

SAMPLING, METRIC ENTROPY, AND DIMENSIONALITY REDUCTION*

DMITRY BATENKOV[†], OMER FRIEDLAND[‡], AND YOSEF YOMDIN[§]

Abstract. Let Q be a relatively compact subset in a Hilbert space V . For a given $\epsilon > 0$ let $N(\epsilon, Q)$ be the minimal number of linear measurements, sufficient to reconstruct any $x \in Q$ with the accuracy ϵ . We call $N(\epsilon, Q)$ the sampling ϵ -entropy of Q . Using dimensionality reduction, as provided by the Johnson–Lindenstrauss lemma, we show that, in an appropriate probabilistic setting, $N(\epsilon, Q)$ is bounded from above by Kolmogorov’s ϵ -entropy $H(\epsilon, Q)$, defined as $H(\epsilon, Q) = \log M(\epsilon, Q)$, with $M(\epsilon, Q)$ being the minimal number of ϵ -balls covering Q . As the main application, we show that piecewise smooth (piecewise analytic) functions in one and several variables can be sampled with essentially the same accuracy rate as their regular counterparts. For univariate piecewise C^k -smooth functions this result, which settles the so-called Eckhoff conjecture, was recently established in D. Batenkov, *Complete Algebraic Reconstruction of Piecewise-smooth Functions from Fourier Data*, arXiv:1211.0680, 2012 via a deterministic “algebraic reconstruction” algorithm.

Key words. sampling, metric entropy, Johnson–Lindenstrauss lemma, dimensionality reduction

AMS subject classifications. 41A46, 46E20, 94A20, 94A12

DOI. 10.1137/130944436

1. Introduction. In this paper we study the “sampling ϵ -entropy” of functional classes in a Hilbert space V . “Samples” or “measurements” of an element $x \in V$ are assumed to be linear functionals on V , i.e., the values of the scalar products

$$(1.1) \quad \lambda_v(x) = \langle x, v \rangle,$$

where $v \in V$ is a known vector. To simplify the presentation we shall assume in this paper that the measurements are exact. See Remark 3.1, where we discuss measurements of a finite accuracy.

Let Q be a relatively compact subset in V . We start with Kolmogorov’s metric entropy, which measures the “information complexity” of Q in a prescribed resolution ϵ .

DEFINITION 1.1. *Let $\epsilon > 0$. The covering number $M(\epsilon, Q)$ is the minimal number of ϵ -balls B_ϵ in V covering Q . $H(\epsilon, Q) = \log_2 M(\epsilon, Q)$ is called the Kolmogorov ϵ -entropy of Q .*

The metric entropy $H(\epsilon, Q)$ of functional classes was introduced and studied in [23, 24, 25]. It is exactly the number of binary bits of information we need to describe any element $x \in Q$ with the accuracy ϵ . Indeed, to get such a description it is necessary and sufficient to specify the ball B_ϵ (or its center) in the minimal ϵ -covering of Q , containing x . The number of the balls is $M(\epsilon, Q)$, and hence the number of bits required is $H(\epsilon, Q) = \log_2 M(\epsilon, Q)$.

*Received by the editors November 7, 2013; accepted for publication (in revised form) December 9, 2014; published electronically February 10, 2015. This research was supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, by ISF grant 639/09, and by the Minerva Foundation.

<http://www.siam.org/journals/sima/47-1/94443.html>

[†]Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel (batenkov@cs.technion.ac.il).

[‡]Institut de Mathématiques de Jussieu, Université Pierre et Marie Curie (Paris 6), 75005 Paris, France (omer.friedland@imj-prg.fr).

[§]Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel (yosef.yomdin@weizmann.ac.il).

In our definition of the “sampling ϵ -entropy” $N(\epsilon, Q)$ we just add the requirement that the information used to reconstruct unknown $x \in Q$ be provided by linear measurements $\lambda_v(x)$ as defined in (1.1).

DEFINITION 1.2. *Let $\epsilon > 0$. The sampling ϵ -entropy $N(\epsilon, Q)$ of Q is defined as the minimal number of exact linear measurements $\lambda_v(x)$, sufficient to reconstruct any $x \in Q$ with the accuracy ϵ . The set of measurements used may depend on Q and on ϵ , but not on the specific $x \in Q$ to be reconstructed.*

However, mostly in this paper we shall work with a randomized version of the sampling entropy.

