mathcal{B} \subseteq [0,1]^m$, and a Radon measure μ on $\mathcal{B}$ such that $\Phi|_{\mathcal{B}}\#\mu$ is m-rectifiable subordinary to $\Phi(\mathcal{B})$ and satisfies

$$\operatorname{Lip}(\Phi|_{[0,1]^m}) \leq m(\lambda_\ell + 1) \tag{199}$$

and

$$W_1(\nu, \Phi|_{\mathcal{B}}\#\mu) \leq \frac{(\lambda_\ell + 1/2)}{N} + (1 - \nu(\mathcal{E}_\ell)) \operatorname{diam}(\mathcal{E}). \tag{200}$$

Moreover, by Corollary 8, there is a $\Sigma \in \mathscr{G}_{3N}^{(m)}$ satisfying

$$\operatorname{Lip}(\Sigma) \leq 6^{m+1} N^m \tag{201}$$

and

$$W_1\Big(\mu, \Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \leq 1/N. \tag{202}$$

We set $\Psi = \Phi \circ \rho \circ \Sigma$ and can therefore conclude that

$$\operatorname{Lip}(\Psi|_{[0,1]}) \leq m(6(N+1))^{m+1} \tag{203}$$

and

$$W_1\Big(\nu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \tag{204}$$

$$\leq W_1\Big(\nu, \Phi|_{\mathcal{B}}\#\mu\Big) + W_1\Big(\Phi|_{\mathcal{B}}\#\mu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \tag{205}$$

$$\leq \frac{\lambda_\ell + 1/2}{N} + (1 - \nu(\mathcal{E}_\ell)) \operatorname{diam}(\mathcal{E}) + \operatorname{Lip}(\Phi|_{[0,1]^m}) W_1\Big(\mu, \Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \tag{206}$$

$$\leq \frac{\lambda_\ell + 1/2}{N} + (1 - \nu(\mathcal{E}_\ell)) \operatorname{diam}(\mathcal{E}) + \frac{m(\lambda_\ell + 1)}{N} \tag{207}$$

$$\leq \frac{(m+1)(\lambda_\ell + 1)}{N} + (1 - \nu(\mathcal{E}_\ell)) \operatorname{diam}(\mathcal{E}), \tag{208}$$

where (206) follows from (200) and Lemma 27 and in (207) we applied (199) and (202). □

We next introduce classes of countably m-rectifiable measures, which will allow us to derive universality result on the generation of the measures in the individual classes.

Definition 8. Let $m, n \in \mathbb{N}$ and $C \in (0, \infty)$. Further, suppose that $\kappa\colon (0,1] \to \mathbb{N}$ is monotonically decreasing. The class $\mathscr{G}(n, m, C, \kappa)$ consists of all probability measures ν on $\mathbb{R}^n$ with the property that there exist a sequence $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ of compact sets $\mathcal{A}_i \subseteq \mathbb{R}^m$, a sequence $\{f_i\}_{i\in\mathbb{N}}$ of Lipschitz mappings $f_i\colon \mathcal{A}_i \to \mathbb{R}^n$ such that

(i) ν decays according to

$$\operatorname{diam}(\mathcal{E})\left(1 - \nu\Big(\bigcup_{k=1}^{\kappa(\varepsilon)} f_k(\mathcal{A}_k)\Big)\right) \leq \varepsilon/2 \quad \text{for all } \varepsilon \in (0,1]; \tag{209}$$

(ii) the mappings are uniformly upper-bounded by

$$\sup_{\varepsilon\in(0,1]} \max_{k=1,\ldots,\kappa(\varepsilon)} \Big\{\|f_k\|_{L_\infty(\mathcal{A}_k)} \varepsilon/\kappa(\varepsilon)\Big\} \leq C; \tag{210}$$

(iii) the Lipschitz constants and the diameters of the sets are upper-bounded as

$$2\max\Big\{\operatorname{diam}(\mathcal{E}), \sup_{k\in\mathbb{N}} \operatorname{diam}(\mathcal{A}_k)\operatorname{Lip}(f_k)\Big\} \leq C; \tag{211}$$

(iv) and ν is countably m-rectifiable subordinary to $\mathcal{E} := \bigcup_{k\in\mathbb{N}} f_k(\mathcal{A}_k)$.

We have the following interpretations of the individual properties on the countably m-rectifiable measure ν in Definition 8. Condition (i) specifies how fast the measure ν decays on the individual m-rectifiable components of the countably m-rectifiable set. A lower decay of the measure allows for a larger increase of the function values in magnitude on the individual components in (ii). Finally, (iii) is a uniform upper bound on the individual Lipschitz constants and the diameters of the sets. The next result states an universality result on the approximation of countably m-rectifiable probability measures.

Corollary 6. Let $m, n \in \mathbb{N}$ and $C \in (0,\infty)$. Further, suppose that $\kappa\colon (0,1] \to \mathbb{N}$ is monotonically decreasing and consider the class $\mathscr{G}(n,m,C,\kappa)$ as defined in Definition 8. Then, for every $\varepsilon \in (0,1]$, there exists a collection $\mathscr{R}(n,m,C,\kappa,\varepsilon)$ of ReLU neural networks of cardinality

$$|\mathscr{R}(n,m,C,\kappa,\varepsilon)| \leq 2^{b(\varepsilon)} \tag{212}$$

with

$$b(\varepsilon) = 3n(3\lceil 2(m+1)(C+1)\kappa(\varepsilon)/\varepsilon\rceil)^m \log(6\lceil 2(m+1)(C+1)\kappa(\varepsilon)/\varepsilon\rceil) \tag{213}$$

such that for every $\nu \in \mathscr{G}(n,m,C,\kappa)$, there is a $\Psi \in \mathscr{R}(n,m,C,\kappa,\varepsilon)$ satisfying

$$\operatorname{Lip}(\Psi|_{[0,1]}) \leq m(6(N+1))^{m+1} \tag{214}$$

and

$$W_1\Big(\nu, \Psi\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \leq \varepsilon. \tag{215}$$

Moreover, every $\Psi \in \mathscr{R}(n,m,C,\kappa,\varepsilon)$ has the same architecture and satisfies (193)–(197) with $N = \lceil C\kappa(\varepsilon)/\varepsilon\rceil$.

Proof. Follows from Corollary 5 with $N = \lceil 2(m+1)(C+1)\kappa(\varepsilon)/\varepsilon\rceil$ and $\mathscr{R}(n,m,C,\kappa,\varepsilon) = \mathscr{J}_N^{(m,n)}$ upon noting that

$$N \geq \max\Big\{2(m+1)(\lambda_{\kappa(\varepsilon)}+1)/\varepsilon, \max_{k=1,\dots,\kappa(\varepsilon)} \|f_k\|_{L_\infty(\mathcal{A}_k)} + \lambda_{\kappa(\varepsilon)}\Big\}. \tag{216}$$

□

Concretely, we get the following examples for the qualitative behaviour of $b(\varepsilon)$ in (213):

(i) If $\kappa(\varepsilon) = \lceil \log(1/\varepsilon)\rceil \Rightarrow b(\varepsilon) = \mathcal{O}\Big(\lceil 1/\varepsilon\rceil^m \log_2^{m+1}(\lceil 1/\varepsilon\rceil)\Big)$;

(ii) if $\kappa(\varepsilon) = \lceil 1/\varepsilon^k\rceil \Rightarrow b(\varepsilon) = \mathcal{O}\Big(\lceil 1/\varepsilon\rceil^{m(k+1)} \log(\lceil 1/\varepsilon\rceil)\Big)$.

Specifically, by (i), if the measure decays exponentially on the individual m-rectifiable components of the countably m-rectifiable set, then the rate at which $b(\varepsilon)$ tends to ∞ as $\varepsilon \to 0$ in (213) is

$$\inf\{s \in (0,\infty) : \lim_{\varepsilon \to 0} \varepsilon^s b(\varepsilon) < \infty\} = m. \tag{217}$$

We therefore have the same qualitative behaviour as for m-rectifiable measures in (171). This is no longer the case if the measure decays polynomially on the individual m-rectifiable components of the countably m-rectifiable set as is demonstrated in (ii).

A Approximations by Uniform Mixtures and Space-Filling Results

This section contains approximation properties of uniform mixtures and quantitative statements on the space-filling approximation of uniform mixtures, which improve and simplify the results in [27]. We start with the definition of (quantized) uniform mixtures.

Definition 9. Let $\mathcal{I} = [0,1]^d$ and $K \in \mathbb{N}$. For every $\ell \in \{1, \ldots, K\}$, set

$$\mathcal{I}_\ell = \begin{cases} [(\ell-1)/K, \ell/K) & \text{if } \ell < K \\ [(\ell-1)/K, \ell/K] & \text{if } \ell = K. \end{cases} \tag{218}$$

Further, let

$$\mathcal{A} = \{k \in \mathbb{N}^d : k_i \leq K \text{ for } i = 1, \ldots, d\} \tag{219}$$

and, for every $k \in \mathcal{A}$, set $\mathcal{J}_k = \mathcal{I}_{k_1} \times \mathcal{I}_{k_2} \times \cdots \times \mathcal{I}_{k_d}$ and designate $\lambda_k = \mathscr{L}^{(d)}|_{\mathcal{J}_k}$. We call a Borel measure κ on $\mathcal{I}$ uniform mixture of resolution K if

$$\kappa = K^d \sum_{k \in \mathcal{A}} w_k \lambda_k \tag{220}$$

with $w_k \in [0,1]$ for all $k \in \mathcal{K}$ and $\sum_{k \in \mathcal{A}} w_k = 1$. For $N \in \mathbb{N}$, the uniform mixture of resolution K in (220) is called $(1/N)$-quanitzed if $w_k \in \delta\mathbb{N}$ for all $k \in \mathcal{A}$.

We next state an approximation property of uniform mixtures in terms of $(1/N)$-quanitzed uniform mixtures.

Lemma 10. Let $\mathcal{I} = [0,1]^d$ be equipped with the metric $\rho(x,y) = \|x-y\|_\infty$. Furhter, let $\kappa = K^d \sum_{k \in \mathcal{A}} w_k \lambda_k$ be a uniform mixture of resolution K on $\mathcal{I} = [0,1]^d$ and let $N \in \mathbb{N}$ with $N \geq K^d$. Then, we can construct a $(1/N)$-quanitzed uniform mixture $\hat{\kappa}$ of resolution K on $\mathcal{I}$ satisfying

$$W_1(\hat{\kappa}, \kappa) \leq 4(K^d - 1)/N. \tag{221}$$

Proof. Lemma 29 implies that we can find a $(\hat{w}_k)_{\mathcal{A}} \in (\mathbb{N}/N)^{|\mathcal{A}|}$ satisfying $\sum_{k \in \mathcal{A}} \hat{w}_k = 1$ and

$$\sum_{k \in \mathcal{A}} |w_k - \hat{w}_k| \leq 4(K^d - 1)/N. \tag{222}$$

Set

$$\hat{\kappa} = K^d \sum_{k\in\mathcal{A}} \hat{w}_k \lambda_k \tag{223}$$

and note that

$$W_1(\hat{\kappa},\kappa) = \sup_{\substack{\psi\colon \mathcal{I}\to\mathbb{R} \\ \mathrm{Lip}(\psi)\le 1}} \left\{ \int \psi \,\mathrm{d}\hat{\kappa} - \int \psi \,\mathrm{d}\kappa \right\} \tag{224}$$

owing to Lemma 25. Since

$$\int (\psi - \psi(0)) \,\mathrm{d}\hat{\kappa} - \int (\psi - \psi(0)) \,\mathrm{d}\kappa = \int \psi \,\mathrm{d}\hat{\kappa} - \int \psi \,\mathrm{d}\kappa, \tag{225}$$

we can restrict, without loss of generality, the supremum in (224) to mappings ψ satisfying $\psi(0) = 0$ and $\mathrm{Lip}(\psi) \le 1$. Now, fix $\psi\colon \mathbb{R}^d \to \mathbb{R}$ satisfying $\psi(0) = 0$ and $\mathrm{Lip}(\psi) \le 1$ arbitrarily. Then, we have

$$\left| \int \psi \,\mathrm{d}\hat{\kappa} - \int \psi \,\mathrm{d}\kappa \right| \le K^d \sum_{k\in\mathcal{A}} |\hat{w}_k - w_k| \int_{\mathcal{J}_k} |\psi| \,\mathrm{d}\lambda_k \tag{226}$$

$$\le \sum_{k\in\mathcal{A}} |\hat{w}_k - w_k| \tag{227}$$

$$\le 4(K^d - 1)/N, \tag{228}$$

where in (227) we used $\big|\psi(x)\big| = \big|\psi(x) - \psi(0)\big| \le \mathrm{Lip}(\psi)\|x\|_\infty \le 1$ for all $x \in \mathcal{I}$ and (228) follows from (222). Using (226)–(228) and the fact that ψ was assumed to be arbitrary in (224), we conclude that (221) holds. □

We next approximate arbitrary Borel probability measures μ by uniform mixtures.

Lemma 11. Let $\mathcal{I} = [0,1]^d$ be equipped with the metric $\rho(x,y) = \|x - y\|_\infty$ and suppose that μ is a Borel probability measure on the metric space $(\mathcal{I},\rho)$ with $\rho(x,y) = \|x-y\|_\infty$. Further, let $K \in \mathbb{N}$ and define the uniform mixture $\tilde{\mu}$ of resolution K according to

$$\tilde{\mu} = K^d \sum_{k\in\mathcal{A}} w_k \lambda_k \tag{229}$$

with $w_k = \mu(\mathcal{J}_k)$ for all $k \in \mathcal{A}$. Then, we have

$$W_1(\mu,\tilde{\mu}) \le 1/K. \tag{230}$$

Proof. Set $\mathcal{B} = \{k \in \mathcal{A} : w_k > 0\}$. Then, we have

$$W_1(\mu,\tilde{\mu}) \le \sum_{k\in\mathcal{B}} w_k W_1\big(\mu|_{\mathcal{J}_k}/w_k, \tilde{\mu}|_{\mathcal{J}_k}/w_k\big) \tag{231}$$

$$\le \sum_{k\in\mathcal{B}} w_k \int \rho \,\mathrm{d}((\mu|_{\mathcal{J}_k}/w_k) \times (\tilde{\mu}|_{\mathcal{J}_k}/w_k)) \tag{232}$$

$$\le 1/K \sum_{k\in\mathcal{B}} w_k \tag{233}$$

$$= 1/K, \tag{234}$$

where (231) follows from Lemma 28. □

We are now in a position to state the an approximation result by uniform mixtures, which improves [27, Theorem VII.2].

Theorem 5. Let $\mathcal{I} = [0,1]^d$ be equipped with the metric $\rho(x,y) = \|x-y\|_\infty$ and suppose that μ is a Borel probability measure on the metric space $(\mathcal{I},\rho)$. Further, let $K \in \mathbb{N}$ and suppose that $N \in \mathbb{N}$ with $N \geq K^d$. Then, we can construct a $(1/N)$-quanitzed uniform mixture $\tilde{\mu}$ of resolution K satisfying

$$W_1(\mu,\tilde{\mu}) \leq 4(K^d-1)/N + 1/K. \tag{235}$$

Proof. Follows from Lemmata 10 and 11. □

It is worth noting that Theorem 5 yields the following metric entropy upper bound on the set of Borel probability measure on $(\mathcal{I},\rho)$.

Corollary 7. Let $\mathscr{F}$ denote the space of Borel probability measures on $(\mathcal{I},\rho)$ and consider the metric space $(\mathscr{F}, W_1)$. Then, we have

$$H_\varepsilon(\mathscr{F}, W_1) \leq (\varepsilon/2)^d \log(24/\varepsilon). \tag{236}$$

Proof. Set

$$\mathcal{A}_d = \{k \in \mathbb{N}^d : \|k\|_\infty \leq K\} \tag{237}$$

and note that the set of $1/(4K^{d+1})$-quatized weights $\{w_k\}_{k\in\mathcal{A}_d}$ satisfies

$$|\{w_k\}_{k\in\mathcal{A}_d}| \leq \binom{4K^{d+1}-1}{K^d-1} \tag{238}$$

$$\leq \binom{4K^{d+1}}{K^d} \tag{239}$$

$$\leq (12K)^{K^d}, \tag{240}$$

where (240) follows from $\binom{n}{k} \leq (en/k)^k$. Now, Theorem 5 with $N = 4K^{d+1}$ implies that for every Borel probability measure μ, there is a $1/(4K^{d+1})$-quatized uniform mixture $\tilde{\mu}$ of resolution K satisfying

$$W_1(\mu,\tilde{\mu}) \leq 2/K. \tag{241}$$

combining (238)–(238) with (241) yields (236). □

We also note the following relation to quanitzation [11]. For every $n \in \mathbb{N}$, let $\mathcal{F}_n$ denote the set of all Borel measurable mappings $f\colon [0,1]^d \to [0,1]^d$ satisfying $\left|f([0,1]^d)\right| \leq n$. For a Borel probability measure μ on $[0,1]^d$, the n-th quantization error of order one can be defined according to [11, Equation (3.1) and Lemma 3.4]

$$V_{n,1}(\mu) = \inf_{f\in\mathcal{F}_n} \widetilde{W}_1(\mu, f\#\mu) \tag{242}$$

with Wasserstein distance

$$\widetilde{W}_1(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int \tilde{\rho}\,\mathrm{d}\pi \tag{243}$$

and $\tilde{\rho}(x,y) = \|x-y\|_1$. The Wasserstein distances $\widetilde{W}_1$ in (243) and W_1 as defined in (486) satisfy $W_1 \le \widetilde{W}_1 \le dW_1$ owing to $\|\cdot\|_\infty \le \|\cdot\|_1 \le d\|\cdot\|_\infty$. The upper quantization dimension of order one of a Borel probability measure μ is defined as

$$\overline{D}_1(\mu) = \limsup_{n\in\mathbb{N}} \frac{\log(n)}{\log(1/V_{n,1}(\mu))}. \tag{244}$$

and Theorem 5 with $N = 4K(K^d-1)$ therefore yields the upper bound

$$\overline{D}_1(\mu) \le d, \tag{245}$$

which is tight provided that μ is regular of dimension d [11, Definition 12.1 and Theorem 12.18].

We next establish that $(1/N)$-quantized uniform mixtures can be approximated with arbitrarily small error as push-forwards of $\mathscr{L}^{(1)}|_{[0,1]}$ using space-filling curves. What is more, the space-filling curves can be realized as ReLU neural networks. A result in this direction was already established in [27, Theorem V.6], but we here present a quantitative version of this result and a much more direct and structured proof.