DEFINITION 1.3. *Assume that for each N a procedure \mathcal{P} of a random choice of N measurements (i.e., of an orthonormal system of N vectors in V) has been fixed. Let $\epsilon > 0$, $0 < p < 1$, and let Q be a relatively compact subset in a Hilbert space V . The sampling (ϵ, p) -entropy $N(\epsilon, p, Q)$ of Q is defined as the minimal number N with the following property: for each fixed $x \in Q$ apply the procedure \mathcal{P} to pick randomly N linear measurements λ_{v_j} , $j = 1, \dots, N$. Then with probability greater than p the element $x \in Q$ can be reconstructed from the measurements values $\lambda_{v_j}(x)$, $j = 1, \dots, N$, with the accuracy ϵ .*

Notice that the random choice according to the procedure \mathcal{P} does not depend on a specific $x \in Q$. Let us stress that this procedure depends only on the ambient Hilbert space V and on the number of measurements N . One specific choice of such a procedure \mathcal{P} is described in section 2, where we present a version of dimensionality reduction, suitable for our purposes, as provided by the Johnson–Lindenstrauss lemma (JLL) [20]. Hence it can be performed “off line,” once and forever, and still we can expect that with probability p any given $x \in Q$ can be ϵ -reconstructed from the chosen set of measurements.

The JLL is a finite-dimensional result, so in order to apply it we have to find for each $\epsilon > 0$ a sufficiently accurate (on Q) finite-dimensional approximation of the Hilbert space V . We achieve it by making the following assumption, which is satisfied in all the specific examples considered below.

Assumption A. There exists an orthonormal basis (v_1, v_2, \dots) in V such that for each $d \in \mathbb{N}$ the subspace $V_d \subset V$ spanned by v_1, v_2, \dots, v_d satisfies the following: for any $x \in Q$ we have

$$|x - \pi_{V_d}x|_2 \leq Cd^{-\beta}|x|_2,$$

where $\beta > 0$, $C > 0$, π_{V_d} is the projection of V on V_d , and $|\cdot|_2$ is the Euclidean norm on V .

For example, Assumption A holds for Q being the class of univariate piecewise C^k functions and the subspaces V_d of d th partial sums of the Fourier basis of $L^2([-\pi, \pi])$, with $\beta = 1$.

Our main result is the following.

THEOREM 1.4. *Let $\epsilon > 0$ and $0 < p < 1$. Let Q be a relatively compact subset in a Hilbert space V satisfying Assumption A. Then*

$$N(\epsilon, p, Q) \leq \frac{20}{1-p}H(\epsilon/6, Q).$$

The proof is given in section 3.

Notice that the parameters β, C in the Assumption A for Q, V do not enter the statement of Theorem 1.4. However, these parameters affect the required measurements accuracy (compare Remark 3.1).

So, at least for exact random measurements, the requirement that the information on $x \in Q$ comes from *linear samples* does not increase significantly the number of samples we need in order to achieve the prescribed reconstruction accuracy.

Below we recall some classical results providing the sampling entropy for C^k -smooth and analytic functions. Based on a recent result of [4] we obtain a sharp bound for the sampling entropy for univariate piecewise C^k -smooth functions. As the main application, we show in section 4 that *multivariate and univariate piecewise smooth functions, as well as univariate piecewise analytic ones, can be sampled with essentially the same accuracy rate as their regular counterparts*. This fact was conjectured in various forms in many publications in the field. See the discussion after Theorem 4.4.

As a direct consequence of the main result of this paper, we confirm a certain very general principle, which compares the sampling entropy of the classical functional classes, defined by certain regularity assumptions, and of their extensions, which preserve complexity, but may completely destroy regularity. Examples considered in section 4 include piecewise regular functions, finite-dimensional “warping,” and adding finite-dimensional nonregular signals. In all these cases *the randomized sampling entropy of the extended nonregular classes remains essentially the same, as for their regular counterparts*.

1.1. Some references. Several results on compressed sensing and its modifications have recently appeared, where the JLL, as well as Kolmogorov’s n -width, play an important role. See [9, 8, 10, 12, 15, 22] and references therein. On the other hand, recent results in [30] provide important connections of the sampling entropy of subsets with their “mean width” and other geometric characteristics. There are important results on various versions of a “nonlinear width” and its connections with the metric entropy (see [6, 11, 13, 14, 29]). It would be interesting to compare these results with the results of the present paper.

2. Johnson–Lindenstrauss lemma. This result [20] (see also [9, 8, 10, 12, 15, 22] and references therein) is central in numerous problems involving dimensionality reduction. We need the following special case of it.

THEOREM 2.1. *Let $Z \subset \mathbb{R}^d$ be a finite set of cardinality m . Let $n = \frac{20}{1-p} \ln m$, with p being a given probability in $(0, 1)$. Let W be a randomly chosen linear n -dimensional subspace in \mathbb{R}^d and let $\pi_W : \mathbb{R}^d \rightarrow W$ be the orthogonal projection onto W . Then with probability greater than p for each $x, y \in Z$ we have*

$$(2.1) \quad \frac{1}{2}|x - y|_2 \leq \sqrt{\frac{d}{n}} |\pi_W x - \pi_W y|_2 \leq 2|x - y|_2.$$

We recall (following [2]) the specific random choice procedure \mathcal{P} used, as well as the main steps of the proof. The orthogonal basis e_1, \dots, e_n of W is formed as follows: e_1 is a random unit vector chosen uniformly in \mathbb{S}^{d-1} . The second vector e_2 is a random unit vector from the space orthogonal to e_1 , etc. For a fixed $x \in \mathbb{R}^d$ with $|x|_2 = 1$ the coordinates of the projection $\pi x \in W$ according to the basis e_1, \dots, e_n are independent random variables with the expectation $\sqrt{\frac{1}{d}}$. Application of the central limit theorem to $|\pi_W x|_2^2$, which is the sum of squares of n coordinates, provides, assuming $n \geq \frac{20}{1-p} \ln m$, the following bound:

$$P \left\{ \frac{1}{2}|x|_2 \leq \sqrt{\frac{d}{n}} |\pi_W x|_2 \leq 2|x|_2 \right\} \geq 1 - \frac{1-p}{m^2}.$$

Finally, applying a union bound over all $\binom{m}{2}$ pairs of distances in Z , we obtain the required estimate.

3. Entropy meets dimensionality reduction. In this section we prove Theorem 1.4. Let $\epsilon > 0$ and $0 < p < 1$. Let Q be a relatively compact subset in a Hilbert space V satisfying Assumption A, with the parameters β, C . Let $\epsilon_1 = \epsilon/6$, and $m = M + 1$, where $M = M(\epsilon_1, Q)$ is the covering number of Q with ϵ_1 -balls. Let $x \in Q$ be an unknown vector. We have to show that for $n = \frac{20}{1-p} \ln m$, with probability greater than p , the vector x can be reconstructed with an error at most ϵ from random n measurements. The reconstruction algorithm proceeds in the following steps:

I. *Preprocessing.*

(1) Let $d = (\frac{RC}{\epsilon_1})^{\frac{1}{\beta}}$, where $R > 0$ is the radius of the minimal ball centered at $0 \in V$, containing Q .

(2) Take the d -dimensional subspace $V_d = \text{span}\{v_1, \dots, v_d\} \subset V$, which exists according to the Assumption A, and let π_{V_d} be the projection on V_d . For any $z \in Q$ we have

$$(3.1) \quad |z - \pi_{V_d} z|_2 \leq C d^{-\beta} |z|_2 \leq \epsilon_1,$$

where the second inequality holds due to our choice of d and since $|z|_2 \leq R$.

(3) Pick at random a subspace $W \subset V_d$ of dimension n , and denote by e_1, \dots, e_n an orthonormal basis of it, and let $\pi_W : V_d \rightarrow W$ be the orthogonal projection of V_d on W .

(4) Put $\hat{\pi} = \sqrt{\frac{d}{n}} \pi_W$, and fix the set of measurements, associated with W , as the set of linear functionals

$$\lambda_i(z) = \langle \hat{\pi} z, e_i \rangle, \quad i = 1, \dots, n,$$

for each $z \in V_d$.