Theorem 6. Let $K \in \mathbb{N}\setminus\{1\}$, set $\mathcal{I} = [0,1]^d$, and consider the following setup:

1) For every $k \in \{1,\dots,K\}$, set

$$\mathcal{I}_k = \begin{cases} [(k-1)/K, k/K) & \text{if } k < K \\ [(k-1)/K, k/K] & \text{if } k = K. \end{cases} \tag{246}$$

Further, for $j = 1,\dots,d$, set

$$\mathcal{A}_j = \{k \in \mathbb{N}^j : \|k\|_\infty \le K\} \tag{247}$$

and, for every $k \in \mathcal{A}_d$, set $\mathcal{J}_k = \mathcal{I}_{k_1} \times \mathcal{I}_{k_2} \times \cdots \times \mathcal{I}_{k_d}$ and designate $\lambda_k = \mathscr{L}^{(d)}|_{\mathcal{J}_k}$. Finally, let $N \in \mathbb{N}$ with $N \ge \min\{K^d, 4\}$ and let μ be an arbitrary $(1/N)$-quantized uniform mixture of resolution K defined according to

$$\mu = K^d \sum_{k\in\mathcal{A}_d} w_k \lambda_k \tag{248}$$

with $w_k \in \mathbb{N}/N$ for all $k \in \mathcal{A}_d$ and $\sum_{k\in\mathcal{A}_d} w_k = 1$.

2) For every $j \in \{1,\dots,d-1\}$ and $(k_1,\dots,k_j) \in \mathcal{A}_j$, define the "marginal weights" as

$$w_{(k_1,\dots,k_j)} = \sum_{k_{j+1}=1}^{K} \sum_{k_{j+2}=1}^{K} \cdots \sum_{k_d=1}^{K} w_k. \tag{249}$$

For every $j \in \{2,\dots,d\}$ and $(k_1,\dots,k_j) \in \mathcal{A}_j$, define the "conditional weights" according to

$$w_{k_j|(k_1\dots k_{j-1})} = \frac{w_{(k_1,\dots,k_j)}}{w_{(k_1,\dots,k_{j-1})}}. \tag{250}$$

3) Set $b_0 = 0$ and, for $k_1 = 1, \dots, K$, let $b_{k_1} = \sum_{i=1}^{k_1} w_{(i)}$ and designate the sets

$$\mathcal{K}_{k_1} = \begin{cases} [b_{k_1-1}, b_{k_1}) & \text{if } k_1 < K \\ [b_{k_1-1}, b_{k_1}] & \text{if } k_1 = K. \end{cases} \tag{251}$$

For every $j \in \{2, \dots, d\}$ and $(k_1, \dots, k_j) \in \mathcal{A}_j$, set $b_{0|(k_1,\dots,k_{j-1})} = 0$,

$$b_{k_j|(k_1,\dots,k_{j-1})} = \sum_{i=1}^{k_j} w_{i|(k_1\dots k_{j-1})}, \tag{252}$$

and designate the sets

$$\mathcal{K}_{k_j|(k_1\dots k_{j-1})} = \begin{cases} [b_{k_j-1|(k_1\dots k_{j-1})}, b_{k_j|(k_1\dots k_{j-1})}) & \text{if } k_j < K \\ [b_{k_j-1|(k_1\dots k_{j-1})}, b_{k_j|(k_1\dots k_{j-1})}] & \text{if } k_j = K. \end{cases} \tag{253}$$

4) Let $\mathcal{K}_k = \mathcal{K}_{k_1} \times \mathcal{K}_{k_2|k_1} \times \cdots \times \mathcal{K}_{k_d|(k_{d-1},\dots,k_1)}$ and define the mapping $f\colon \mathcal{I} \to \mathcal{I}$ according to $f = \sum_{k\in\mathcal{A}} f^{(k)} \chi_{\mathcal{K}_k}$ with

$$f^{(k)}(x) = \begin{pmatrix} f_1^{(k_1)}(x_1) \\ f_2^{(k_1,k_2)}(x_2) \\ \vdots \\ f_d^{(k_1,k_2,\dots,k_d)}(x_d) \end{pmatrix} = \begin{pmatrix} \frac{x_1 - b_{k_1}}{Kw_{(k_1)}} + \frac{k_1}{K} \\ \frac{x_2 - b_{k_2|k_1}}{Kw_{k_2|k_1}} + \frac{k_2}{K} \\ \vdots \\ \frac{x_d - b_{k_d|(k_1,\dots,k_{d-1})}}{Kw_{k_d|(k_1,\dots,k_{d-1})}} + \frac{k_d}{K} \end{pmatrix} \tag{254}$$

for all $x \in \mathcal{K}_k$ and $k \in \mathcal{A}_d$.

5) Fix $s \in \mathbb{N}$ and define $\tilde{f}^{(s)}\colon \mathbb{R} \to \mathbb{R}^d$ according to $\tilde{f}^{(s)} = (\tilde{f}_1, \tilde{f}_2^{(s)}, \dots, \tilde{f}_d^{(s)})^\mathsf{T}$ with

$$\tilde{f}_1(x) = x \quad \text{for all } x \in (-\infty, 0), \tag{255}$$

$$\tilde{f}_1(x) = \sum_{k_1 \in \mathcal{A}_1} f_1^{(k_1)} \chi_{\mathcal{K}_{k_1}}(x) \quad \text{for all } x \in [0, 1], \tag{256}$$

and

$$\tilde{f}_1(x) = 1/(Kw_{(K)})(x-1) + 1 \quad \text{for all } x \in (1, \infty), \tag{257}$$

and

$$\tilde{f}_j^{(s)} = \sum_{(k_1,\dots,k_{j-1})\in\mathcal{A}_{j-1}} \left(\hat{f}_j^{(k_1,\dots,k_{j-1})} \circ g_s \circ L_{k_{j-1}}\right) \tag{258}$$

$$\circ \left(\hat{f}_{j-1}^{(k_1,\dots,k_{j-2})} \circ g_s \circ L_{k_{j-2}}\right) \tag{259}$$

$$\vdots \tag{260}$$

$$\circ \left(\hat{f}_2^{(k_1)} \circ g_s \circ L_{k_1}\right) \circ \tilde{f}_1, \tag{261}$$

where g_s is as defined in Definition 22, $L_\ell(x) = Kx - \ell + 1$ for all $x \in \mathbb{R}$ and $\ell \in \mathbb{N}$, and, for every $j = 2, \ldots, d$ and $(k_1, \ldots, k_{j-1}) \in \mathcal{A}_{j-1}$, we set

$$\hat{f}_j^{(k_1,\ldots,k_{j-1})}(x) = x \quad \text{for all } x \in (-\infty, 0), \tag{262}$$

$$\hat{f}_j^{(k_1,\ldots,k_{j-1})}(x) = \sum_{k_j=1}^{K} f_j^{(k_1,\ldots,k_j)} \chi_{\mathcal{K}_{k_j|(k_1,\ldots,k_{j-1})}}(x) \quad \text{for all } x \in [0,1], \tag{263}$$

and

$$\hat{f}_j^{(k_1,\ldots,k_{j-1})}(x) = 1/(Kw_{k_j|(k_1,\ldots,k_{j-1})})(x-1) + 1 \quad \text{for all } x \in (1, \infty). \tag{264}$$

Then, the following properties hold.

(i) The mapping f in 4) is a bijection with $f(\mathcal{K}_k) = \mathcal{J}_k$ for all $k \in \mathcal{A}_d$ satisfying $\mu = f\#\mathscr{L}^{(d)}|_{\mathcal{I}}$.

(ii) For every $j = 1 \ldots, d$, define the cubes $\mathcal{C}_{r_j}^{k_j}$ for $k_j = 1, \ldots, K$ and $r_j = 1, \ldots, 2^{s-1}$ as follows. Set

$$\mathcal{C}_{r_j}^{k_j} = \Big[\frac{k_j - 1}{K} + \frac{r_j - 1}{K2^{s-1}},\ \frac{k_j - 1}{K} + \frac{r_j}{K2^{s-1}}\Big) \tag{265}$$

if $k_j < K$ or $k_j = K$ and $r_j < 2^{s-1}$ and

$$\mathcal{C}_{2^{s-1}}^{K} = \Big[1 - \frac{1}{K2^{s-1}},\ 1\Big]. \tag{266}$$

Further, set

$$\mathcal{B}_{r_1}^{k_1} = \tilde{f}_1^{-1}(\mathcal{C}_{r_1}^{k_1}) \quad \text{for } k_1 = 1, \ldots, K \text{ and } r_1 = 1, \ldots, 2^{s-1} \tag{267}$$

and define inductively

$$\mathcal{B}_{r_1,\ldots,r_j}^{k_1,\ldots,k_j} = \tilde{f}_j^{(s)^{-1}}(\mathcal{C}_{r_j}^{k_j}) \cap \mathcal{B}_{r_1,\ldots,r_{j-1}}^{k_1,\ldots,k_{j-1}} \tag{268}$$

for all $(k_1, \ldots, k_j) \in \mathcal{A}_j$, $r_1, \ldots, r_j \in \{1, \ldots, 2^{s-1}\}$, and $j = 2, \ldots, d$.. Then, we have

$$(\tilde{f}^{(s)}\#(\mathscr{L}^{(1)}|_{[0,1]}))\big(\mathcal{C}_{r_1}^{k_1} \times \mathcal{C}_{r_2}^{k_2} \cdots \times \mathcal{C}_{r_d}^{k_d}\big) = \mu\big(\mathcal{C}_{r_1}^{k_1} \times \mathcal{C}_{r_2}^{k_2} \cdots \times \mathcal{C}_{r_d}^{k_d}\big) \tag{269}$$

for all $(k_1, \ldots, k_d) \in \mathcal{A}_d$ and $r_1, r_2, \ldots, r_d \in \{1, \ldots, 2^{s-1}\}$.

(iii) We have

$$W_1(\tilde{f}^{(s)}\#(\mathscr{L}^{(1)}|_{[0,1]}), \mu) \leq \frac{1}{K2^{s-1}} \tag{270}$$

with

$$\mathrm{Lip}(\tilde{f}^{(s)}) \leq \frac{2^{s(d-1)}N}{K}. \tag{271}$$

(iv) We can construct a ReLU neural network $\Phi^{(s)}_{N,K} \in \mathcal{N}_{1,d}$ with

$$\mathcal{L}(\Phi^{(s)}_{N,K}) = 3 + (d-1)(5+s) \tag{272}$$

$$\mathcal{M}(\Phi^{(s)}_{N,K}) \leq 4d(2K+3s+4)K^d/(K-1) \tag{273}$$

$$\mathcal{W}(\Phi^{(s)}_{N,K}) \leq 4K^d \tag{274}$$

$$\mathcal{K}(\Phi^{(s)}_{N,K}) \subseteq \mathcal{D}_{N,K} \tag{275}$$

$$\mathcal{B}(\Phi^{(s)}_{N,K}) \leq N, \tag{276}$$

where we set

$$\mathcal{D}_{N,K} = \{a/b : a \in \mathbb{Z}, b \in \mathbb{N}, |a| \leq N, \text{ and } b \leq NK\}, \tag{277}$$

satisfying

$$\Phi^{(s)}_{N,K} = \tilde{f}^{(s)} \tag{278}$$

and

$$\mathrm{Lip}(\Phi^{(s)}_{N,K}) \leq \frac{2^{s(d-1)}N}{K}. \tag{279}$$

The architecture of $\Phi^{(s)}_{N,K}$ only depend on s, d, and K but not on μ and N.

Proof. We first prove (i). By construction, we have

$$f_1^{(k_1)}(\mathcal{K}_{k_1}) = \mathcal{I}_{k_1} \tag{280}$$

$$f_2^{(k_1,k_2)}(\mathcal{K}_{k_2|k_1}) = \mathcal{I}_{k_2} \tag{281}$$

$$\vdots \tag{282}$$

$$f_d^{(k_1,k_2,\dots,k_d)}(\mathcal{K}_{k_d|(k_{d-1},\dots,k_1)}) = \mathcal{I}_{k_d}, \tag{283}$$

which implies $f(\mathcal{K}_k) = \mathcal{J}_k$ for all $k \in \mathcal{A}_d$. Since $\mathcal{I} = \bigcup_{k\in\mathcal{A}} \mathcal{J}_k$, this implies in turn that f is onto. Now, $f^{(k)}$ is injective on $\mathcal{K}_k$ for all $k \in \mathcal{A}_d$. Moreover, since the sets $\mathcal{J}_k$ are pairwise disjoint and $f(\mathcal{K}_k) = \mathcal{J}_k$ for all $k \in \mathcal{A}$, f is injective on $\bigcup_{k\in\mathcal{A}_d} \mathcal{K}_k$. Finally, since

$$\bigcup_{k_1=1}^{K} \mathcal{K}_{k_1} = [0, b_K] = [0,1] \tag{284}$$

and

$$\bigcup_{k_j=1}^{K} \mathcal{K}_{k_j|(k_{j-1},\dots,k_1)} = [0, b_{K|(k_1,\dots,k_{j-1})}] \tag{285}$$

$$= [0,1] \quad \text{for } j = 2,\dots,d, \tag{286}$$

we conclude that $\bigcup_{k\in\mathcal{A}_d} \mathcal{K}_k = \mathcal{I}$ so that f is injective on $\mathcal{I}$. Summarizing, $f\colon \mathcal{I} \to \mathcal{I}$ is a bijection with $f(\mathcal{K}_k) = \mathcal{J}_k$ for all $k \in \mathcal{A}_d$. Now, fix $k \in \mathcal{A}_d$ arbitrarily and suppose that $\mathcal{B}_k \subseteq \mathcal{J}_k$. Then, we have

$$\mathscr{L}^{(1)}(f_1^{-1}(\mathcal{B}_k)) = K w_{k_1} \mathscr{L}^{(1)}(\mathcal{B}_k \cap \mathcal{I}_{k_1}) \tag{287}$$

and

$$\mathscr{L}^{(1)}(f_j^{-1}(\mathcal{B}_k)) = K w_{k_j|(k_1,\dots,k_{j-1})} \mathscr{L}^{(1)}(\mathcal{B}_k \cap \mathcal{I}_{k_j}) \quad \text{for } j = 2, \dots, d, \tag{288}$$

which yields

$$\mathscr{L}^{(d)}(f^{-1}(\mathcal{B}_k)) = K^d w_{k_1} \Big(\prod_{j=2}^{d} w_{k_j|(k_1,\dots,k_{j-1})} \Big) \lambda_k(\mathcal{B}_k) \tag{289}$$

$$= K^d w_k \lambda_k(\mathcal{B}_k). \tag{290}$$

We thus have

$$\mathscr{L}^{(d)}(f^{-1}(\mathcal{B})) = \sum_{k \in \mathcal{A}_d} \mathscr{L}^{(d)}((f^{-1}(\mathcal{B} \cap \mathcal{J}_k)) \tag{291}$$

$$= \sum_{k \in \mathcal{A}_d} K^d w_k \lambda_k(\mathcal{B}) \tag{292}$$

$$= \mu(\mathcal{B}) \quad \text{for all } \mathcal{B} \subseteq \mathcal{I}, \tag{293}$$

where (292) follows from (289)–(290), which proves $\mu = f\#(\mathscr{L}^{(d)}|_{\mathcal{I}})$. We next prove (ii). Suppose first that $d = 1$. Since $f_1^{k_1}$ has slope $1/(K w_{(k_1)})$, we have

$$\mathscr{L}^{(1)}(\mathcal{B}_{r_1}^{k_1}) = K w_{(k_1)} \mathscr{L}^{(1)}(\mathcal{C}_{r_1}^{k_1}) = w_{(k_1)}/2^{s-1} \tag{294}$$

for $k_1 = 1, \dots, K$ and $r_1 = 1, \dots, 2^{s-1}$, which establishes (269) for $d = 1$. Next, suppose that $d > 1$. We divide the proof into several steps.

Step 1: We prove by induction that, for every $j \in \{2, \dots, d\}$, we have

$$\tilde{f}_j^{(s)}(x) = \Big(\hat{f}_j^{(k_1,\dots,k_{j-1})} \circ g_s \circ L_{k_{j-1}} \Big) \circ \Big(\hat{f}_{j-1}^{(k_1,\dots,k_{j-2})} \circ g_s \circ L_{k_{j-2}} \Big) \tag{295}$$

$$\circ \cdots \circ \Big(\hat{f}_2^{(k_1)} \circ g_s \circ L_{k_1} \Big) \circ \tilde{f}_1(x) \quad \text{for all } x \in \mathcal{B}_{r_1,\dots,r_{j-1}}^{k_1,\dots,k_{j-1}}, \tag{296}$$

$(k_1, \dots, k_{j-1}) \in \mathcal{A}_{j-1}$, and $r_1, \dots, r_{j-1} \in \{1, \dots, 2^{s-1}\}$. Suppose first that $j = 2$ and fix k_1, r_1, and $x \in \mathcal{B}_{r_1}^{k_1}$ arbitrarily. Then, we have

$$\tilde{f}_2^{(s)}(x) = \sum_{(\ell_1) \in \mathcal{A}_1} \hat{f}_2^{(\ell_1)} \circ g_s \circ L_{\ell_1} \circ \tilde{f}_1(x). \tag{297}$$

Since $\mathcal{B}_{r_1}^{k_1} = \tilde{f}_1^{\,-1}(\mathcal{C}_{r_1}^{k_1})$, we have $\tilde{f}_1(x) \in \mathcal{C}_{r_1}^{k_1} \subseteq \mathcal{I}_{k_1}$ so that $(\hat{f}_2^{(\ell_1)} \circ g_s \circ L_{\ell_1} \circ \tilde{f}_1)(x) = 0$ if $\ell_1 \neq k_1$ owing to $g_s(x) = 0$ for all $x \notin (0, 1)$ and $\hat{f}_2^{(\ell_1)}(0) = 0$. We thus have

$$\tilde{f}_2^{(s)}(x) = \hat{f}_2^{(k_1)} \circ g_s \circ L_{k_1} \circ \tilde{f}_1(x) \quad \text{for all } x \in \mathcal{B}_{r_1}^{k_1}. \tag{298}$$

Now, suppose that (295)-(296) holds for $j < d$ and fix $(k_1, \dots, k_j) \in \mathcal{A}_j$, $r_1, \dots, r_j \in \{1, \dots, s^{s-1}\}$, and $x \in \mathcal{B}_{r_1,\dots,r_j}^{k_1,\dots,k_j}$ arbitrarily. Suppose that $(\ell_1, \dots, \ell_{j-1}) \in \mathcal{A}_{j-1}$ with $(\ell_1, \dots, \ell_{j-1}) \neq (k_1, \dots, k_{j-1})$. Then, we have

$$\Big(\hat{f}_{j+1}^{(\ell_1,\dots,\ell_j)} \circ g_s \circ L_{\ell_{j+1}} \Big) \circ \Big(\hat{f}_j^{(\ell_1,\dots,\ell_{j-1})} \circ g_s \circ L_{\ell_{j-1}} \Big) \tag{299}$$

$$\circ \cdots \circ \Big(\hat{f}_2^{(\ell_1)} \circ g_s \circ L_{\ell_1} \Big) \circ \tilde{f}_1^{(s)}(x) \tag{300}$$

$$= \Big(\hat{f}_{j+1}^{(\ell_1,\dots,\ell_j)} \circ g_s \circ L_{\ell_{j+1}} \Big)(0) \tag{301}$$

$$= 0 \quad \text{for } \ell_j = 1, \dots, K, \tag{302}$$

where (301) follows from $\mathcal{B}^{k_1,\dots,k_j}_{r_1,\dots,r_j} \subseteq \mathcal{B}^{k_1,\dots,k_{j-1}}_{r_1,\dots,r_{j-1}}$, the fact that (295)-(296) holds for j, and because all the functions in the sum in (258)–(261) are nonnegative and (302) is by $g_s(x) = 0$ for all $x \notin (0,1)$ and $\hat{f}^{(\ell_1,\dots,\ell_j)}_{j+1}(0) = 0$. We therefore have

$$\tilde{f}^{(s)}_{j+1}(x) = \sum_{\ell_j=1}^{K} \left(\hat{f}^{(k_1,\dots,k_{j-1},\ell_j)}_{j+1} \circ g_s \circ L_{\ell_j}\right) \circ \left(\hat{f}^{(k_1,\dots,k_{j-1})}_{j} \circ g_s \circ L_{k_{j-1}}\right) \tag{303}$$

$$\circ \dots \circ \left(\hat{f}^{(k_1)}_{2} \circ g_s \circ L_{k_1}\right) \circ \tilde{f}^{(s)}_1(x) \tag{304}$$

$$= \sum_{\ell_j=1}^{K} \left(\hat{f}^{(k_1,\dots,k_{j-1},\ell_j)}_{j+1} \circ g_s \circ L_{\ell_j}\right) \circ \tilde{f}^{(s)}_j(x) \tag{305}$$

$$= \left(\hat{f}^{(k_1,\dots,k_j)}_{j+1} \circ g_s \circ L_{k_j}\right) \circ \tilde{f}^{(s)}_j(x), \tag{306}$$

where (305) follows again from $\mathcal{B}^{k_1,\dots,k_j}_{r_1,\dots,r_j} \subseteq \mathcal{B}^{k_1,\dots,k_{j-1}}_{r_1,\dots,r_{j-1}}$ and the fact that (295)-(296) holds for j and in (306) we used $\tilde{f}^{(s)}_j(x) \in \mathcal{C}^{k_j}_{r_j} \subseteq \mathcal{I}_{k_j}$ if $x \in \mathcal{B}^{k_1,\dots,k_j}_{r_1,\dots,r_j}$, $g_s(x) = 0$ for all $x \notin (0,1)$, and $\hat{f}^{(k_1,\dots,k_{j-1},\ell_j)}_{j+1}(0) = 0$ for $\ell_j = 1,\dots,K$. Plugging the expression for $\tilde{f}^{(s)}_j$ from (295)-(296) into (306) establishes (295)-(296) for $j+1$.

Step 2: We prove by induction that, for every $j \in \{2,\dots,d\}$, we have

$$\tilde{f}^{(s)}_j(x) = \left(\hat{f}^{(k_1,\dots,k_{j-1})}_{j} \circ H_{r_{j-1}} \circ L_{k_{j-1}}\right) \circ \left(\hat{f}^{(k_1,\dots,k_{j-2})}_{j-1} \circ H_{r_{j-2}} \circ L_{k_{j-2}}\right) \tag{307}$$

$$\circ \dots \circ \left(\hat{f}^{(k_1)}_{2} \circ H_{r_1} \circ L_{k_1}\right) \circ \tilde{f}_1(x) \quad \text{for all } x \in \mathcal{B}^{k_1,\dots,k_{j-1}}_{r_1,\dots,r_{j-1}}, \tag{308}$$

$(k_1,\dots,k_{j-1}) \in \mathcal{A}_{j-1}$, and $r_1,\dots,r_{j-1} \in \{1,\dots,2^{s-1}\}$, where, for $r = 1,\dots,2^{s-1}$, the continuous piecewise linear function $H_r \colon \mathbb{R} \to \mathbb{R}$ is defined according to

$$H_r(x) = \begin{cases} 2^s x - 2(r-1) & \text{if } x \in (-\infty, (2r-1)/2^s] \\ -2^s x + 2r & \text{if } x \in ((2r-1)/2^s, \infty]. \end{cases} \tag{309}$$

Suppose first that $j = 2$ and fix k_1, r_1, and $x \in \mathcal{B}^{k_1}_{r_1}$ arbitrarily. Then, we have

$$\tilde{f}^{(s)}_2(x) = \hat{f}^{(k_1)}_2 \circ g_s \circ L_{k_1} \circ \tilde{f}_1(x) \tag{310}$$

$$= \sum_{r=1}^{2^{s-1}} \hat{f}^{(k_1)}_2 \circ h_r \circ L_{k_1} \circ \tilde{f}_1(x) \tag{311}$$

$$= \hat{f}^{(k_1)}_2 \circ H_{r_1} \circ L_{k_1} \circ \tilde{f}_1(x), \tag{312}$$

where (310) is by (295)-(296), in (311) we applied Lemma 32, and (312) follows from $\mathcal{B}^{k_1}_{r_1} = \tilde{f}_1^{-1}(\mathcal{C}^{k_1}_{r_1})$,

$$L_{k_1}(\mathcal{C}^{k_1}_{r_1}) \subseteq [(r_1-1)/2^{s-1}, r_1/2^{s-1}], \tag{313}$$

and $H_{r_1}(x) = h_{r_1}(x)$ for all $x \in [(r_1-1)/2^{s-1}, r_1/2^{s-1}]$. Now, suppose that (307)-(308) holds for $j < d$ and fix $(k_1,\dots,k_j) \in \mathcal{A}_j$, $r_1,\dots,r_j \in \{1,\dots,2^{s-1}\}$, and $x \in \mathcal{B}^{k_1,\ldots,k_j}_{r_1,\ldots,r_j}$ arbitrarily. Then, we have

$$\tilde{f}^{(s)}_{j+1}(x) = \left(\hat{f}^{(k_1,\ldots,k_j)}_{j+1} \circ g_s \circ L_{k_j}\right) \circ \tilde{f}^{(s)}_j(x) \tag{314}$$

$$= \sum_{r=1}^{2^{s-1}} \left(\hat{f}^{(k_1,\ldots,k_j)}_{j+1} \circ h_r \circ L_{k_j}\right) \circ \tilde{f}^{(s)}_j(x) \tag{315}$$

$$= \hat{f}^{(k_1,\ldots,k_j)}_{j+1} \circ H_{r_j} \circ L_{k_j} \circ \tilde{f}^{(s)}_j(x), \tag{316}$$

where (314) is by (303)–(306), in (315) we applied Lemma 32, and (316) follows from $\mathcal{B}^{k_1,\ldots,k_j}_{r_1,\ldots,r_j} \subseteq \tilde{f}^{(s)}_j{}^{-1}(\mathcal{C}^{k_j}_{r_j})$,

$$L_{k_j}(\mathcal{C}^{k_j}_{r_j}) \subseteq [(r_j-1)/2^{s-1}, r_j/2^{s-1}], \tag{317}$$

and $H_{r_j}(x) = h_{r_j}(x)$ for all $x \in [(r_j-1)/2^{s-1}, r_j/2^{s-1}]$. Plugging the expression for $\tilde{f}^{(s)}_j$ from (307)–(308) into (316) establishes (307)–(308) for $j+1$.

Step 3: We prove by induction that, for every $j \in \{2,\ldots,d\}$, we have

$$\Bigg(\left(\hat{f}^{(k_1,\ldots,k_{j-1})}_j \circ H_{r_{j-1}} \circ L_{k_{j-1}}\right) \circ \left(\hat{f}^{(k_1,\ldots,k_{j-2})}_{j-1} \circ H_{r_{j-2}} \circ L_{k_{j-2}}\right) \tag{318}$$

$$\circ \cdots \circ \left(\hat{f}^{(k_1)}_2 \circ H_{r_1} \circ L_{k_1}\right) \circ \tilde{f}_1\Bigg)^{-1}(\mathcal{A}) \tag{319}$$

$$= \Bigg(\left(f^{(k_1,\ldots,k_j)}_j \circ H_{r_{j-1}} \circ L_{k_{j-1}}\right) \circ \left(f^{(k_1,\ldots,k_{j-1})}_{j-1} \circ H_{r_{j-2}} \circ L_{k_{j-2}}\right) \tag{320}$$

$$\circ \cdots \circ \left(f^{(k_1,k_2)}_2 \circ H_{r_1} \circ L_{k_1}\right) \circ f^{k_1}_1\Bigg)^{-1}(\mathcal{A}) \tag{321}$$

for all $(k_1,\ldots,k_j) \in \mathcal{A}_j$, $r_1,\ldots,r_j \in \{1,\ldots,2^{s-1}\}$, and $\mathcal{A} \subseteq \overline{\mathcal{I}}_{k_j}$, with H_r as defined in (309) for $r = 1,\ldots,2^{s-1}$. Suppose that $j = 2$ and fix $(k_1,k_2) \in \mathcal{A}_2$, $r_1, r_2 \in \{1,\ldots,2^{s-1}\}$, and $\mathcal{A} \subseteq \overline{\mathcal{I}}_{k_2}$ arbitrarily. Then, (i) implies

$$\hat{f}^{(k_1)}_2{}^{-1}(\mathcal{A}) = f^{(k_1,k_2)}_2{}^{-1}(\mathcal{A}) \subseteq [0,1]. \tag{322}$$

Moreover, we have

$$L^{-1}_{k_1}(H^{-1}_{r_1}([0,1])) \subseteq \overline{\mathcal{C}}^{k_1}_{r_1} \subseteq \overline{\mathcal{I}}_{k_1} \tag{323}$$

so that (i) yields

$$\left(\hat{f}^{(k_1)}_2 \circ H_{r_1} \circ L_{k_1} \circ \tilde{f}_1\right)^{-1}(\mathcal{A}) = \left(f^{(k_1,k_2)}_2 \circ H_{r_1} \circ L_{k_1} \circ f^{k_1}_1\right)^{-1}(\mathcal{A}), \tag{324}$$

which establishes (318)–(321) for $j = 2$. Next, suppose that (318)–(321) holds for $j < d$ and fix $(k_1,\ldots,k_{j+1}) \in \mathcal{A}_{j+1}$, $r_1,\ldots,r_{j+1} \in \{1,\ldots,2^{s-1}\}$, and $\mathcal{A} \subseteq \overline{\mathcal{I}}_{k_{j+1}}$ arbitrarily. Then, (i) implies

$$\hat{f}^{(k_1,\ldots,k_j)}_{j+1}{}^{-1}(\mathcal{A}) = f^{(k_1,\ldots,k_{j+1})}_{j+1}{}^{-1}(\mathcal{A}) \subseteq [0,1]. \tag{325}$$

Moreover, we have

$$L_{k_j}^{-1}(H_{r_j}^{-1}([0,1])) = \overline{\mathcal{C}}_{r_j}^{k_j} \subseteq \overline{\mathcal{I}}_{k_j}. \tag{326}$$

Now, (325) and (326) imply

$$L_{k_j}^{-1}(H_{r_j}^{-1}(\hat{f}_{j+1}^{(k_1,\dots,k_j)^{-1}}(\mathcal{A}))) \subseteq \overline{\mathcal{I}}_{k_j}. \tag{327}$$

Finally, combining (325), (327), and (318)–(321) establishes (318)–(321) for $j+1$.

Step 4: Combining the results from Step 2 and Step 3, we obtain

$$\tilde{f}_j^{(s)}(x) = \left(f_j^{(k_1,\dots,k_j)} \circ H_{r_{j-1}} \circ L_{k_{j-1}}\right) \circ \left(f_{j-1}^{(k_1,\dots,k_{j-1})} \circ H_{r_{j-2}} \circ L_{k_{j-2}}\right) \tag{328}$$

$$\circ \cdots \circ \left(f_2^{(k_1,k_2)} \circ H_{r_1} \circ L_{k_1}\right) \circ f_1^{(k_1)}(x) \quad \text{for all } x \in \mathcal{B}_{r_1,\dots,r_j}^{k_1,\dots,k_j}, \tag{329}$$

$(k_1,\dots,k_j) \in \mathcal{A}_j$, and $r_1,\dots,r_j \in \{1,\dots,2^{s-1}\}$, and $j = 2,\dots,d$, with H_r as defined in (309) for $r = 1,\dots,2^{s-1}$.

Step 5: We establish

$$\mathscr{L}^{(1)}(\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d}) \leq \frac{w_{k_1,\dots,k_d}}{2^{d(s-1)}} \tag{330}$$

for all $(k_1,\dots,k_d) \in \mathcal{A}_d$ and $r_1,\dots,r_d \in \{1,\dots,2^{s-1}\}$. To this end, fix $(k_1,\dots,k_d) \in \mathcal{A}_d$ and $r_1,\dots,r_d \in \{1,\dots,2^{s-1}\}$ arbitrarily and note that (328)–(329) for $j = d$ implies

$$(J\tilde{f}_d^{(s)})|_{\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d}} = \frac{2^{s(d-1)}}{Kw_{(k_1,\dots,k_d)}} \tag{331}$$

and, since $\left|H_r^{-1}(x)\right| \leq 2$ for all $x \in \mathbb{R}$ and $r = 1,\dots,2^{s-1}$,

$$\left|\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d} \cap \left\{\tilde{f}_d^{(s)^{-1}}(x)\right\}\right| \leq 2^{d-1} \quad \text{for all } x \in [0,\infty). \tag{332}$$

We therefore have

$$\mathscr{L}^{(1)}(\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d}) = \frac{Kw_{(k_1,\dots,k_d)}}{2^{s(d-1)}} \int_{\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d}} (J\tilde{f}_d^{(s)})\, \mathrm{d}\mathscr{L}^{(1)} \tag{333}$$

$$= \frac{Kw_{(k_1,\dots,k_d)}}{2^{s(d-1)}} \int_{\mathcal{C}_{r_d}^{k_d}} \left|\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d} \cap \left\{\tilde{f}_d^{(s)^{-1}}(x)\right\}\right| \mathrm{d}\mathscr{L}^{(1)} \tag{334}$$

$$\leq \frac{Kw_{(k_1,\dots,k_d)}}{2^{(s-1)(d-1)}} \mathscr{L}^{(1)}(\mathcal{C}_{r_d}^{k_d}) \tag{335}$$

$$= \frac{w_{k_1,\dots,k_d}}{2^{d(s-1)}}, \tag{336}$$

where (333) follows from (331), in (334) we applied the area formula Theorem 8, $\tilde{f}_d^{(s)}(\mathcal{B}_{r_1,\dots,r_d}^{k_1,\dots,k_d}) \subseteq \mathcal{C}_{r_d}^{k_d}$, and (iv) in Lemma 21, and (335) is by (332), which proves (330).

Step 6: We prove by induction that, for every $j \in \{2, \ldots, d\}$, we have

$$\tilde{f}_j^{(s)}([0,1]) \subseteq [0,1] \tag{337}$$

and

$$\bigcup_{k_1,\ldots,k_j=1}^{K} \bigcup_{r_1,\ldots,r_j=1}^{2^{s-1}} \mathcal{B}_{r_1,\ldots,r_j}^{k_1,\ldots,k_j} = [0,1]. \tag{338}$$

Suppose that $j = 2$. Then, we have

$$\begin{aligned}
\bigcup_{k_1=1}^{K} \bigcup_{r_1=1}^{2^{s-1}} \mathcal{B}_{r_1}^{k_1} &= \bigcup_{k_1=1}^{K} \bigcup_{r_1=1}^{2^{s-1}} \tilde{f_1}^{-1}(\mathcal{C}_{r_1}^{k_1}) &(339)\\
&= \tilde{f_1}^{-1}\left(\bigcup_{k_1=1}^{K} \bigcup_{r_1=1}^{2^{s-1}} \mathcal{C}_{r_1}^{k_1}\right) &(340)\\
&= \tilde{f_1}^{-1}([0,1]) &(341)\\
&= [0,1], &(342)
\end{aligned}$$

where in (342) we used the fact that

$$\tilde{f}_1(x) \in [0,1] \quad \text{for all } x \in [0,1], \tag{343}$$

which follows from (256) and $f_1^{(k_1)}(x) \in [0,1]$ for all $x \in \mathcal{K}_{k_1}$ and $k_1 = 1, \ldots, K$. Now, pick $x \in [0,1]$ arbitrarily. Then, (339)–(342) implies that there exists a k_1 and an r_1 such that $x \in \mathcal{B}_{r_1}^{k_1}$. We therefore have

$$\begin{aligned}
\tilde{f}_2^{(s)}(x) &= \hat{f}_2^{(k_1)} \circ H_{r_1} \circ L_{k_1} \circ \tilde{f}_1(x) &(344)\\
&\in \hat{f}_2^{(k_1)} \circ H_{r_1} \circ L_{k_1}(\mathcal{C}_{r_1}^{k_1}) &(345)\\
&\subseteq \hat{f}_2^{(k_1)}([0,1]) &(346)\\
&= [0,1], &(347)
\end{aligned}$$

where (344) follows from (307)–(308) and in (346) we used (313), which proves (337) for $j = 2$. Next, note that

$$\begin{aligned}
\bigcup_{k_1,k_2=1}^{K} \bigcup_{r_1,r_2=1}^{2^{s-1}} \mathcal{B}_{r_1,r_2}^{k_1,k_2} &= \bigcup_{k_1,k_2=1}^{K} \bigcup_{r_1,r_2=1}^{2^{s-1}} \left(\mathcal{B}_{r_1}^{k_1} \cap \tilde{f}_2^{(s)^{-1}}(\mathcal{C}_{r_2}^{k_2})\right) &(348)\\
&= \bigcup_{k_1=1}^{K} \bigcup_{r_1=1}^{2^{s-1}} \mathcal{B}_{r_1}^{k_1} \cap \left(\bigcup_{k_2=1}^{K} \bigcup_{r_2=1}^{2^{s-1}} \tilde{f}_2^{(s)^{-1}}(\mathcal{C}_{r_2}^{k_2})\right) &(349)\\
&= [0,1] \cap \tilde{f}_2^{(s)^{-1}}\left(\bigcup_{k_2=1}^{K} \bigcup_{r_2=1}^{2^{s-1}} \mathcal{C}_{r_2}^{k_2}\right) &(350)\\
&= [0,1] \cap \tilde{f}_2^{(s)^{-1}}([0,1]) &(351)\\
&= [0,1], &(352)
\end{aligned}$$