(5) Let $M(\epsilon_1, Q)$ be a minimal ϵ_1 -cover of Q . Denote by $\{x_j\}_{j=1}^M \subset Q$ the centers of the covering ϵ_1 -balls. Denote the points $y_j = \hat{\pi} x_j \in W$, $j = 1, \dots, M$, and denote $y = \hat{\pi} x$.

$$\begin{array}{c} \{x\}, \{x_j\}_{j=1}^M \subset Q \subset V \\ \downarrow \pi_{V_d} \text{ Assumption A} \\ \{\pi_{V_d} x, \pi_{V_d} x_1, \dots, \pi_{V_d} x_M\} = Z \subset V_d \\ \downarrow \hat{\pi} \text{ JLL} \\ \{y\}, \{y_j\}_{j=1}^M \subset W \end{array}$$

Notice that for any $z \in Q$ we have

$$\hat{\pi} \circ \pi_{V_d} z = \sqrt{\frac{d}{n}} \pi_W \circ \pi_{V_d} z = \sqrt{\frac{d}{n}} \pi_W z = \hat{\pi} z,$$

since a composition of orthogonal projections on the larger and on the smaller subspaces coincides with the projection on the smaller one.

II. *Reconstruction of x .*

(1) For the unknown vector $x \in Q$ perform the measurements $\lambda_i(x)$, $i = 1, \dots, n$.

(2) Let $y = \hat{\pi}x \in W$. The coordinates of y in W are exactly the measurements $\lambda_i(x)$, so y is known.

(3) Take one of the points y_j , $j = 1, \dots, M$, say, y_{j_1} in the ball in W of radius $2\epsilon_1$ around y . As we shall prove below, such points exist.

(4) We take $x_{j_1} \in Q$ as the desired ϵ -approximation of x .

Completion of the proof of Theorem 1.4. It remains to show that with probability greater than p , two assertions hold:

(a) Points y_j exist in the ball in W of radius $2\epsilon_1$ around y .

(b) For each one of them, say, for y_{j_1} , we have $|x - x_{j_1}|_2 \leq \epsilon$.

Consider the point $\pi_{V_d}x \in V_d$. We have $y = \hat{\pi}x = \hat{\pi} \circ \pi_{V_d}x \in W$, by the same reasoning as above. Now, define $Z \subset V_d$ as

$$Z = \{\pi_{V_d}x, \pi_{V_d}x_1, \dots, \pi_{V_d}x_M\}.$$

So $Z \subset V_d$ is a finite set of cardinality $m = |Z| = M + 1$. Let us recall that $\pi_{V_d}x$, as well as x , is unknown, while $\pi_{V_d}x_j$'s are known points. Now by Theorem 2.1 with probability p for any $u, v \in Z$ we have

$$(3.2) \quad \frac{1}{2}|u - v|_2 \leq |\hat{\pi}u - \hat{\pi}v|_2 \leq 2|u - v|_2.$$

In addition, notice that there exists j_0 such that

$$(3.3) \quad |\pi_{V_d}x - \pi_{V_d}x_{j_0}|_2 \leq \epsilon_1.$$

Indeed, $\|\pi_{V_d}\| = 1$, and $x \in Q$ belongs to one of the covering balls, so the center of this ball x_{j_0} satisfies $|x - x_{j_0}|_2 \leq \epsilon_1$.

Thus, by the definition of y, y_{j_0} and the upper bound of (3.2) and (3.3) we have

$$|y - y_{j_0}|_2 = |\hat{\pi}x - \hat{\pi}x_{j_0}|_2 \leq 2|\pi_{V_d}x - \pi_{V_d}x_{j_0}|_2 \leq 2\epsilon_1.$$

This proves assertion (a) above, i.e., the existence of points y_j in the ball in W of radius $2\epsilon_1$ around y .