where (350) is by (339)–(342) and (352) follows from (337) for $j = 2$, which establishes (338) for $j = 2$. Now, suppose that (337) and (338) hold for $j < d$ and pick $x \in [0,1]$ arbitrarily. Then, there must exist a $(k_1, \dots, k_j) \in \mathcal{A}_j$ and $r_1, \dots, r_j \in \{1, \dots, 2^{s-1}\}$ such that $x \in \mathcal{B}^{k_1,\dots,k_j}_{r_1,\dots,r_j}$. We therefore have

$$\begin{aligned}\tilde{f}^{(s)}_{j+1}(x) &= \hat{f}^{(k_1,\dots,k_j)}_{j+1} \circ H_{r_j} \circ L_{k_j} \circ \tilde{f}^{(s)}_j(x) && (353)\\ &\in \hat{f}^{(k_1)}_{j+1} \circ H_{r_j} \circ L_{k_j}(\mathcal{C}^{k_j}_{r_j}) && (354)\\ &\subseteq \hat{f}^{(k_1)}_{j+1}([0,1]) && (355)\\ &= [0,1], && (356)\end{aligned}$$

where (353) follows from (314)–(316), which proves (337) for $j+1$. Finally, note that

$$\begin{aligned}&\bigcup_{k_1,\dots,k_{j+1}=1}^{K} \bigcup_{r_1,\dots,r_{j+1}=1}^{2^{s-1}} \mathcal{B}^{k_1,\dots,k_{j+1}}_{r_1,\dots,r_{j+1}} && (357)\\ &= \bigcup_{k_1,\dots,k_{j+1}=1}^{K} \bigcup_{r_1,\dots,r_{j+1}=1}^{2^{s-1}} \Big(\mathcal{B}^{k_1,\dots,k_j}_{r_1,\dots,r_j} \cap ((\tilde{f}^{(s)}_{j+1})^{-1}(\mathcal{C}^{k_{j+1}}_{r_{j+1}}))\Big) && (358)\\ &= \bigcup_{k_1,\dots,k_j=1}^{K} \bigcup_{r_1,\dots,r_j=1}^{2^{s-1}} \mathcal{B}^{k_1,k_2,\dots,k_j}_{r_1,r_2,\dots,r_j} \cap \Big(\bigcup_{k_{j+1}=1}^{K} \bigcup_{r_{j+1}=1}^{2^{s-1}} (\tilde{f}^{(s)}_{j+1})^{-1}(\mathcal{C}^{k_{j+1}}_{r_{j+1}})\Big) && (359)\\ &= [0,1] \cap (\tilde{f}^{(s)}_{j+1})^{-1}\Big(\bigcup_{k_{j+1}=1}^{K} \bigcup_{r_{j+1}=1}^{2^{s-1}} \mathcal{C}^{k_{j+1}}_{r_{j+1}}\Big) && (360)\\ &= [0,1] \cap (\tilde{f}^{(s)}_{j+1})^{-1}([0,1]) && (361)\\ &= [0,1], && (362)\end{aligned}$$

where (360) is by (338) for j and (362) follows from (337) for $j+1$, which establishes (338) for $j+1$.

Step 7: We prove (iii). To establish (269), fix $(k_1, \dots, k_d) \in \mathcal{A}_d$ and $r_1, r_2, \dots, r_d \in \{1, \dots, 2^{s-1}\}$ arbitrarily. Then, we have

$$\mu\big(\mathcal{C}^{k_1}_{r_1} \times \mathcal{C}^{k_2}_{r_2} \cdots \times \mathcal{C}^{k_d}_{r_d}\big) = \frac{w_{(k_1,\dots,k_d)}}{2^{d(s-1)}}. \tag{363}$$

Combining (330) with (338) yields

$$(\tilde{f}^{(s)}_d \# \mathscr{L}^{(1)}|_{[0,1]})\big(\mathcal{C}^{k_1}_{r_1} \times \mathcal{C}^{k_2}_{r_2} \cdots \times \mathcal{C}^{k_d}_{r_d}\big) = \mathscr{L}^{(1)}|_{[0,1]}(\mathcal{B}^{k_1,\dots,k_j}_{r_1,\dots,r_j}) = \frac{w_{k_1,\dots,k_d}}{2^{d(s-1)}} \tag{364}$$

upon noting that $(\tilde{f}^{(s)}_d \# \mathscr{L}^{(1)}|_{[0,1]})(\mathcal{I}) = 1$ and

$$\sum_{k_1,\dots,k_d=1}^{K} \sum_{r_1,\dots,r_d=1}^{2^{s-1}} \frac{w_{(k_1,\dots,k_d)}}{2^{d(s-1)}} = 1. \tag{365}$$

Now, (270) follows from Lemma 11 with $K2^{s-1}$ in place of K. It remains to establish (271). To this end, note that (255)–(257) in combination with

Theorem 7 implies that $\tilde{f}_1$ is a piecewise linear function with

$$\text{Lip}(\tilde{f}_1) \leq \frac{N}{K} \tag{366}$$

and (254) and (328)–(329) in combination with Theorem 7 implies that $\tilde{f}_j^{(s)}$ is a piecewise linear function with

$$\text{Lip}(\tilde{f}_j^{(s)}) \leq \frac{2^{s(d-1)}N}{K} \quad \text{for } j = 2, \dots, d. \tag{367}$$

Combining (366) and (334) proves (271).

It remains to establish (iv). We first realize the individual mappings appearing in the composition of the mappings $\tilde{f}_j$ as ReLU neural networks.

(a) Consider the affine mappings $W_1 \colon \mathbb{R} \to \mathbb{R}^3$, $W_2 \colon \mathbb{R}^3 \to \mathbb{R}^3$, and $W_3 \colon \mathbb{R}^3 \to \mathbb{R}$ defined according to

$$W_1(x) = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix} x - \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \tag{368}$$

$$W_2(x) = \begin{pmatrix} 2 & -2 & 2 \\ 4 & -4 & 4 \\ 2 & -2 & 2 \end{pmatrix} x - \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \tag{369}$$

$$W_3(x) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} x, \tag{370}$$

respectively, and define, for every $s \in \mathbb{N}$, the ReLU neural network $\Phi_s \in \mathcal{N}_{1,1}$ according to

$$\Phi_s = W_3 \circ \rho \circ \underbrace{W_2 \circ \rho \dots W_2 \circ \rho}_{(s-1) \text{ times}} W_1 \tag{371}$$

with $\mathcal{L}(\Phi_s) = s + 1$, $\mathcal{M}(\Phi_s) = 11s - 3$, $\mathcal{K}(\Phi_s) = \{0, 1, 2, 4, -1, -2, -4\} \subseteq \mathcal{D}_{N,K}$, and $\mathcal{W}(\Phi_s) = 3$. Then, we have $\Phi_s = g_s$.

(b) For every $\ell \in \{1, \dots, K\}$, consider the affine mapping $L_\ell \colon \mathbb{R} \to \mathbb{R}$ defined according to $L_\ell(x) = Kx - \ell + 1$. Further, define the affine mappings $V_1 \colon \mathbb{R} \to \mathbb{R}^2$ and $V_\ell \colon \mathbb{R} \to \mathbb{R}$ according to

$$V_1(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} x \tag{372}$$

$$V_2(x) = \begin{pmatrix} K & -K \end{pmatrix} x - \ell + 1 \tag{373}$$

and define, for every $\ell \in \{1, \dots, K\}$, the ReLU neural network $\Theta_\ell \in \mathcal{N}_{1,1}$ according to

$$\Theta_\ell = V_2 \circ \rho \circ V_1 \tag{374}$$

with $\mathcal{L}(\Theta_\ell) = 2$, $\mathcal{M}(\Theta_\ell) = 5$, $\mathcal{K}(\Theta_\ell) = \{-K, \dots, K\} \subseteq \mathcal{D}_{N,K}$, and $\mathcal{W}(\Theta_\ell) = 2$. Then, we have $\Theta_\ell = L_\ell$.

(c) Consider the mapping $\tilde{f}_1$ as defined in (255)–(257). Further, consider the affine mappings $U_1\colon \mathbb{R} \to \mathbb{R}^{2K-1}$ and $U_2\colon \mathbb{R}^{2K-1} \to \mathbb{R}$ defined according to

$$U_1(x) = \begin{pmatrix} -x \\ x \\ x - b_1 \\ x - b_1 \\ \vdots \\ x - b_{K-1} \\ x - b_{K-1} \end{pmatrix} \tag{375}$$

and

$$U_2(x) = \begin{pmatrix} -x \\ x/(Kw_1) \\ x/(Kw_2) \\ -x/(Kw_1) \\ \dots \\ x/(Kw_K) \\ -x/(Kw_{K-1}) \end{pmatrix}^{\mathsf{T}}, \tag{376}$$

respectively and define the ReLU neural network $\Psi_1 \in \mathcal{N}_{1,1}$ according to

$$\Psi_1 = U_2 \circ \rho \circ U_1 \tag{377}$$

with $\mathcal{L}(\Psi_1) = 2$, $\mathcal{M}(\Psi_1) = 4K-1$, and $\mathcal{W}(\Psi_1) = 2K$, and $\mathcal{K}(\Psi_1) \subseteq \mathcal{D}_{N,K}$. Then, Lemma 30 implies $\Psi_1 = \tilde{f}_1$.

(d) Fix $(k_1, \dots, k_{j-1}) \in \mathcal{A}_{j-1}$ and $j = 2, \dots, d$ arbitrarily, consider the mapping $\hat{f}_j^{(k_1,\dots,k_{j-1})}$ as defined in (262)–(264). Further, consider the affine mappings $U_1^{(k_1,\dots,k_{j-1})}\colon \mathbb{R} \to \mathbb{R}^{2K-1}$ and $U_2^{(k_1,\dots,k_{j-1})}\colon \mathbb{R}^{2K-1} \to \mathbb{R}$ defined according to

$$U_1^{(k_1,\dots,k_{j-1})}(x) = \begin{pmatrix} -x \\ x \\ x - b_{1|(k_1,\dots,k_{j-1})} \\ x - b_{1|(k_1,\dots,k_{j-1})} \\ \vdots \\ x - b_{K-1|(k_1,\dots,k_{j-1})} \\ x - b_{K-1|(k_1,\dots,k_{j-1})} \end{pmatrix} \tag{378}$$

and

$$U_2^{(k_1,\dots,k_{j-1})}(x) = \begin{pmatrix} -x \\ x/(Kw_{1|(k_1,\dots,k_{j-1})}) \\ x/(Kw_{2|(k_1,\dots,k_{j-1})}) \\ -x/(Kw_{1|(k_1,\dots,k_{j-1})}) \\ \vdots \\ x/(Kw_{K|(k_1,\dots,k_{j-1})}) \\ -x/(Kw_{K-1|(k_1,\dots,k_{j-1})}) \end{pmatrix}^{\mathsf{T}}, \tag{379}$$

respectively and define the ReLU neural network $\Psi_1^{(k_1,\dots,k_{j-1})} \in \mathcal{N}_{1,1}$ according to

$$\Psi_j^{(k_1,\dots,k_{j-1})} = U_2^{(k_1,\dots,k_{j-1})} \circ \rho \circ U_1^{(k_1,\dots,k_{j-1})} \tag{380}$$

with $\mathcal{L}(\Psi_j^{(k_1,\dots,k_{j-1})}) = 2$, $\mathcal{M}(\Psi_j^{(k_1,\dots,k_{j-1})}) = 4K-1$, $\mathcal{W}(\Psi_j^{(k_1,\dots,k_{j-1})}) = 2K$, and $\mathcal{K}(\Psi_j^{(k_1,\dots,k_{j-1})}) \subseteq \mathcal{D}_{N,K}$. Then, Lemma 30 implies $\Psi_j^{(k_1,\dots,k_{j-1})} = \hat{f}_j^{(k_1,\dots,k_{j-1})}$.

(e) Fix $j = 2,\dots,d$ arbitrarily and consider the mapping $\tilde{f}_j^{(s)}$ as defined in (258)–(261). Combining the ReLU neural networks in (a)–(d) by adding a void ReLU activation function ρ between all the individual ReLU networks, we can construct, for every $(k_1,\dots,k_{j_1}) \in \mathcal{A}_{j-1}$, a ReLU neural network $\Psi_{(k_1,\dots,k_{j-1})}^{(s)}$ with[1]

$$\mathcal{L}(\Psi_{(k_1,\dots,k_{j-1})}^{(s)}) = 2 + (d-1)(5+s) \tag{381}$$

$$\mathcal{M}(\Psi_{(k_1,\dots,k_{j-1})}^{(s)}) = (d-j)(5+s) + (j-1)(4K+11s+1) + 4K - 1 \tag{382}$$

$$\mathcal{W}(\Psi_{(k_1,\dots,k_{j-1})}^{(s)}) = 2K \tag{383}$$

$$\mathcal{K}(\Psi_{(k_1,\dots,k_{j-1})}^{(s)}) \subseteq \mathcal{D}_{N,K} \tag{384}$$

such that

$$\tilde{f}_j^{(s)} = \sum_{(k_1,\dots,k_{j-1})\in\mathcal{A}_{j-1}} \Psi_{(k_1,\dots,k_{j-1})}^{(s)}. \tag{385}$$

(f) Fix $j = 2,\dots,d$ arbitrarily and consider the mapping $\tilde{f}_j^{(s)}$ as defined in (258)–(261). Combining Lemma 2 with (e) and realizing the sum in (385) as inner product with the all-ones vector of length K^{j-1}, we can construct a ReLU neural network $\Psi_j^{(s)}$ with

$$\mathcal{L}(\Psi_j^{(s)}) = 3 + (j-1)(5+s) \tag{386}$$

$$\mathcal{M}(\Psi_j^{(s)}) = \Big((d-j)(5+s) + (j-1)(4K+11s+1) + 4K\Big)K^{j-1} \tag{387}$$

$$\mathcal{W}(\Psi_j^{(s)}) = 2K^j \tag{388}$$

$$\mathcal{K}(\Psi_j^{(s)}) \subseteq \mathcal{D}_{N,K} \tag{389}$$

such that $\tilde{f}_j^{(s)} = \Psi_j^{(s)}$.

(g) Consider the ReLU neural networks $\Psi_j^{(s)}$ constructed in (f). Then, the ReLU neural network $\Phi_s = P(\Psi_1^{(s)},\dots,\Psi_d^{(s)})$ has the desired properties

[1] By adding void components of the form $1 \circ \rho \dots 1 \circ \rho$ of lengths $(d-j)(5+s)$, we can make all networks to have depth $2 + (d-1)(5+s)$.

in (iv) upon noting that

$$\sum_{j=1}^{d}\Big((d-j)(5+s)+(j-1)(4K+11s+1)+4K\Big)K^{j-1} \tag{390}$$
$$\leq 4d(3s+2K+4)K^d/(K-1) \tag{391}$$

and

$$2\sum_{j=1}^{d}K^j = 2K(K^d-1)/(K-1) \tag{392}$$
$$\leq 4K^d. \tag{393}$$

Finally, (279) follows from (271) and (278).

□

We next combine Theorem 5 and Theorem 6 to obtain space-filling approximations of arbitrary Borel measures on $[0,1]^d$.

Corollary 8. Let $\mathcal{I}=[0,1]^d$ be equipped with the metric $\rho(x,y)=\|x-y\|_\infty$. Further, let $K\in\mathbb{N}\setminus\{1\}$ and set

$$\mathcal{D}_K=\{a/b : a\in\mathbb{Z}, b\in\mathbb{N}, |a|\leq 4K^{d+1} \text{ and } b\leq 4K^{d+2}\}. \tag{394}$$

Then, we can construct a collection $\mathscr{G}_K^{(d)}\subseteq\mathcal{N}_{1,d}$ of ReLU neural networks with $\big|\mathscr{G}_K^{(d)}\big|\leq(12K)^{K^d}$ with the property that for every Borel probability measure μ on $(\mathcal{I},\rho)$, there is a $\Sigma\in\mathscr{G}_K^{(d)}$ satisfying

$$W_1\Big(\mu,\Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big)\leq 3/K. \tag{395}$$

Moreover, every $\Sigma\in\mathscr{G}_K^{(d)}$ satisfies

$$\mathcal{L}(\Sigma)=3+6(d-1) \tag{396}$$
$$\mathcal{M}(\Sigma)\leq 44dK^d \tag{397}$$
$$\mathcal{W}(\Sigma)\leq 4K^d \tag{398}$$
$$\mathcal{B}(\Sigma)=4K^{d+1} \tag{399}$$
$$\mathcal{K}(\Sigma)\subseteq\mathcal{D}_K. \tag{400}$$

and

$$\mathrm{Lip}(\Sigma)\leq 2^{d+1}K^d. \tag{401}$$

Proof. Fix an arbitrary Borel probability measure μ on $(\mathcal{I},\rho)$ and let $\tilde{\mu}$ be as in Theorem 5 with $N=4K^{d+1}$. Further, consider the ReLU neural network $\Sigma:=\Phi^{(1)}_{4K^{d+1},K}$ from (iv) in Theorem 6 when applied to $\tilde{\mu}$ with $N=4K^{d+1}$ and $s=1$. Then, we have

$$W_1\Big(\mu,\Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big)\leq W_1(\mu,\tilde{\mu})+W_1\Big(\tilde{\mu},\Sigma\#(\mathcal{L}^{(1)}|_{[0,1]})\Big) \tag{402}$$
$$\leq (K^d-1)/K^{d+1}+2/K \tag{403}$$
$$\leq 3/K, \tag{404}$$

where (403) follows from Theorem 5 and (iii)–(iv) in Theorem 6. Moreover, Σ satisfyies (395)–(400) owing to (iv) in Theorem 6 upon noting that $(2K+7)/(K-1) \leq 11$. Finally, set

$$\mathcal{A}_d = \{k \in \mathbb{N}^d : \|k\|_\infty \leq K\} \tag{405}$$

and note that Σ is completely determined by the set of $1/N$-quatized weights $\{w_k\}_{k\in\mathcal{A}_d}$ of $\tilde{\mu}$. But

$$|\{w_k\}_{k\in\mathcal{A}_d}| \leq \binom{N-1}{K^d-1} \tag{406}$$

$$\leq \binom{N}{K^d} \tag{407}$$

$$= \binom{4K^{d+1}}{K^d} \tag{408}$$

$$\leq (12K)^{K^d}, \tag{409}$$

where (409) follows from $\binom{n}{k} \leq (en/k)^k$, which establishes $\left|\mathscr{G}_K^{(d)}\right| \leq (12K)^{K^d}$. □

Corollary 8 implies that, for fixed d and $\varepsilon \in (0,1]$, the number of ReLU neural networks required to approximate an arbitrary μ on $\mathcal{I}$ with error no larger than ε through push-forward of $\mathcal{L}^{(1)}|_{[0,1]}$ is upper-bounded by

$$(12\lceil 3/\varepsilon \rceil)^{\lceil 3/\varepsilon \rceil^d}. \tag{410}$$

A Properties of Lipschitz Mappings

In this section, we state basic definitions and properties of Lipschitz mappings required in the current paper. We start with the definition of Lipschitz constants, gradients, and generalized Jacobian determinants of Lipschitz mappings.

Definition 10. Let $\mathcal{A} \subseteq \mathbb{R}^m$ and suppose that $f\colon \mathcal{A} \to \mathbb{R}^n$ is Lipschitz. For every $p \in [1,\infty]$, we set

$$\mathrm{Lip}^{(p)}(f) = \inf\left\{ s \in [0,\infty) : \sup_{\substack{x,y\in\mathcal{A}\\ x\neq y}} \frac{\|f(x)-f(y)\|_\infty}{\|x-y\|_p} \leq s \right\} \tag{411}$$

and designate $\mathrm{Lip}(f) := \mathrm{Lip}^{(\infty)}(f)$. We define the gradient of f according to $\nabla f = (\partial_{x_1} f, \partial_{x_2} f, \ldots, \partial_{x_m} f)$ at each point where ∇f exists and write

$$Jf = \begin{cases} \sqrt{\det((\nabla f)^\mathsf{T} \nabla f))} & \text{if } m \leq n \\ \sqrt{\det((\nabla f)\nabla f^\mathsf{T})} & \text{if } n > m. \end{cases} \tag{412}$$

for the generalized Jacobian determinant of f.

Rademacher's theorem [18, Theorem 5.1.11] ensures that the ∇f and Jf in Definition 10 are defined up to a set of Lebesgue measure zero and Lebesgue-measurable. The following result states that on convex open domains, the set of Lipschitz functions equals the set of local Sobolev functions.

Theorem 7. Let $\mathcal{O} \subseteq \mathbb{R}^m$ be open and convex and consider an arbitrary function $f\colon \mathcal{O} \to \mathbb{R}$. Then, the following properties hold.

(i) f is Lipschitz if and only if $f \in W^{1,\infty}(\mathcal{O}')$ for all open and bounded sets $\mathcal{O}' \subseteq \mathcal{O}$;

(ii) If f is Lipschitz, then

$$\mathrm{Lip}^{(1)}(f) = \| \|\nabla f\|_\infty \|_{L_\infty(\mathcal{O})}. \tag{413}$$

Proof. The proof follows along the same lines as the proof of [17, Theorem 3.31]. Suppose that f is Lipschitz. Then, ∇f exists up to a set of measure zero thanks to Rademacher's theorem [18, Theorem 5.1.11]. Moreover, we have

$$\begin{aligned} \left|e_i^\mathsf{T} \nabla f(x)\right| &= \lim_{t\to 0} \frac{\left|f(x+te_i) - f(x)\right|}{t} && (414)\\ &\le \mathrm{Lip}^{(1)}(f) \quad \text{for almost all } x \in \mathcal{O} \text{ and } i = 1,\dots,m, && (415) \end{aligned}$$

which implies $\| \|\nabla f\|_\infty \|_{L_\infty(\mathbb{R}^m)} \le \mathrm{Lip}^{(1)}(f)$ and, in turn, that $f \in W^{1,\infty}(\mathcal{O}')$ for all open and bounded sets $\mathcal{O}' \subseteq \mathcal{O}$.

Now, fix an open, bounded, and convex set $\mathcal{O}' \subseteq \mathcal{O}$ arbitrarily and suppose that $f \in W^{1,\infty}(\mathcal{O}')$. Further, fix $p \in (m, \infty)$ arbitrarily. Since $\mathcal{O}'$ is bounded, we have $f \in W^{1,p}(\mathcal{O}')$ thanks to [1, Theorem 2.14] so that [1, Part II of Theorem 4.12] implies that f has a representative that is uniformly continuous on $\mathcal{O}'$. Let $\hat{f}$ be the zero extension of that representative outside $\mathcal{O}'$. Now, set

$$\eta(x) = \begin{cases} ce^{-1/(1-\|x\|_2)} & \text{if } \|x\|_2 < 1 \\ 0 & \text{else,} \end{cases} \tag{416}$$

where $c \in (0,\infty)$ is chosen so that $\int \eta \,\mathrm{d}\mathscr{L}^m = 1$. Further, for every $\varepsilon \in (0,\infty)$, set $\eta_\varepsilon = \eta(\,\cdot\,/\varepsilon)/\varepsilon^m$, $\mathcal{O}'_\varepsilon = \{x \in \mathcal{O}' : \sup_{y\in\partial\mathcal{O}} \|x-y\|_2 > 2\varepsilon\}$, and $f_\varepsilon = \hat{f} * \eta_\varepsilon$, where $*$ denotes convolution. Now, [1, Item (e) of Theorem 2.29] implies that f_ε converges uniformly to f on $\mathcal{O}'$. Next, note that

$$\begin{aligned} \|\nabla f_\varepsilon\|_\infty &= \max_{i=1,\dots,m} |\partial_{x_i} f_\varepsilon| && (417)\\ &= \max_{i=1,\dots,m} |(D_{x_i} f) * \eta_\varepsilon| && (418)\\ &\le \Big(\max_{i=1,\dots,m} |D_{x_i} f| \Big) * \eta_\varepsilon && (419)\\ &= \|Df\|_\infty * \eta_\varepsilon, && (420) \end{aligned}$$

where (418) is by [17, Theorem 1.19] with $D_{x_i} f$ denoting the weak derivative of f and in (420) we set $Df = (D_{x_1} f, \dots, D_{x_m} f)^\mathsf{T}$. We therefore have

$$\begin{aligned} \| \|\nabla f_\varepsilon\|_\infty \|_{L_p(\mathcal{O}'_\varepsilon)} &\le \| \|Df\|_\infty \|_{L_p(\mathcal{O}'_\varepsilon)} \|\eta_\varepsilon\|_{L_1(\mathcal{O}'_\varepsilon)} && (421)\\ &\le \| \|Df\|_\infty \|_{L_p(\mathcal{O})} \quad \text{for all } \varepsilon \in (0,\infty), && (422) \end{aligned}$$

where in (421) we applied Young's inequality for convolutions [1, Corollary 2.25] in combination with (417)–(420). By the arbitrariness of p, (421)–(422) together with [1, Theorem 2.14] yields

$$\| \|\nabla f_\varepsilon\|_\infty \|_{L_\infty(\mathcal{O}'_\varepsilon)} \le \| \|Df\|_\infty \|_{L_\infty(\mathcal{O})} \quad \text{for all } \varepsilon \in (0,\infty). \tag{423}$$

Now, fix $x, y \in \mathcal{O}$ arbitrarily and pick an open, convex, and bounded set $\mathcal{O}' \subseteq \mathcal{O}$ with $x, y \in \mathcal{O}'$. Futher, let $\varepsilon_0 \in (0, \infty)$ be sufficiently small so that $x, y \in \mathcal{O}'_{\varepsilon_0}$. Then, we have

$$|f_\varepsilon(y) - f_\varepsilon(x)| = \left| \int_0^1 \nabla f_\varepsilon(tx + (1-t)y)^\mathsf{T}(x - y)\,\mathrm{d}t \right| \tag{424}$$

$$\leq \|\|\nabla f_\varepsilon\|_\infty\|_{L_\infty(\mathcal{O}'_\varepsilon)} \|x - y\|_1 \quad \text{for all } \varepsilon \in (0, \varepsilon_0), \tag{425}$$

where (424) follows from the fundamental theorem of calculus [31, Proposition 1.6.41]. Taking the limit $\varepsilon \to 0$ in (424)–(425) and using uniform convergence of $f_\varepsilon \to f$ and (423), it follows that $\big|f(y) - f(x)\big| \leq \|\|Df\|_\infty\|_{L_\infty(\mathcal{O})}\|x - y\|_1$. Since x, y were assumed to be arbitrary, we can conclude that f is Lipschitz with $\mathrm{Lip}^{(1)}(f) \leq \|\|Df\|_\infty\|_{L_\infty(\mathcal{O})}$. Finally, we can replace Df by ∇f thanks to Rademacher's theorem [18, Theorem 5.1.11]. □

We next state an elementary property of Lipschitz functions on convex and closed domains.

Lemma 12. Let $\mathcal{A}, \mathcal{B}_1, \ldots, \mathcal{B}_n \subseteq \mathbb{R}^m$ be convex and closed with $\mathcal{A} = \bigcup_{i=1}^n \mathcal{B}_i$. Further, let $f \colon \mathcal{A} \to \mathbb{R}$ be Lipschitz. Then, we have

$$\mathrm{Lip}^{(1)}(f) \leq \max_{i=1,\ldots,n} \mathrm{Lip}^{(1)}(f|_{\mathcal{B}_i}). \tag{426}$$

Proof. Fix $x, y \in \mathcal{A}$ with $x \neq y$ arbitrarily and set $\mathscr{L} = \{ty + (1-t)x : t \in [0,1]\}$. We next establish that there exist $x_0, \ldots, x_k \in \mathcal{L}$ satisfying the following properties: (i) $x_0 = x$, (ii) $x_k = y$, and (iii) for every $i \in \{1, \ldots, k\}$, there exists a $\mathcal{B}_{j_i}$ such that $x_i, x_{i-1} \in \mathcal{B}_{j_i}$. Set $x_0 = x$. There exists an $\varepsilon_1 \in (0, 1]$ and a $\mathcal{B}_{i_1}$ such that $\{ty + (1-t)x_0 : t \in [0, \varepsilon_1]\} \subseteq \mathcal{B}_{i_1}$ owing to Lemma 13 below. If $y \in \mathcal{B}_{i_1}$, then set $x_1 = y$ and we are done. If $y \notin \mathcal{B}_{i_1}$, then set $t_1 = \max\{t \in [0,1] : ty + (1-t)x_0 \in \mathcal{B}_{i_1}\}$ and $x_1 = t_1 y + (1-t_1)x_0$. Now, there exists an $\varepsilon_2 \in (0, 1]$ and a $\mathcal{B}_{i_2}$ such that $\{ty + (1-t)x_1 : t \in [0, \varepsilon_1]\} \subseteq \mathcal{B}_2$ owing to Lemma 13 below. By construction, $\mathcal{B}_{i_2} \neq \mathcal{B}_{i_1}$. If $y \in \mathcal{B}_{i_2}$, then set $x_2 = y$ and we are done. If $y \notin \mathcal{B}_{i_2}$, then set $t_2 = \max\{t \in [0,1] : ty + (1-t)x_1 \in \mathcal{B}_{i_2}\}$ and $x_2 = t_2 y + (1-t_2)x_1$. Now, continue this way to find points $x_1, \ldots, x_j$ with $x_0, x_1 \in \mathcal{B}_{i_1}$, $\ldots$, $x_{j-1}, x_j \in \mathcal{B}_{i_j}$. Now, the sets $\mathcal{B}_{i_1}, \ldots, \mathcal{B}_{i_j}$ satisfy $\mathcal{B}_{i_s} \neq \mathcal{B}_{i_{s-1}}$ for $s = 1, \ldots, j$ by construction, which implies that they are pairwise disjoint owing to the assumption that they are all convex. Since there are a finite number of sets $\mathcal{B}_i$, this procedure has to stop after a finite number of steps $k \leq n$. We therefore have

$$|f(y) - f(x)| \leq \sum_{i=1}^k |f(x_i) - f(x_{i-1})| \tag{427}$$

$$\leq \sum_{i=1}^k \mathrm{Lip}^{(1)}(f|_{\mathcal{B}_{j_i}})\|x_i - x_{i-1}\|_1 \tag{428}$$

$$\leq \max_{i=1,\ldots,n} \mathrm{Lip}^{(1)}(f|_{\mathcal{B}_i}) \sum_{i=1}^k \|x_i - x_{i-1}\|_1 \tag{429}$$

$$= \max_{i=1,\ldots,n} \mathrm{Lip}^{(1)}(f|_{\mathcal{B}_i})\|y - x\|_1, \tag{430}$$

which establishes (426) as x, y were assumed to be arbitrary. □

Lemma 13. Let $\mathcal{A}, \mathcal{B}_1, \dots, \mathcal{B}_n \subseteq \mathbb{R}^m$ be convex and closed with $\mathcal{A} = \bigcup_{i=1}^n \mathcal{B}_i$. Then, for every $x, y \in \mathcal{A}$, there exists an $\varepsilon \in (0,1]$ and a $\mathcal{B}_{i_0}$ such that $\{ty + (1-t)x : t \in [0,\varepsilon]\} \subseteq \mathcal{B}_{i_0}$.

Proof. Toward a contradiction, suppose that there exist $x, y \in \mathcal{A}$ such that the statement of the lemma is false and set $\mathcal{C} = \bigcup_{i\in\{1,\dots n\}:x\in\mathcal{B}_i} \mathcal{B}_i$ and $\mathcal{D} = \bigcup_{i\in\{1,\dots n\}:x\notin\mathcal{B}_i} \mathcal{B}_i$. Now, fix $k \in \mathbb{N}$ and a $\mathcal{B}_i$ with $x \in \mathcal{B}_i$ arbitrarily. Since the statement is false, we can find a $t_k^{(i)} \in (0, 1/k]$ such that $t_k^{(i)} y + (1 - t_k^{(i)})x \notin \mathcal{B}_i$. As $\mathcal{B}_i$ is convex and $x \in \mathcal{B}_i$, this implies $ty + (1-t)x \notin \mathcal{B}_i$ for all $t \in [t_k^{(i)}, 1]$. By the arbitrariness of $\mathcal{B}_i$, this implies in turn that $x_k := (1/k)y + (1 - 1/k)x \in \mathcal{D}$. Since k was assumed to be arbitrary, we can conclude that $x_k \in \mathcal{D}$ for all $x \in \mathbb{N}$. But $x \notin \mathcal{D}$ and $\lim_{n\to\infty} = x$, which is a contradiction to the fact that $\mathcal{D}$ as the finite union of closed sets is closed. □

Finally, we need the following version of the area formula.

Theorem 8. [18, Corollary 5.1.13] Let $m, n \in \mathbb{N}$ with $m \leq n$. Furher, let $\mathcal{A} \subseteq \mathbb{R}^m$ be Lebesgue measurable and let $f = (f_1, \dots, f_n)^\mathsf{T} \colon \mathcal{A} \to \mathbb{R}^n$ be Lipschitz. Finally, let $g \colon \mathcal{A} \to [0,\infty)$ be Lebesgue measurable and define $h \colon \mathbb{R}^n \to [0,\infty]$ according to

$$h(y) = \sum_{x\in\mathcal{A}\cap f^{-1}(y)} g(x). \tag{431}$$

Then, h is $\mathcal{H}^m$-measurable and

$$\int_{\mathcal{A}} g\, Jf \,\mathrm{d}\mathcal{L}^m = \int_{\mathbb{R}^n} h \,\mathrm{d}\mathcal{H}^m. \tag{432}$$

B Tools from General Measure Theory

This section summarizes the main definitions and properties in general measure theory needed in this work. We follow mainly the exposition in [23, Section 1].

Definition 11. Let $(\mathcal{X}, \rho)$ be a metric space and designate $\mathscr{P}(\mathcal{X}) = \{\mathcal{A} : \mathcal{A} \subseteq \mathcal{X}\}$. A measure μ is mapping $\mu \colon \mathscr{P}(\mathcal{X}) \to [0,\infty]$ satisfying

(i) $\mu(\emptyset) = 0$;

(ii) $\mu(\mathcal{E}) \leq \mu(\mathcal{F})$ for all $\mathcal{E} \subseteq \mathcal{F} \subseteq \mathcal{X}$;

(iii) $\mu\Big(\bigcup_{i\in\mathbb{N}} \mathcal{E}_i\Big) \leq \sum_{i\in\mathbb{N}} \mu(\mathcal{E}_i)$ for all $\mathcal{E}_1, \mathcal{E}_2, \dots \subseteq \mathcal{X}$;

Definition 12. Let $(\mathcal{X}, \rho)$ be a metric space equipped with a measure μ. If $\mathcal{A} \subseteq \mathcal{X}$ satisfies $\mu(\mathcal{X} \setminus \mathcal{A}) = 0$, then we say that μ is supported on $\mathcal{A}$. We set

$$\operatorname{spt}(\mu) = \mathcal{X} \setminus \bigcup \{\mathcal{V} \subseteq \mathcal{X} : \mathcal{V} \text{ is open and } \mu(\mathcal{V}) = 0\}, \tag{433}$$

which is the smallest closed support set.

Definition 13. Let $(\mathcal{X}, \rho)$ be a metric space equipped with a measure μ. A set $\mathcal{A} \subseteq \mathcal{X}$ is called μ-measurable if

$$\mu(\mathcal{E}) = \mu(\mathcal{E} \cap \mathcal{A}) + \mu(\mathcal{E} \setminus \mathcal{A}) \quad \text{for all } \mathcal{E} \subseteq \mathcal{X}. \tag{434}$$

We set $\mathscr{F}(\mu) = \{\mathcal{A} \subseteq \mathcal{X} : \mathcal{A} \text{ is } \mu\text{-measurable}\}$.

Theorem 9. [23, Theorem 1.4] Let $(\mathcal{X}, \rho)$ be a metric space equipped with a measure μ. Then, $\mathcal{M}(\mu)$ is a σ-algebra containing all sets of measure zero. Moreover, μ is countably additive on $\mathcal{M}(\mu)$.

Definition 14. Let $(\mathcal{X}, \rho)$ be a metric space equipped with a measure μ. Then, μ is called

(i) locally finite if, for every $x \in \mathcal{X}$, there exists an $r \in (0, \infty)$ such that $\mu(\{y \in \mathcal{X} : d(x, y) \leq r\}) < \infty$;

(ii) Borel if $\mathscr{M}(\mu)$ contains all Borel sets;

(iii) Borel regular if it is Borel and for every $\mathcal{E} \subseteq \mathcal{X}$, there exists a Borel set $\mathcal{B} \subseteq \mathcal{X}$ such that $\mathcal{E} \subseteq \mathcal{B}$ and $\mu(\mathcal{E}) = \mu(\mathcal{B})$;

(iv) a Radon measure if it is Borel and satisfies the following properties:

   (a) $\mu(\mathcal{K}) < \infty$ for all compact sets $\mathcal{K} \subseteq \mathcal{X}$;

   (b) $\mu(\mathcal{V}) = \sup\{\mu(\mathcal{K}) : \mathcal{K} \subseteq \mathcal{V}, \mathcal{K} \text{ is compact}\}$ for all open sets $\mathcal{V} \subseteq \mathcal{X}$;

   (c) $\mu(\mathcal{A}) = \inf\{\mu(\mathcal{V}) : \mathcal{A} \subseteq \mathcal{V}, \mathcal{V} \text{ is open}\}$ for all sets $\mathcal{A} \subseteq \mathcal{X}$.