Let y_{j_1} be one of such points. It remains to show that $|x - x_{j_1}|_2 \leq \epsilon$. Indeed,

$$|x - x_{j_1}|_2 \leq |x - \pi_{V_d}x|_2 + |\pi_{V_d}x - \pi_{V_d}x_{j_1}|_2 + |\pi_{V_d}x_{j_1} - x_{j_1}|_2.$$

The first and third terms, by (3.1), do not exceed ϵ_1 . Now, since y_{j_1} is in the ball in W of radius $2\epsilon_1$ around y , we have $|y - y_{j_1}|_2 \leq 2\epsilon_1$. Applying the lower bound of (3.2), we conclude that

$$|\pi_{V_d}x - \pi_{V_d}x_{j_1}|_2 \leq 2|\hat{\pi}x - \hat{\pi}x_{j_1}|_2 = 2|y - y_{j_1}|_2 \leq 2 \cdot 2\epsilon_1 = 4\epsilon_1.$$

Finally, $|x - \pi_{V_d}x_{j_1}|_2 \leq 6\epsilon_1 = \epsilon$, which completes the proof of Theorem 1.4. \square

Remark 3.1. Let us assume that our measurements $\lambda_v(x) = \langle x, v \rangle$ are normalized, i.e., $|v|_2 = 1$, and have a finite accuracy $\delta > 0$. In order to keep the reconstruction accuracy ϵ we must (according to our proof of Theorem 1.4 given above) assume that $\delta \leq \frac{\epsilon}{\sqrt{d}}$, where d is the dimension of the ambient space. This is because the linear mapping $\hat{\pi}$ providing the dimensionality reduction in the JLL, Theorem 2.1, is an orthogonal projection amplified by, essentially, \sqrt{d} . Hence this is the expected amplification factor of the measurement noise. The dimension $d = (\frac{1}{\epsilon})^{\frac{1}{\beta}}$ in typical examples is of order $\frac{1}{\epsilon}$. Altogether we have to require $\delta \asymp \epsilon^{\frac{3}{2}}$, where “ \asymp ” denotes

equivalence up to constants, not depending on ϵ . Whether such a high accuracy of the samples is indeed necessary is one of the open problems in our approach. Another open problem concerns the true necessity of the random choice of the measurements, which seems to be unavoidable in our specific algorithm above.

Remark 3.2. We do not provide in this paper lower bounds for the sampling entropy. Simple information considerations show that $N(\epsilon, Q)$ cannot be much smaller than $H(\epsilon, Q)$. Indeed, let us estimate the maximal amount of information provided by one measurement with the range $[-1, 1]$ and a possible error of order δ . With this accuracy we can surely identify only the measurement values quantified up to 2δ , and this gives us $\log(\frac{1}{\delta})$ binary bits of information. Thus, N measurements provide at most $N \log(\frac{1}{\delta})$ bits. But to specify any given $x \in Q$ up to accuracy ϵ we need exactly $H(\epsilon, Q)$ bits, by the definition of the Kolmogorov entropy, no matter what the source of the information. Therefore, independently of the specific reconstruction method we apply, we must have $N \log(\frac{1}{\delta}) \geq H(\epsilon, Q)$, or

$$(3.4) \quad N(\epsilon, Q) \geq H(\epsilon, Q) \log^{-1} \left(\frac{1}{\delta} \right).$$

Under the assumption $\delta \asymp \epsilon^{\frac{3}{2}}$, as in Remark 3.1, we get $N(\epsilon, Q) \geq \frac{2}{3} H(\epsilon, Q) \log^{-1}(\frac{1}{\epsilon})$.

Remark 3.3. For exact measurements, in a fixed finite-dimensional setting, accurate bounds for the sampling entropy are provided in [30] via the “mean width” of Q . It would be interesting to compare the results of the present paper and of [30] in a situation where the dimension of the ambient space grows as the required accuracy bound ϵ tends to zero.

4. Main examples. In this section we state some known and prove some new results on the Kolmogorov entropy of important functional classes. Through Theorem 1.4 we obtain estimates for the corresponding sampling entropy, comparing them with the known values, when available.

4.1. Classical regularity spaces. The following result has been established in [23, 24, 25, 33].

THEOREM 4.1. *Let $C^k(n, K) \subset C_0(I^n)$ be a (relatively compact) subset of $C_0(I^n)$ consisting of all C^k -smooth functions on I^n with all the derivatives up to order k bounded by K . Then*

$$H(\epsilon, C^k(n, K)) \asymp \left(\frac{1}{\epsilon} \right)^{n/k}.$$

Moreover, let $A(\eta, K) \subset C_0([- \pi, \pi])$ consist of periodic real analytic functions on $[- \pi, \pi]$, extendable into a complex η -neighborhood of $[- \pi, \pi]$ and uniformly bounded there by K . Then

$$H(\epsilon, A(\eta, K)) \asymp \log^2 \left(\frac{1}{\epsilon} \right).$$

Kolmogorov’s results have been obtained for the C_0 ambient metric. However, for the classes $C^k(n, K)$ and $A(\eta, K)$ consisting of functions with uniformly bounded derivatives, the L^2 and C_0 norms are equivalent, so the results remain true also in the L^2 ambient norm.

Now the following corollary of our general Theorem 1.4 recovers, in principle, the classical sampling accuracy for smooth functions. For analytic ones it gives a

less sharp result: indeed, sampling with the first $N \asymp \log(\frac{1}{\epsilon})$ Fourier coefficients reconstructs analytic function with an accuracy ϵ .

COROLLARY 4.2.

$$N(\epsilon, p, C^k(n, K)) \leq \frac{C(n, k, K)}{1-p} \left(\frac{1}{\epsilon}\right)^{n/k}, \quad N(\epsilon, p, A(\eta, K)) \leq \frac{C_1(\eta, K)}{1-p} \log^2\left(\frac{1}{\epsilon}\right).$$

4.2. Piecewise smooth functions. In this section we show that the metric entropy of piecewise smooth functions is essentially the same as for their smooth counterparts. This is intuitively clear in one variable: in addition to the “smooth information” we have to specify the positions of the jumps, which requires a negligible amount of additional information. In several variables we have to be more careful: the complexity of the boundaries of the smooth regions can significantly contribute to the overall entropy.

To avoid “boundary effects,” we shall consider piecewise C^k -smooth functions $f(x_1, \dots, x_n)$ defined on the torus T^n which we identify with the n -dimensional box $I^n = [-\pi, \pi]^n$ with the opposite faces glued together.

Assumption B. We shall assume that the jumps of f occur at certain C^k -smooth, pairwise disjoint, compact hypersurfaces $\Sigma_j \subset T^n$, $j = 1, \dots, K$. The complement in T^n of $\Sigma = \cup_{j=1}^K \Sigma_j$ is a union of certain n -dimensional manifolds D_s whose boundaries ∂D_s are unions of some of Σ_j . So we assume that $f = f_s$ on each D_s , with f_s being C^k -smooth function on D_s with all its derivatives up to order k uniformly bounded by K_2 in the corresponding domains D_s . This assumption implies the upper bound $2K_2$ for the jumps of all the derivatives of f along Σ . In addition, let ρ_1 denote the minimal distance between the jump submanifolds Σ_j . Next, denote for each point $x \in \Sigma$ by $n(x)$ the unit normal vector to Σ at x . Consider a new orthogonal coordinate system (y_1, \dots, y_n) at x such that the axis Oy_n is parallel to $n(x)$. Locally the hypersurface Σ near x can be represented by the equation $y_n = \eta(y_1, \dots, y_{n-1})$, with η a C^k -smooth function. We shall assume that for each point $x \in \Sigma$ this representation is valid in a ball B of a radius at least ρ_2 in the coordinates y_1, \dots, y_{n-1} , with all the derivatives of η up to order k uniformly bounded by K_1 in B . The constants ρ_2 and K_1 are closely related to the injectivity radius of Σ , and they can be estimated through the upper bound on its curvature.

DEFINITION 4.3. *The class $PC^k(n) = PC^k(n, \rho_1, \rho_2, K_1, K_2) \subset L^2(T^n)$ consists of all the piecewise C^k -smooth functions on the torus T^n , satisfying Assumptions B with the parameters ρ_1, ρ_2, K_1, K_2 .*

THEOREM 4.4. *The Kolmogorov ϵ -entropy of $PC^k(n) \subset L^2(T^n)$ satisfies*

$$H(\epsilon, PC^k(n)) \asymp \left(\frac{1}{\epsilon}\right)^m,$$

where $m = \frac{1}{k}$ for $n = 1$ and $m = \frac{2(n-1)}{k}$ for $n \geq 2$.