Lemma 14. [23, Corollary] A measure on $\mathbb{R}^n$ is a Radon measure if and only if it is a locally finite Borel measure.

We have the following properties for sums and limits of Borel regular measures.

Lemma 15. Let μ_1 and μ_2 be Borel regular measures on the metric space $(\mathcal{X}, \rho)$. Then, $\mu_1 + \mu_2$ is a Borel regular measure.

Proof. It follows immediately from Definition 11 that $\mu_1 + \mu_2$ is a measure. Fix a Borel set $\mathcal{B}$ and a set $\mathcal{E} \subseteq \mathcal{X}$ arbitrarily. Then, we have

$$(\mu_1 + \mu_2)(\mathcal{E}) = \mu_1(\mathcal{E}) + \mu_2(\mathcal{E}) \tag{435}$$
$$= \mu_1(\mathcal{E} \cap \mathcal{B}) + \mu_1(\mathcal{E} \setminus \mathcal{B}) + \mu_2(\mathcal{E} \cap \mathcal{B}) + \mu_2(\mathcal{E} \setminus \mathcal{B}) \tag{436}$$
$$= (\mu_1 + \mu_2)(\mathcal{E} \cap \mathcal{B}) + (\mu_1 + \mu_2)(\mathcal{E} \setminus \mathcal{B}), \tag{437}$$

which implies that $\mu_1 + \mu_2$ is a Borel measure. Finally, fix a set $\mathcal{A} \subseteq \mathcal{X}$ arbitrarily. Then, there exist Borel sets $\mathcal{B}_1, \mathcal{B}_2$ satisfying $\mathcal{A} \subseteq \mathcal{B}_i$ and $\mu_i(\mathcal{A}) = \mu_i(\mathcal{B}_i)$ for $i = 1, 2$. Now, set $\mathcal{B} = \mathcal{B}_1 \cap \mathcal{B}_2$. Then, we have $\mathcal{A} \subseteq \mathcal{B}$ and $\mu_k(\mathcal{A}) \leq \mu_k(\mathcal{B}) \leq \mu_k(\mathcal{B}_k) = \mu_k(\mathcal{A})$, which implies $(\mu_1 + \mu_2)(\mathcal{A}) = (\mu_1 + \mu_2)(\mathcal{B})$. Hence, $\mu_1 + \mu_2$ is Borel regular. □

Lemma 16. Let μ_1 and μ_2 be Borel regular measures on the metric space $(\mathcal{X}, \rho)$ satisfying $\mu_1(\mathcal{B}) = \mu_2(\mathcal{B})$ for all Borel sets $\mathcal{B} \subseteq \mathcal{X}$. Then, $\mu_1 = \mu_2$.

Proof. Fix an arbitrary set $\mathcal{A} \subseteq \mathcal{X}$. Then, there must exist Borel sets $\mathcal{B}_k \subseteq \mathcal{X}$ satisfying $\mathcal{A} \subseteq \mathcal{B}$ and $\mu_k(\mathcal{B}_k)$ for $k = 1, 2$. Set $\mathcal{B} = \mathcal{B}_1 \cap \mathcal{B}_2$. Then, we have

$$\mu_k(\mathcal{A}) \leq \mu_k(\mathcal{B}) \leq \mu_k(\mathcal{B}_k) = \mu_k(\mathcal{A}) \quad \text{for } k = 1, 2, \tag{438}$$

which implies

$$\mu_1(\mathcal{A}) = \mu_1(\mathcal{B}) = \mu_2(\mathcal{B}) = \mu_2(\mathcal{A}). \tag{439}$$

□

Lemma 17. Let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence of measures μ_k on the metric space $(\mathcal{X}, \rho)$. Then, the following properties hold. Suppose that $\mu_k(\mathcal{E}) \leq \mu_{k+1}(\mathcal{E})$ for all $k \in \mathbb{N}$ and $\mathcal{E} \subseteq \mathcal{X}$ and define $\mu_\infty\colon \mathscr{P}(\mathcal{X}) \to [0, \infty]$ setwise according to

$$\mu_\infty(\mathcal{E}) = \lim_{k\to\infty} \mu_k(\mathcal{E}) \quad \text{for all } \mathcal{E} \subseteq \mathcal{X}. \tag{440}$$

Then, μ_∞ is a measure. If, in addition, μ_k is Borel (regular) for all $k \in \mathbb{N}$, so is μ_∞.

Proof. We first establish that μ_∞ is a measue. (i) and (ii) in Definition 11 follow immediately from (440). To establish (iii) in Definition 11, fix an arbitrary collection $\{\mathcal{E}_i : i \in \mathbb{N}\}$ of sets $\mathcal{E}_i \subseteq \mathcal{X}$ and set $\mathcal{E} = \bigcup_{i\in\mathbb{N}} \mathcal{E}_i$. Since

$$\mu_k(\mathcal{E}) \leq \sum_{\ell\in\mathbb{N}} \mu_k(\mathcal{E}_\ell) \leq \sum_{\ell\in\mathbb{N}} \mu_\infty(\mathcal{E}_\ell) \text{ for all } k \in \mathbb{N}, \tag{441}$$

we conclude that

$$\mu_\infty(\mathcal{E}) = \lim_{k\to\infty} \mu_k(\mathcal{E}) \leq \sum_{k\in\mathbb{N}} \mu_\infty(\mathcal{E}_k). \tag{442}$$

This proves (iii) in Definition 11 and, in turn, that μ_∞ is a measure.

Next, suppose that μ_k is a Borel measure for all $k \in \mathbb{N}$, and fix a Borel set $\mathcal{B}$ and a set $\mathcal{E} \subseteq \mathcal{X}$ arbitrarily. Then, we have

$$\begin{aligned}
\mu_\infty(\mathcal{E}) &= \lim_{k\to\infty} \mu_k(\mathcal{E}) && (443)\\
&= \lim_{k\to\infty} (\mu_k(\mathcal{E} \cap \mathcal{B}) + \mu_k(\mathcal{E} \setminus \mathcal{B})) && (444)\\
&= \lim_{k\to\infty} \mu_k(\mathcal{E} \cap \mathcal{B}) + \lim_{k\to\infty} \mu_k(\mathcal{E} \setminus \mathcal{B}) && (445)\\
&= \mu_\infty(\mathcal{E} \cap \mathcal{B}) + \mu_\infty(\mathcal{E} \setminus \mathcal{B}), && (446)
\end{aligned}$$

which prove that μ_∞ is a Borel measure.

Finally, suppose that μ_k is Borel regular for all $k \in \mathbb{N}$ and fix $\mathcal{E} \subseteq \mathcal{X}$ arbitrarily. Then, for every $k \in \mathbb{N}$, there exists a Borel set $\mathcal{B}_k$ satisfying $\mathcal{E} \subseteq \mathcal{B}_k$ and $\mu_k(\mathcal{E}) = \mu_k(\mathcal{B}_k)$. Set $\mathcal{B} = \bigcap_{k\in\mathbb{N}} \mathcal{B}_k$. Then, we have $\mathcal{E} \subseteq \mathcal{B}$ and $\mu_k(\mathcal{E}) \leq \mu_k(\mathcal{B}) \leq \mu_k(\mathcal{B}_k) = \mu_k(\mathcal{E})$ for all $k \in \mathbb{N}$, which implies

$$\mu_\infty(\mathcal{E}) = \lim_{k\to\infty} \mu_k(\mathcal{E}) = \lim_{k\to\infty} \mu_k(\mathcal{B}) = \mu_\infty(\mathcal{B}). \tag{447}$$

Therefore, μ_∞ is Borel regular. □

Definition 15. Let $(\mathcal{X}, \rho)$ be a metric space equipped with a measure μ and $\mathcal{C} \subseteq \mathcal{X}$. Then, we defined $\mu|_{\mathcal{C}} \colon \mathscr{P}(\mathcal{X}) \to [0, \infty]$ according to

$$\mu|_{\mathcal{C}}(\mathcal{E}) = \mu(\mathcal{E} \cap \mathcal{C}) \quad \text{for all } \mathcal{E} \in \mathscr{P}(\mathcal{X}). \tag{448}$$

Theorem 10. [23, Theorem 1.9] Let $(\mathcal{X}, \rho)$, μ, and $\mathcal{C}$ be as in Definition 15. Then, the following properties hold.

(i) $\mu|_{\mathcal{C}}$ is a measure and $\mathscr{M}(\mu) \subseteq \mathscr{M}(\mu|_{\mathcal{C}})$.

(ii) If, in addition, μ is Borel regular and $C \in \mathscr{M}(\mu)$ with $\mu(\mathcal{C}) < \infty$, then $\mu|_{\mathcal{C}}$ is Borel regular.

Definition 16. Let $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \sigma)$ be a metric space and consider a mapping. $f \colon \mathcal{X} \to \mathcal{Y}$.

(i) f is called Borel if $f^{-1}(\mathcal{B})$ is a Borel set in $\mathcal{X}$ for all Borel sets in $\mathcal{Y}$;

(ii) if μ is a measure on $\mathcal{X}$, then f is called μ-measurable if $f^{-1}(\mathcal{V}) \in \mathscr{M}(\mu)$ for all open sets $\mathcal{V} \subseteq \mathcal{Y}$.

Definition 17. Let $(\mathcal{X}, \rho)$ be a metric space equipped with a measure μ and let $f \colon \mathcal{X} \to [0, \infty]$ be μ-measurable. We define $f\mu \colon \mathscr{P}(\mathcal{X}) \to [0, \infty]$ according to

$$f\mu(\mathcal{E}) = \inf \left\{ \int_{\mathcal{B}} f \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \subseteq \mathcal{B} \right\}. \tag{449}$$

Lemma 18. Let $(\mathcal{X}, \rho)$, μ, and f be as in Definition 17. Then, $f\mu$ is a measure satisfying $f\mu \ll \mu$ and $\mathscr{M}(\mu) \subseteq \mathscr{M}(f\mu)$. Moreover, we have

$$(f\mu)|_{\mathcal{A}} = f(\mu|_{\mathcal{A}}) \quad \text{for all } \mathcal{A} \in \mathscr{M}(\mu). \tag{450}$$

If, in addition, μ is Borel (regular), so is $f\mu$.

Proof. (i) and (ii) in Definition 11 follow immediately from (449). To establish (iii) in Definition 11, fix an arbitrary collection $\{\mathcal{E}_i : i \in \mathbb{N}\}$ of sets $\mathcal{E}_i \subseteq \mathcal{X}$ and set $\mathcal{E} = \bigcup_{i \in \mathbb{N}} \mathcal{E}_i$. Then, we have

$$f\nu(\mathcal{E}) = \inf \left\{ \int_{\mathcal{B}} f \, \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \subseteq \mathcal{B} \right\} \tag{451}$$

$$\leq \inf \left\{ \int_{\bigcup_{i \in \mathbb{N}} \mathcal{B}_i} f \, \mathrm{d}\mu : \mathcal{B}_i \in \mathscr{M}(\mu), \mathcal{E}_i \subseteq \mathcal{B}_i \text{ for all } i \in \mathbb{N} \right\} \tag{452}$$

$$\leq \inf \left\{ \sum_{i \in \mathbb{N}} \int_{\mathcal{B}_i} f \, \mathrm{d}\mu : \mathcal{B}_i \in \mathscr{M}(\mu), \mathcal{E}_i \subseteq \mathcal{B}_i \text{ for all } i \in \mathbb{N} \right\} \tag{453}$$

$$= \sum_{i \in \mathbb{N}} \inf \left\{ \int_{\mathcal{B}_i} f \, \mathrm{d}\mu : \mathcal{B}_i \in \mathscr{M}(\mu), \mathcal{E}_i \subseteq \mathcal{B}_i \right\} \tag{454}$$

$$= \sum_{i \in \mathbb{N}} f\nu(\mathcal{E}_i), \tag{455}$$

where in (453) we applied [6, Corollary 4.9]. We next prove $f\mu \ll \mu$. To this end, suppose that $\mathcal{A} \subseteq \mathcal{X}$ with $\mu(\mathcal{A}) = 0$. Since this implies $\mathcal{A} \in \mathscr{M}(\mu)$ and, in turn, $f\mu(\mathcal{A}) = \int_{\mathcal{A}} f \, \mathrm{d}\mu$, we conclude that $f\mu(\mathcal{A}) = 0$ thanks to [6, Corollary 4.10]. Next, we establish $\mathscr{M}(\mu) \subseteq \mathscr{M}(f\mu)$. To this end, fix $\mathcal{A} \in \mathscr{M}(\mu)$ and $\mathcal{E} \subseteq \mathcal{X}$ arbitrary. Then, we have

$$f\mu(\mathcal{E}) \leq f\mu(\mathcal{E} \cap \mathcal{A}) + f\mu(\mathcal{E} \setminus \mathcal{A}) \tag{456}$$

thanks to (ii) in Definition 11, an, by (449), we obtain

$$\begin{aligned}
f\mu(\mathcal{E}) &= \inf\left\{\int_{\mathcal{B}} f \, \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \subseteq \mathcal{B}\right\} && (457)\\
&= \inf\left\{\int_{\mathcal{B}\cap\mathcal{A}} f \, \mathrm{d}\mu + \int_{\mathcal{B}\setminus\mathcal{A}} f \, \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \subseteq \mathcal{B}\right\} && (458)\\
&\geq \inf\left\{\int_{\mathcal{B}} f \, \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \cap \mathcal{A} \subseteq \mathcal{B}\right\} && (459)\\
&\quad + \inf\left\{\int_{\mathcal{B}} f \, \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \setminus \mathcal{A} \subseteq \mathcal{B}\right\} && (460)\\
&= f\mu(\mathcal{E} \cap \mathcal{A}) + f\mu(\mathcal{E} \setminus \mathcal{A}), && (461)
\end{aligned}$$

where in (458) we applied [6, Corollary 4.9]. Now, (450) follows from

$$\begin{aligned}
(f\mu)|_{\mathcal{A}}(\mathcal{E}) &= \inf\left\{\int_{\mathcal{B}} f \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \cap \mathcal{A} \subseteq \mathcal{B}\right\} && (462)\\
&= \inf\left\{\int_{\mathcal{B}\cap\mathcal{A}} f \mathrm{d}\mu : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \subseteq \mathcal{B}\right\} && (463)\\
&= \inf\left\{\int_{\mathcal{B}} f \mathrm{d}\mu|_{\mathcal{A}} : \mathcal{B} \in \mathscr{M}(\mu), \mathcal{E} \subseteq \mathcal{B}\right\} && (464)\\
&= f(\mu|_{\mathcal{A}})(\mathcal{E}) \quad \text{for all } \mathcal{E} \subseteq \mathcal{X}. && (465)
\end{aligned}$$

If μ is Borel, so is $f\mu$ since $\mathscr{M}(\mu) \subseteq \mathscr{M}(f\mu)$. Finally, suppose that μ is Borel regular and fix $\mathcal{E} \subseteq \mathcal{X}$ arbitrarily. Then, there exists a Borel set $\mathcal{B}$ satisfying $\mathcal{E} \subseteq \mathcal{B}$ and $\mu(\mathcal{E}) = \mu(\mathcal{B})$. Now, (449) implies $f\mu(\mathcal{E}) \leq f\mu(\mathcal{B})$. To establish $f\mu(\mathcal{E}) \geq f\mu(\mathcal{B})$, fix $\mathcal{A} \in \mathscr{M}(\mu)$ with $\mathcal{E} \subseteq \mathcal{A}$ arbitrarily. Then, we have

$$\mu(\mathcal{E}) \leq \mu(\mathcal{B} \cap \mathcal{A}) \leq \mu(\mathcal{B}) = \mu(\mathcal{E}), \tag{466}$$

which implies $\mu(\mathcal{B} \cap \mathcal{A}) = \mu(\mathcal{B})$ and, in turn,

$$\mu(\mathcal{B} \setminus \mathcal{A}) = \mu(\mathcal{B}) - \mu(\mathcal{B} \cap \mathcal{A}) = 0 \tag{467}$$

since $\mathcal{A} \in \mathscr{M}(\mu)$. We thus have $f\mu(\mathcal{B} \setminus \mathcal{A}) = 0$ thanks to $f\mu \ll \mu$ so that

$$f\mu(\mathcal{B}) = f\mu(\mathcal{B} \setminus \mathcal{A}) - f\mu(\mathcal{B} \cap \mathcal{A}) = f\mu(\mathcal{B} \setminus \mathcal{A}) \leq f\mu(\mathcal{A}) \tag{468}$$

owing to $\mathcal{A} \in \mathscr{M}(f\mu)$. Since $\mathcal{A}$ was assumed to be arbitrary, (468) implies $f\mu(\mathcal{E}) \geq f\mu(\mathcal{B})$. □

Definition 18. Let $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \sigma)$ be metric spaces, let μ be measure on $\mathcal{X}$, and suppose that $f: \mathcal{X} \to \mathcal{Y}$ is μ-measurable. Then, we define $f\#\mu: \mathscr{P}(\mathcal{Y}) \to [0, \infty]$ according to

$$f\#\mu(\mathcal{E}) = \mu(f^{-1}(\mathcal{E})) \quad \text{for all } \mathcal{E} \subseteq \mathcal{Y}. \tag{469}$$

Lemma 19. Let $(\mathcal{X}, \rho)$, $(\mathcal{Y}, \sigma)$, μ, and f be as in Definition 18. Then, $f\#\mu$ is a measure, and $\mathcal{A} \in \mathscr{M}(f\#\mu)$ provided that $f^{-1}(\mathcal{A}) \in \mathscr{M}(\mu)$. If, in addition, μ is a Borel measure and f is Borel measurable, then $f\#\mu$ is a Borel measure.

Proof. It follows immediately form (469) that $f\#\mu$ is a measure by verifying the defining properties in Definition 11. Suppose that $\mathcal{A} \subseteq \mathcal{Y}$ satisfies $f^{-1}(\mathcal{A}) \in \mathscr{M}(\mu)$. Then, we have

$$\begin{aligned} f\#\mu(\mathcal{E}) &= \mu(f^{-1}(\mathcal{E})) && (470)\\ &= \mu(f^{-1}(\mathcal{E}) \cap f^{-1}(\mathcal{A})) + \mu(f^{-1}(\mathcal{E}) \setminus f^{-1}(\mathcal{A})) && (471)\\ &= \mu(f^{-1}(\mathcal{E} \cap \mathcal{A})) + \mu(f^{-1}(\mathcal{E} \setminus \mathcal{A})) && (472)\\ &= f\#\mu(\mathcal{E} \cap \mathcal{A}) + f\#\mu(\mathcal{E} \setminus \mathcal{A}) \quad \text{for all } \mathcal{E} \subseteq \mathcal{Y}, && (473) \end{aligned}$$

which implies $\mathcal{A} \in \mathscr{M}(f\#\mu)$. Finally, suppose that μ is a Borel measure and fix a Borel set $\mathcal{B} \subseteq \mathcal{Y}$ arbitrarily. Now, suppose that μ f is Borel measurable, and fix a Borel set $\mathcal{B} \subseteq \mathcal{Y}$ arbitrarily. Then, $f^{-1}(\mathcal{B})$ is a Borel set and, in particular, μ-measurable. Thus, $\mathcal{B} \in \mathscr{M}(f\#\mu)$, which establishes that $f\#\mu$ is a Borel measure. □

Theorem 11. [23, Theorem 1.18] Let $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \sigma)$ be be separable metric spaces. if $f: \mathcal{X} \to \mathcal{Y}$ is continuous and μ is a Radon measure on $\mathcal{X}$ with compact support $\operatorname{spt}(\nu)$, then $f\#\mu$ is a Radon measure. Moreover, $\operatorname{spt}(f\#\mu) = f(\operatorname{spt}(\mu))$.

Theorem 12. [23, Theorem 1.20] Let $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \sigma)$ be metric spaces and suppose that $f: \mathcal{X} \to \mathcal{Y}$ is a continuous surjection. Then, for every Radon measure ν on $\mathcal{Y}$, there exists a Radon measure μ on $\mathcal{X}$ satisfying $\nu = f\#\mu$.

Corollary 9. Let $\mathcal{A} \subseteq \mathbb{R}^m$ be compact and $f: \mathcal{A} \to \mathbb{R}^n$ Lipschitz. Further, suppose that ν is a Radon measure on $\mathbb{R}^n$ supported on the m-rectifiable set $\mathcal{E} = f(\mathcal{A})$. Then, there exists a Radon measure μ on $\mathcal{A}$ such that $\nu = f\#\mu$.

Proof. Equip $f(\mathcal{A})$ with the measure $\tilde{\nu}$ defined according to $\tilde{\nu}(\mathcal{C}) = \nu(\mathcal{C})$ for all $\mathcal{E} \subseteq f(\mathcal{A})$. Clearly, $\tilde{\nu}$ is a Radon measure. Moreover, by Theorem 12, there exists a Radon measure μ on $\mathcal{A}$ such that $\tilde{\nu} = f\#\mu$. Finally, we have

$$\begin{aligned} \nu(\mathcal{B}) = \nu(\mathcal{B} \cap \mathcal{E}) = \tilde{\nu}(\mathcal{B} \cap \mathcal{E}) &= \mu(f^{-1}(\mathcal{B} \cap \mathcal{E})) && (474)\\ &= \mu(f^{-1}(\mathcal{B})) = (f\#\mu)(\mathcal{B}) && (475) \end{aligned}$$

for all Borel sets $\mathcal{B} \subseteq \mathbb{R}^m$, which establishes $\nu = f\#\mu$ thanks to Lemma 16. □

Definition 19. Let $(\mathcal{X}, \rho)$ be a metric space. For every $s \in [0, \infty)$, the Hausdorff measure $\mathscr{H}^s: \mathscr{P}(\mathcal{X}) \to [0, \infty]$ is defined according to

$$\mathscr{H}^s(\mathcal{E}) = \lim_{\delta \to 0} \mathscr{H}^s_\delta(\mathcal{E}) \tag{476}$$

with

$$\mathscr{H}^s_\delta(\mathcal{E}) = \frac{\pi^{s/2}}{2^s\Gamma(1+s/2)} \inf\Big\{\sum_{i\in\mathbb{N}} : \mathcal{E} \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{F}_i, \operatorname{diam}(\mathcal{E}_i) \leq \delta\Big\} \tag{477}$$

for all $\mathcal{E} \subseteq \mathcal{X}$ and $\delta \in (0,\infty)$.

The Hausdorff measures $\mathscr{H}^s$ are Borel regular measures [23, Corollary 4.5].

Lemma 20. Let $(\mathcal{X},\rho)$ be a metric space equipped with the Hausdorff measure $\mathscr{H}^s$ and suppose that $\mathcal{A} \subseteq \mathcal{X}$ is $\mathscr{H}^s$-measurable with $\mathscr{H}^s(\mathcal{A}) < \infty$. Then, $\mathscr{H}^s|_{\mathcal{A}}$ is a Radon measure.

Proof. By assumption, $\mathscr{H}^s|_{\mathcal{A}}$ is a finite measure. Moreover, it is Borel regular thanks to (i) in Theorem 10 and Borel regularity of $\mathscr{H}^s$. Therefore, $\mathscr{H}^s|_{\mathcal{A}}$ is a Radon measure. ☐

Lemma 21. [2, Proposition 2.49] [8, Lemma 3.2] The Hausdorff measures $\mathscr{H}^s$ on $\mathbb{R}^n$ have the following properties:

(i) For every $s_1, s_2 \in [0,\infty)$ with $s_1 < s_2$ and $\mathcal{E} \subseteq \mathbb{R}^n$, $\mathscr{H}^{s_2}(\mathcal{E}) > 0$ implies $\mathscr{H}^{s_1}(\mathcal{E}) = \infty$;

(ii) If $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz, then

$$\mathscr{H}^s(f(\mathcal{E})) \leq \operatorname{Lip}^s(f)\mathscr{H}^s(\mathcal{E}) \quad \text{for all } s \in [0,\infty) \text{ and } \mathcal{E} \subseteq \mathbb{R}^m; \tag{478}$$

(iii) If $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz and $\mathcal{A} \subseteq \mathbb{R}^m$ is Lebesgue measurable, then $f(\mathcal{A}) \in \mathscr{M}(\mathscr{H}^m)$;

(iv) $\mathscr{H}^n(\mathcal{B}) = \mathscr{L}^n(\mathcal{B})$ for all Borel sets $\mathcal{B} \subseteq \mathbb{R}^n$.

Definition 20. [18, Definition 2.5.1] Let $(\mathcal{X},\rho)$ be a metric space. The Hausdorff dimension of a set $\mathcal{E} \subseteq \mathcal{X}$ is defined according to

$$\dim_{\mathrm{H}}(\mathcal{E}) = \sup\{s \in [0,\infty] : \mathscr{H}^s(\mathcal{E}) > \infty\}. \tag{479}$$

If μ is a measure on $\mathcal{X}$, then we set $\dim_{\mathrm{H}}(\mu) = \dim_{\mathrm{H}}(\operatorname{spt}(\mu))$.

Lemma 22. [18, Definition 2.5.1] Let $(\mathcal{X},\rho)$ be a metric space and $\mathcal{E} \subseteq \mathcal{X}$. Then, the Hausdorff dimension satisfies

$$\begin{aligned}\dim_{\mathrm{H}}(\mathcal{E}) &= \sup\{s \in [0,\infty] : \mathscr{H}^s(\mathcal{E}) = \infty\} &(480)\\ &= \inf\{s \in [0,\infty] : \mathscr{H}^s(\mathcal{E}) < \infty\} &(481)\\ &= \inf\{s \in [0,\infty] : \mathscr{H}^s(\mathcal{E}) = 0\}. &(482)\end{aligned}$$

C Properties of Wasserstein Distance

This section summarizes the main definitions and properties of Wasserstein distance needed in this work. We follow mainly the exposition in [3, Chapter 7]

Definition 21. Let $(\mathcal{X}, \rho)$ be a separable complete metric space. We denote by $\mathcal{P}(\mathcal{X})$ the set of all Borel probability measures $\mathcal{X}$. A coupling π between $\mu, \nu \in \mathcal{P}(\mathcal{X})$ is a Borel probability measure on $\mathcal{X} \times \mathcal{X}$ satisfying

$$\pi(\mathcal{A}, \mathcal{X}) = \mu(\mathcal{A}) \quad \text{for all } \mathcal{A} \subseteq \mathcal{X} \tag{483}$$

$$\pi(\mathcal{X}, \mathcal{B}) = \nu(\mathcal{B}) \quad \text{for all } \mathcal{B} \subseteq \mathcal{X}. \tag{484}$$

We denote by $\Pi(\mu, \nu)$ the set of all couplings between $\mu, \nu \in \mathcal{P}(\mathcal{X})$. For every $p \in [1, \infty)$, we set

$$\mathcal{P}_p(\mathcal{X}) = \left\{ \mu \in \mathcal{P}(\mathcal{X}) \; \exists x_0 \in \mathcal{X} \text{ with } \int \rho^p(x_0, \cdot) \mathrm{d}\mu < \infty \right\}. \tag{485}$$

Finally, for every $p \in [1, \infty)$, $\alpha \in (0, \infty)$, and $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$, the Wasserstein distance of order p between $\alpha\mu$ and $\alpha\nu$ is defined according to

$$W_p(\alpha\mu, \alpha\nu) = \alpha^{1/p} \left(\inf_{\pi \in \Pi(\mu,\nu)} \int \rho^p \, \mathrm{d}\pi \right)^{1/p}. \tag{486}$$

Lemma 23. Let $(\mathcal{X}, \rho)$ be a separable complete metric space and suppose that $\mathcal{S} \subseteq \mathcal{X}$ is bounded. Then, we have $W_p(\mu, \nu) \leq \mathrm{diam}(\mathcal{S})$ for all $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ satisfying $\mathrm{spt}(\mu) \subseteq \mathcal{S}$ and $\mathrm{spt}(\nu) \subseteq \mathcal{S}$.

Proof. Fix $\pi \in \Pi(\mu, \nu)$ arbitrarily. Then, $\pi|_{\mathcal{S}\times\mathcal{S}}/\pi(\mathcal{S} \times \mathcal{S})$ is also a coupling, and we have

$$W_p^p(\mu, \nu) \leq \frac{1}{\pi(\mathcal{S} \times \mathcal{S})} \int \rho^p \, \mathrm{d}\pi|_{\mathcal{S}\times\mathcal{S}} \leq \mathrm{diam}^p(\mathcal{S}). \tag{487}$$

□

We have the following relation between Wasserstein distances of different order.

Lemma 24. Let $(\mathcal{X}, \rho)$ be a separable complete metric space and suppose that $p, q \in [1, \infty)$ with $q > p$. Further, assume that $\mathrm{spt}(\mu) \subseteq \mathcal{S}$ and $\mathrm{spt}(\nu) \subseteq \mathcal{S}$. Then, we have

$$W_p(\mu, \nu) \;\leq\; W_q(\mu, \nu) \;\leq\; \mathrm{diam}^{q-p}(\mathcal{S}) W^{p/q}(\mu, \nu) \quad \text{for all } \mu, \nu \in \mathcal{P}_p(\mathcal{X}). \tag{488}$$

Proof. Jensen's inequality [28, Theorem 2.3] applied to the convex function $\phi\colon (0, \infty) \to (0, \infty)$, $t \to t^{q/p}$ yields $W_q^q(\mu, \nu) \geq W_p^q(\mu, \nu)$ and

$$W_q^q(\mu, \nu) \leq \frac{\mathrm{diam}^{q-p}(\mathcal{S})}{\pi(\mathcal{S} \times \mathcal{S})} \inf_{\pi \in \Pi(\mu,\nu)} \int \rho^p \mathrm{d}\pi|_{\mathcal{S}\times\mathcal{S}} \tag{489}$$

$$= \mathrm{diam}^{q-p}(\mathcal{S}) \, W_p^p(\mu, \nu). \tag{490}$$

□

Lemma 24 implies that for bounded separable complete metric space, Wasserstein distances of different order are all comparable in the sense of (488). For the particular case where $p = 1$, we have the following dual characterization.

Lemma 25. [3, Equation (7.1.2)] Let $(\mathcal{X}, \rho)$ be a separable complete metric space Then, we have

$$W_1(\mu,\nu) = \sup_{\substack{\psi\colon \mathcal{X}\to\mathbb{R}\\ \mathrm{Lip}(\psi)\le 1}} \left\{ \int \psi \,\mathrm{d}\mu - \int \psi\,\mathrm{d}\nu \right\} \quad \text{for all } \mu,\nu \in \mathcal{P}_1(\mathcal{X}). \tag{491}$$

We need the following properties of the Wasserstein distance of order one under push forwards of measures.

Lemma 26. Let $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \sigma)$ be separable complete metric spaces and let $p \in [1, \infty)$ and $\alpha \in (0, \infty)$. Further, let $\mu \in \mathcal{P}_p(\mathcal{X})$ and suppose that $f, g\colon \mathcal{X} \to \mathcal{Y}$ are μ-measurable. Then, we have

$$W_1(f\#(\alpha\mu), g\#(\alpha\mu)) \le \alpha \int \sigma(f,g)\,\mathrm{d}\mu. \tag{492}$$

Proof. Since

$$W_1(f\#(\alpha\mu), g\#(\alpha\mu)) = W_1(\alpha(f\#\mu), \alpha(g\#\mu)) = \alpha W_1(f\#\mu, g\#\mu) \tag{493}$$

and

$$\int \sigma(f,g)\,\mathrm{d}(\alpha\mu) = \alpha \int \sigma(f,g)\,\mathrm{d}\mu \tag{494}$$

we can assume, without loss of generality, that $\alpha = 1$. We have

$$W_1(f\#\mu, g\#\mu) = \sup_{\substack{\psi\colon \mathcal{X}\to\mathbb{R}\\ \mathrm{Lip}(\psi)\le 1}} \left\{ \int \psi\,\mathrm{d}(f\#\mu) - \int \psi\,\mathrm{d}(g\#\mu) \right\} \tag{495}$$

$$= \sup_{\substack{\psi\colon \mathcal{X}\to\mathbb{R}\\ \mathrm{Lip}(\psi)\le 1}} \left\{ \int \psi\circ f\,\mathrm{d}\mu - \int \psi\circ g\,\mathrm{d}\mu \right\} \tag{496}$$

$$\le \sup_{\substack{\psi\colon \mathcal{X}\to\mathbb{R}\\ \mathrm{Lip}(\psi)\le 1}} \left\{ \int |\psi\circ f - \psi\circ g|\,\mathrm{d}\mu \right. \tag{497}$$

$$\le \int \sigma(f,g)\,\mathrm{d}\mu, \tag{498}$$

where (495) follows from Lemma 25. □

Lemma 27. Let $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \sigma)$ be a separable complete metric spaces and let $p \in [1, \infty)$ and $\alpha \in (0, \infty)$. Further, let $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ be Borel measures and suppose that $f\colon \mathcal{X} \to \mathcal{Y}$ is Lipschitz. Then, we have

$$W_1(f\#(\alpha\mu), f\#(\alpha\nu)) \le \mathrm{Lip}(f) W_1(\alpha\mu, \alpha\nu). \tag{499}$$

Proof. Since

$$W_1(f\#(\alpha\mu), f\#(\alpha\nu)) = W_1(\alpha(f\#\mu), \alpha(f\#\nu)) = \alpha W_1(f\#\mu, f\#\nu) \tag{500}$$

and

$$W_1(\alpha\mu, \alpha\nu) = \alpha W_1(\mu,\nu) \tag{501}$$

we can assume, without loss of generality, that $\alpha = 1$. Now, define $g\colon \mathcal{X} \times \mathcal{X} \to \mathcal{Y} \to \mathcal{Y}$ according to

$$g(x, y) = (f(x), f(y)). \tag{502}$$

Now, fix $\pi \in \Pi(\mu, \nu)$ arbitrarily. Then, we have

$$(g\#\pi)(\mathcal{A} \times \mathcal{Y}) = \pi(f^{-1}(\mathcal{A}) \times \mathcal{X}) = \mu(f^{-1}(\mathcal{A})) = (f\#\mu)(\mathcal{A}) \tag{503}$$

and

$$(g\#\pi)(\mathcal{Y} \times \mathcal{A}) = \pi(\mathcal{X} \times f^{-1}(\mathcal{A})) = \nu(f^{-1}(\mathcal{A})) = (f\#\nu)(\mathcal{A}), \tag{504}$$

which implies $g\#\pi \in \Pi(f\#\mu, f\#\nu)$ so that

$$\begin{aligned} W_1(f\#\mu, f\#\nu) &\le \int \sigma(x, y) \, \mathrm{d}(g\#\pi) && (505) \\ &= \int \sigma(f(x), f(y)) \, \mathrm{d}\pi && (506) \\ &\le \mathrm{Lip}(f) \int \rho(x, y) \, \mathrm{d}\pi, && (507) \end{aligned}$$

which establishes (499) for $\alpha = 1$ as π was assumed to be arbitrary. □

Finally, we need the following result for sequences of measures.