Proof. We shall construct explicitly an ϵ -net in the set $PC^k(n) \subset L^2(T^n)$ containing approximately minimal number of elements.

Let us start with dimension $n = 1$. Assumption B implies the uniform bound s on the number of the jump points of $f \in PC^k(1)$. First we quantize, up to accuracy ϵ^2 , the positions of these jump points. The number ϵ^2 appears here, instead of ϵ , since a shift of the jump points to a distance of order h produces an L^2 norm of order \sqrt{h} for the difference of the corresponding piecewise smooth functions. There are at most $(\frac{2\pi}{\epsilon^2})^s$ quantized jump configurations. For each configuration we construct an ϵ -net of

C^k functions on each continuity segment Δ_j . Taking all the jump configurations, and picking independently elements of ϵ -nets on each continuity segment Δ_j , we construct an ϵ -net in $L^2([-\pi, \pi])$ for the entire set $PC^k(1)$. The number of the elements in this net is the product of the number of quantized jump configurations and of the number of the elements in the nets for each of the continuity segments. After taking the logarithm we get, by the Kolmogorov result, that $H(\epsilon, PC^k(1)) \asymp (\frac{1}{\epsilon})^{\frac{1}{k}}$. Indeed, the contribution of the logarithm of the number of quantized jump configurations is at most $s \log(\frac{2\pi}{\epsilon})$, which is negligible with respect to the main asymptotic rate.

In several variables, we repeat the same construction. The jump manifolds Σ are, by Assumption B, C^k smooth $(n-1)$ -dimensional submanifolds in T^n , with uniformly bounded derivatives up to order k (in appropriate coordinate charts η). By Kolmogorov's result in dimension $n-1$, after quantization with the step ϵ^2 and taking logarithm, we get $\log M_0 \asymp (\frac{1}{\epsilon})^{\frac{2(n-1)}{k}}$ for the number M_0 of the quantized jump configurations. For the number M_j of the elements in the ϵ -net on each continuity domain D_j we have $\log M_j \asymp (\frac{1}{\epsilon})^{n/k}$. But now, in contrast with the one-dimensional case, we have $\frac{2(n-1)}{k} > \frac{n}{k}$ for $n \geq 3$. So the asymptotic behavior is determined by $m = \max(\frac{2(n-1)}{k}, \frac{n}{k})$, which is equal to $\frac{n}{k}$ for $n = 1$, and to $\frac{2(n-1)}{k}$ for $n \geq 2$. This completes the proof of Theorem 4.4. \square

By Theorem 1.4 we conclude that for piecewise smooth functions in one and two variables the sampling entropy is the same as for their smooth counterparts, while starting with three variables it becomes strictly larger. This is an artifact of the L^2 norm: it requires double accuracy in positioning of the jumps. In an appropriate pseudonorm, combining the C^0 -distance of smooth pieces and the C^0 -distance of the jump submanifolds, the Kolmogorov ϵ -entropy, and hence the sampling ϵ -entropy, remain of order $(\frac{1}{\epsilon})^{n/k}$. Presumably, the same happens if we consider piecewise regular functions in Besov classes, whose entropy and nonlinear width have been studied, in particular, in [14].

This fact was conjectured in various forms in many publications in the field. Let us consider in somewhat more detail one specific setting of this conjecture.

Periodic C^k functions on $[-\pi, \pi]$ are approximated by their partial Fourier sums up to the frequency N with the accuracy of order N^{-k-1} . For real analytic functions the approximation accuracy is of order q^{-N} , $q < 1$. From this fact it follows that the deterministic sampling ϵ -entropy of these classes is of order $(\frac{1}{\epsilon})^{\frac{1}{k+1}}$ and $\log(\frac{1}{\epsilon})$, respectively. However, if discontinuity points appear, this rate drops to $\frac{1}{N}$, no matter what the signal regularity between the jumps. It can be shown that if the position of the jumps is free and not known in advance, any linear method cannot provide better accuracy than $\frac{1}{N}$ (see, for example, [16]).