Lemma 28. Let $(\mathcal{X}, \rho)$ be a separable complete metric space satisfying $\sup_{x,y\in\mathcal{X}} \rho(x, y) < \infty$ and let $p \in [1, \infty)$. Further, let $(\mu_i)_{i\in\mathbb{N}}$ and $(\nu_i)_{i\in\mathbb{N}}$ be two sequences of Borel regular probability measures on $\mathcal{X}$ and let $(a_i)_{i\in\mathbb{N}}$ be a sequence in $[0, 1]$ satisfying $\sum_{i\in\mathbb{N}} a_i = 1$. Then, the measures

$$\mu = \sum_{i\in\mathbb{N}} a_i \mu_i \tag{508}$$

and

$$\nu = \sum_{i\in\mathbb{N}} a_i \nu_i \tag{509}$$

are Borel regular, in $\mathcal{P}_p(\mathcal{X})$, and satisfy

$$W_1(\mu, \nu) \le \sum_{i\in\mathbb{N}} a_i W_1(\mu_i, \nu_i). \tag{510}$$

Proof. Set $C = \sup_{x,y\in\mathcal{X}} \rho(x, y)$ and note that $C < \infty$ implies $\mu_i, \nu_i \in \mathcal{P}_p(\mathcal{X})$ for all $i \in \mathbb{N}$. Now, Lemma 17 implies that μ and ν are both Borel regular measures, where convergence in (508) and (509) is setwise. Moreover, $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ since, for fixed $x_0 \in \mathcal{X}$, we have

$$\int \rho(x, x_0) \, \mathrm{d}\mu \le C\mu(\mathcal{X}) = C \sum_{i\in\mathbb{N}} a_i \mu_i(\mathcal{X}) = C \tag{511}$$

and similar for ν. Now, for every $i \in \mathbb{N}$, fix $\pi_i \in \Pi(\mu_i, \nu_i)$ arbitrarily. Then, again by Lemma 17,

$$\pi = \sum_{i \in \mathbb{N}} a_i \pi_i \tag{512}$$

is a Borel regular measure, where convergence in (512) is setwise. Moreover, we have

$$\pi(\mathcal{A} \times \mathcal{X}) = \sum_{i \in \mathbb{N}} a_i \pi_i(\mathcal{A} \times \mathcal{X}) = \sum_{i \in \mathbb{N}} a_i \mu_i(\mathcal{A}) = \mu(\mathcal{A}) \tag{513}$$

and

$$\pi(\mathcal{X} \times \mathcal{A}) = \sum_{i \in \mathbb{N}} a_i \pi_i(\mathcal{X} \times \mathcal{A}) = \sum_{i \in \mathbb{N}} a_i \nu_i(\mathcal{A}) = \nu(\mathcal{A}), \tag{514}$$

which implies $\pi \in \Pi(\mu, \nu)$. Since the coupling π_i was assumed to be arbitrary, we conclude that

$$W_1(\mu, \nu) \le \inf_{\substack{\pi_i \in \Pi(\mu_i, \nu_i) \\ \text{for all } i \in \mathbb{N}}} \int \rho^p \, \mathrm{d}\Big(\sum_{i \in \mathbb{N}} a_i \pi_i \Big) \tag{515}$$

$$= \sum_{i \in \mathbb{N}} a_i \inf_{\pi_i \in \Pi(\mu_i, \nu_i)} \int \rho^p \, \mathrm{d}\pi_i \tag{516}$$

$$= \sum_{i \in \mathbb{N}} a_i W_1(\mu_i, \nu_i), \tag{517}$$

where (516) follows from (512) and approxixmation of ρ through simple functions. □

D Auxiliary Results

Lemma 29. Let $n, N \in \mathbb{N}$ with $N \ge n$, set $\delta = 1/N$, and let $m_1, \dots, m_n \in [0, 1]$ with $\sum_{k=1}^{n} m_k = 1$. Then, we can construct $\hat{w}_1, \dots, \hat{w}_n \in \delta\mathbb{N}$ satisfying $\sum_{k=1}^{n} \hat{w}_k = 1$ and

$$\sum_{k=1}^{n} |\hat{w}_k - w_k| \le 4\delta(n-1). \tag{518}$$

Proof. Set $\hat{p}_k = \delta \lfloor w_k/\delta \rfloor$ for $k = 1, \dots, n-1$ and $\hat{p}_n = 1 - \sum_{k=1}^{n-1} \hat{p}_k$. By construction, we have $\hat{p}_k \in \delta\mathbb{N}_0$ with $\sum_{k=1}^{n} \hat{p}_k = 1$. Moreover, the $\hat{p}_k$ satisfy

$$\sum_{k=1}^{n} |\hat{p}_k - w_k| = 2\delta \sum_{k=1}^{n-1} |\lfloor w_k/\delta \rfloor - w_k/\delta| \le 2\delta(n-1). \tag{519}$$

Next, set

$$\mathcal{C} = \{k \in \{1, \dots, n\} : \hat{p}_k = 0\} \qquad \text{and} \qquad \mathcal{C}^{\perp} = \{1, \dots, n\} \setminus \mathcal{C} \tag{520}$$

and let $\alpha = |\mathcal{C}|\delta$ and $\beta = |\mathcal{C}^\perp|\delta$. Since

$$\sum_{k \in \mathcal{C}^\perp} \hat{p}_k = 1 \geq \alpha + \beta, \tag{521}$$

we can find $\alpha_1, \alpha_2 \ldots, \alpha_{|\mathcal{C}^\perp|} \in \delta\mathbb{N}_0$ satisfying

$$\sum_{k \in \mathcal{C}^\perp} \alpha_k = \alpha \tag{522}$$

such that

$$\hat{w}_k := \hat{p}_k - \alpha_k \in \delta\mathbb{N} \quad \text{for all } k \in \mathcal{C}^\perp. \tag{523}$$

Setting

$$\hat{w}_k = \hat{p}_k + \delta \qquad \text{for all } k \in \mathcal{C}, \tag{524}$$

we have $\sum_{k=1}^n \hat{w}_k = \sum_{k=1}^n \hat{p}_k = 1$ by construction and

$$\sum_{k=1}^n |\hat{w}_k - w_k| \leq \sum_{k=1}^n |\hat{w}_k - \hat{p}_k| + \sum_{k=1}^n |\hat{p}_k - w_k| \tag{525}$$

$$\leq 2\delta(n-1) + 2|\mathcal{C}|\delta \tag{526}$$

$$\leq 4\delta(n-1), \tag{527}$$

where in (526) we used (519) and (522)–(524), which establishes (518). □

Lemma 30. Let $w_1, \ldots, w_K \in (0, 1]$ with $\sum_{k=1}^K w_k = 1$ and set $a_0 = 1/(Kw_1)$ and $b_0 = 0$. For $k = 1, \ldots, K$, set $a_k = 1/(Kw_{k+1}) - 1/(Kw_k)$, $b_k = \sum_{j=1}^k w_j$, and

$$\mathcal{K}_k = \begin{cases} [b_{k-1}, b_k) & \text{if } k < K \\ [b_{k-1}, \infty) & \text{if } k = K. \end{cases} \tag{528}$$

Define $f\colon \mathbb{R} \to \mathbb{R}$ according to

$$f(x) = \begin{cases} x & \text{for all } x \in (-\infty, 0) \\ (x - b_k)/(Kw_k) + k/K & \text{for all } x \in \mathcal{K}_k \text{ and } k = 1, \ldots, K \end{cases} \tag{529}$$

and $g\colon \mathbb{R} \to \mathbb{R}$ as

$$g(x) = -\rho(-x) + \sum_{k=1}^K \rho(x - b_{k-1}) a_{k-1}. \tag{530}$$

Then, $f = g$.

Proof. We have

$$g(x) = -\rho(-x) = x = f(x) \quad \text{for all } x \in (-\infty, 0). \tag{531}$$

Next, we establish $f(x) = g(x)$ for all $x \in [0, \infty)$. To this end, fix $k \in \{1, \ldots, K\}$ arbitrarily and note that

$$g(x) = x \sum_{j=1}^{k} a_{j-1} - \sum_{j=1}^{k} a_{j-1} b_{j-1} \quad \text{for all } x \in \mathcal{K}_k. \tag{532}$$

Now,

$$\sum_{j=1}^{k} a_{j-1} = 1/(K w_k) \tag{533}$$

and

$$K \sum_{j=1}^{k} a_{j-1} b_{j-1} \tag{534}$$
$$= w_1(1/w_2 - 1/w_1) + (w_1 + w_2)(1/w_3 - 1/w_2) \tag{535}$$
$$+ \cdots + (w_1 + w_2 + \cdots + w_{k-1})(1/w_k - 1/w_{k-1}) \tag{536}$$
$$= 1 - k + (w_1/w_2 + (w_1 + w_2)/w_3 + \cdots + (w_1 + \cdots + w_{k-1})/w_k) \tag{537}$$
$$- (w_1/w_2 + (w_1 + w_2)/w_3 + \cdots + (w_1 + \cdots + w_{k-2})/w_{k-1}) \tag{538}$$
$$= 1 - k + b_{k-1}/(w_k). \tag{539}$$

Using (533) and (534)–(539) in (532) yields

$$g(x) = (x - b_{k-1})/(K w_k) + (k-1)/K \tag{540}$$
$$= (x - b_k)/(K w_k) + k/K \tag{541}$$
$$= f(x) \quad \text{for all } x \in \mathcal{K}_k. \tag{542}$$

□

E Properties of the Sawtooth Function

We start with the definition of the sawtooth function.

Definition 22. The sawtooth function $g \colon \mathbb{R} \to [0, 1]$ is defined according to

$$g(x) = \begin{cases} 2x & \text{if } x \in [0, 1/2) \\ 2(1-x) & \text{if } x \in [1/2, 1] \\ 0 & \text{else.} \end{cases} \tag{543}$$

For every $s \in \mathbb{N}$, we set

$$g_s = \underbrace{g \circ g \circ \cdots \circ g}_{s \text{ times}}. \tag{544}$$

We next state some elementary properties of g_s.

Lemma 31. For every $s \in \mathbb{N}$, we have $g_s(x) = 0$ for all $x \in \mathbb{R} \setminus (0, 1)$ and

$$g_s(x) = \begin{cases} 2^s x - \lfloor 2^s x \rfloor & \text{if } \lfloor 2^s x \rfloor \text{ is even} \\ 1 - 2^s x + \lfloor 2^s x \rfloor & \text{if } \lfloor 2^s x \rfloor \text{ is odd.} \end{cases} \tag{545}$$

Proof. Since $g_s(x) = 0$ whenever $g(x) = 0$, we conclude that $g_s(x) = 0$ for all $x \in \mathbb{R} \setminus [0, 1]$. We prove (545) by induction. For $s = 1$, (545) follows from $\lfloor 2x \rfloor = 0$ for all $x \in [0, 1/2)$ and $\lfloor 2x \rfloor = 1$ for all $x \in [1/2, 1]$ and the definition of g. Suppose that (545) holds for $s - 1$ and set $\mathcal{I}_k = [(k-1)/2^s, k/2^s)$ for $k \in \mathbb{N}$. Then, for every $m \in \mathbb{N}$, we have

$$x \in \mathcal{I}_{4m} \Rightarrow \lfloor 2^{s-1}x \rfloor = 2m-1, \lfloor 2^s x \rfloor = 4m-1, \text{ and } g_{s-1}(x) \leq 1/2 \tag{546}$$
$$x \in \mathcal{I}_{4m-1} \Rightarrow \lfloor 2^{s-1}x \rfloor = 2m-1, \lfloor 2^s x \rfloor = 4m-2, \text{ and } g_{s-1}(x) > 1/2 \tag{547}$$
$$x \in \mathcal{I}_{4m-2} \Rightarrow \lfloor 2^{s-1}x \rfloor = 2m-2, \lfloor 2^s x \rfloor = 4m-3, \text{ and } g_{s-1}(x) \geq 1/2 \tag{548}$$
$$x \in \mathcal{I}_{4m-3} \Rightarrow \lfloor 2^{s-1}x \rfloor = 2m-2, \lfloor 2^s x \rfloor = 4m-4, \text{ and } g_{s-1}(x) < 1/2, \tag{549}$$

which implies

$$g_s(x) = \begin{cases} 2g_{s-1}(x) = 4m - 2^s x = 1 - 2^s x + \lfloor 2^s x \rfloor & \text{if } x \in \mathcal{I}_{4m} \\ 2 - 2g_{s-1}(x) = 2^s x - 4m + 2 = 2^s x - \lfloor 2^s x \rfloor & \text{if } x \in \mathcal{I}_{4m-1} \\ 2 - 2g_{s-1}(x) = 2^s x - 4m + 4 = 1 - 2^s x + \lfloor 2^s x \rfloor & \text{if } x \in \mathcal{I}_{4m-2} \\ 2g_{s-1}(x) = 2^s x - 4m + 4 = 2^s x - \lfloor 2^s x \rfloor & \text{if } x \in \mathcal{I}_{4m-4}. \end{cases} \tag{550}$$

□

Lemma 32. For every $s \in \mathbb{N}$, we have

$$g_s = \sum_{k=1}^{2^{s-1}} h_k \tag{551}$$

with $h_k \colon \mathbb{R} \to [0, 1]$, $h_k(x) = g(2^{s-1}x - k + 1)$ for $k = 1, \dots, 2^{s-1}$. The mappings h_k satisfy $h_k(x) = 0$ for all $x \notin \mathcal{F}_k := ((k-1)/2^{s-1}, k/2^{s-1})$. Moreover, for every mapping $f \colon [0, 1] \to \mathbb{R}$ satisfying $f(0) = 0$, we have

$$f \circ g_s = \sum_{k=1}^{2^{s-1}} f \circ h_k. \tag{552}$$

Proof. It follows immediately from Definition 22 that $h_k(x) = 0$ for all $x \notin \mathcal{F}_k$. Moreover, by Lemma 31, we have

$$g_s(k/2^{s-1}) = 2k - \lfloor 2k \rfloor = 0 \qquad \text{for all } k \in \mathbb{Z}. \tag{553}$$

Since the intervals $\mathcal{F}_k$ are pairwise disjoint, it is therefore sufficient to establish $h_k(x) = g_s(x)$ for all $x \in \mathcal{F}_k$ and $k = 1, 2, \dots, 2^{s-1}$. Now, fix $k \in \{1, 2, \dots, 2^{s-1}\}$ arbitrarily and let $\mathcal{F}_k^{(1)} = ((2k-2)/2^s, (2k-1)/2^s)$ and $\mathcal{F}_k^{(2)} = [(2k-1)/2^s, 2k/2^s)$ so that $\mathcal{F}_k = \mathcal{F}_k^{(1)} \cup \mathcal{F}_k^{(2)}$. since $\lfloor 2^s x \rfloor = 2k - 2$ for all $x \in \mathcal{F}_k^{(1)}$ and $\lfloor 2^s x \rfloor = 2k - 1$ for all $x \in \mathcal{F}_k^{(2)}$, we have

$$h_k(x) = \begin{cases} 2^s x - 2k + 2 = 2^s x - \lfloor 2^s x \rfloor & \text{for all } x \in \mathcal{F}_k^{(1)} \\ 2k - 2^s x = 1 - 2^s x + \lfloor 2^s x \rfloor & \text{for all } x \in \mathcal{F}_k^{(2)}, \end{cases} \tag{554}$$

which implies $h_k(s) = g_s(x)$ for all $x \in \mathcal{F}_k$ owing to Lemma 31. To establish (552), fix $k \in \{1, 2, \ldots, 2^{s-1}\}$ arbitrarily. Then, we we have

$$f(g_s(x)) = f\left(\sum_{k=1}^{2^{s-1}} h_k(x)\right) \tag{555}$$

$$= f(h_k(x)) \tag{556}$$

$$= \sum_{k=1}^{2^{s-1}} f(h_k(x)) \quad \text{for all } x \in \mathcal{F}_k, \tag{557}$$

where (555) follows from (551), in (556) is by the fact that $h_k(x) = 0$ for all $x \notin \mathcal{F}_k$, and in (557) used $f(0) = 0$ and again $h_k(x) = 0$ for all $x \notin \mathcal{F}_k$. □

References

[1] R. Adams and J. Fournier, *Sobolev Spaces*, Academic Press, New York, NY, 2nd ed., 2003.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Univ. Press, New York, NY, 2000.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Springer, New York, NY, 2nd ed., 2008.

[4] M. Arjovsky and L. Bottou, *Towards principled methods for training generative adversarial networks*, in International Conference on Learning Representations, 2017, pp. 1–15.

[5] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory, 39 (1993), pp. 930–945.

[6] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*, Wiley, New York, NY, 1995.

[7] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei, *Deep neural network approximation theory*, IEEE Trans. Inform. Theory, 67 (2021), pp. 2581–2623.

[8] L. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, New York, NY, revised ed., 2015.

[9] K. Falconer, *Fractal Geometry*, Wiley, New York, NY, 3rd ed., 2014.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial networks*, Advances in Neural Information Processing Systems, 3 (2014), pp. 1–9.

[11] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*, Springer, Berlin, Germany, 2000.

[12] I. Gühring, G. Kutyniok, and P. Petersen, *Error bounds for approximations with deep ReLu networks in $W^{s,p}$ norms*, Analysis and Applications, 18 (2020), pp. 803–859.

[13] G. E. Hinton, P. Dayan, and M. Revow, *Modeling the manifolds of images of handwritten digits*, IEEE Trans. Neural Netw., 8 (1997), pp. 65–74.

[14] R. Hong and A. Kratsios, *Bridging the gap between approximation and learning via optimal approximation by ReLU MLPs of maximal regularity*, arXiv:2409.12335v1, (2024).

[15] S. Karnik, R. Wang, and M. Iwen, *Neural network approximation of continuous functions in high dimensions with applications to inverse problems*, J. Comput. Appl. Math., 438 (2024), pp. 1–20.

[16] D. P. Kingma and M. Welling, *Auto-encoding variational Bayes*, in Proceedings of the International Conference on Learning Representations, 2014, pp. 1–14.

[17] J. Kinnunen, *Sobolev Spaces*. https://math.aalto.fi/~jkkinnun/files/sobolev_spaces.pdf.

[18] S. G. Krantz and H. R. Parks, *Geometric Integration Theory*, Birkhäuser, Princeton, NJ, 2008.

[19] H. Lee, R. Ge, T. Ma, A. Risteski, and S. Arora, *On the ability of neural nets to express distributions*, in Proceedings of the 2017 Conference on Learning Theory, vol. 65 of Proceedings of Machine Learning Research, Jul. 2017, pp. 1271–1296.

[20] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, *On the self-similar nature of Ethernet traffic (extended version)*, IEEE/ACM Trans. Netw., 2 (1994), pp. 1–15.

[21] H. Lu, Y. Fainman, and R. Hecht-Nielsen, *Image manifolds*, in Proc. SPIE 3307, Applications of Artificial Neural Networks in Image Processing III, San Jose, CA, Apr 1998, pp. 3307–3307.

[22] A. J. Lzenman, *Introduction to manifold learning*, WIREs Comput. Stat., 4 (2012), pp. 439–446.

[23] P. Mattila, *Geometry of Sets and Measures in Euclidean Space: Fractals and Rectifiability*, Cambridge Univ. Press, Cambridge, UK, 1999.

[24] E. J. McShane, *Extension of range of functions*, Bulletin of the American Mathematical Society, 40 (1934), pp. 837–842.

[25] A. Ng and M. Jordan, *On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes*, in Advances in Neural Information Processing Systems, vol. 14, MIT Press, 2002.

[26] Y. Pan, C. Hutter, and H. Bölcskei, *Metric-entropy limits on nonlinear dynamical system learning*, Information Theory, Probability and Statistical Learning: A Festschrift in Honor of Andrew Barron, 2 (2024), pp. 1–44.

[27] D. Perekrestenko, L. Eberhard, and H. Bölcskei, *High-dimensional distribution generation through deep neural networks*, Partial Differential Equations and Applications, 2 (2021), pp. 1–44.

[28] J. E. Pečarić, F. Proschan, and Y. L. Tong, *Convex Functions, Partial Ordering, and Statistical Applications*, Academic Press Ind., London, UK, 1992.

[29] E. Riegler, G. Koliander, D. Stotz, and H. Bölcskei, *Completion of matrices with low description complexity*, Mathematical Statistics and Learning, (2024 (submitted)).

[30] N. Sochen and Y. Y. Zeevi, *Representation of colored images by manifolds embedded in higher dimensional non-Euclidean space*, in Proc. IEEE Int. Conf. on Image Process., Chicago, IL, Oct 1998, pp. 166–170.

[31] T. Tao, *An Introduction to Measure Theory*, AMS, Providence, RI, 2011.

[32] M. J. Telgarsky, *Benefits of depth in neural networks*, in 29th Annual Conference on Learning Theory, vol. 49 of Proceedings of Machine Learning Research, Jun. 2016, pp. 1517–1539.

[33] Y. Yang, L. Zheng, and Y. Wang, *On the capacity of deep generative networks for approximating distributions*, Neutal Networks, 145 (2022), pp. 144–154.