Fourier reconstruction of piecewise smooth functions is important in many theoretical and applied problems, from numerical solution of nonlinear PDEs developing shock waves to medical imaging. Many recent publications have been devoted to various methods of such reconstruction (see [1, 5, 17, 18, 19, 26] and references therein for a very partial sample). While it is known that the "smooth" sampling rate can be achieved inside the continuity intervals, the problem of accurate reconstruction of the positions of the jump points (and the signal nearby) turns out to be more difficult. See, however, [27], where piecewise smooth functions which are continuous but have discontinuous derivatives are reconstructed from pointwise samples with full accuracy.

It was conjectured by Eckhoff and other authors ([17, 26] and references therein) that the "smooth" accuracy rate N^{-k-1} of Fourier reconstruction of univariate C^k functions from their first N Fourier coefficients can be achieved also in a piecewise

smooth case (including the positions of the jumps) if we allow nonlinear processing of the Fourier data. In spite of many attempts, this conjecture remained largely open. In [5] we have shown that a modification of the original Eckhoff's method (based on algebraic reconstruction of the positions and amplitudes of the jumps) produces half the accuracy, $N^{-\frac{k}{2}}$. Recently Eckhoff's conjecture was finally settled in [4], by a further significant improvement of the original Eckhoff's method. For the deterministic sampling ϵ -entropy of piecewise C^k -smooth functions this result gives the "smooth order" $(\frac{1}{\epsilon})^{\frac{1}{k+1}}$. However, the analytic version of this problem remains open. So, the result of the next section is apparently a first confirmation of the Eckhoff-type conjecture for univariate piecewise analytic functions.

4.3. Piecewise analytic functions. To simplify the presentation, we restrict ourselves to the case of one variable. A similar result remains true in a multidimensional situation.

THEOREM 4.5. *Let $PA = PA(\kappa, \eta, K) \subset L^2([- \pi, \pi])$ be the set of piecewise real analytic functions on $[- \pi, \pi]$ with at most κ discontinuity points, such that each analytic piece can be extended into a complex η -neighborhood of the corresponding real segment and is bounded there by K . Then*

$$H(\epsilon, PA) \asymp \log^2 \left(\frac{1}{\epsilon} \right).$$

Proof. The proof is exactly as for Theorem 4.4, one-dimensional case. \square

By Theorem 1.4 we conclude that for piecewise analytic functions in one variable the sampling entropy is of order $\log^2(\frac{1}{\epsilon})$. It is an open question whether this bound can be improved to the classical order of $\log(\frac{1}{\epsilon})$.

4.4. Finite-dimensional warping. Let a finite-dimensional family of transformations $\Psi_\tau : [- \pi, \pi] \rightarrow [- \pi, \pi]$ be given, with the parameter $\tau = (\tau_1, \dots, \tau_s) \in I^s = [0, 1]^s$. We shall assume that $\Psi_\tau(x)$ is Lipschitzian, both in x and in the parameter τ .

Let a relatively compact $Q \subset L^2([- \pi, \pi])$ be given, with all $f \in Q$ being Lipschitzian, with the same Lipschitz constant. Define $Q \circ \Psi := \{g = f \circ \Psi_\tau, f \in Q, \tau \in I^s\}$. Since Ψ_τ are Lipschitzian, $Q \circ \Psi$ is also a relatively compact subset in $L^2([- \pi, \pi])$.

THEOREM 4.6. *For each $\epsilon > 0$ we have $H(\epsilon, Q \circ \Psi) \leq H(\epsilon, Q) + C \log(\frac{1}{\epsilon})$. If Ψ_τ are homeomorphisms of $[- \pi, \pi]$, then $H(\epsilon, Q \circ \Psi) \leq H(\epsilon, Q)$.*

Proof. This is a very special case of the results proved in [23, 25, 31] on the metric entropy of composition classes. Notice that in general $H(\epsilon, Q \circ \Psi)$ may be strictly smaller than $H(\epsilon, Q)$. This happens, in particular, if Ψ maps the entire interval $[- \pi, \pi]$ into one point. \square

In particular, let C^k, A denote the classes of C^k -smooth and real analytic functions on $[- \pi, \pi]$ (respectively).

COROLLARY 4.7. *For Ψ_τ a finite-dimensional family of homeomorphisms we have*

$$H(\epsilon, C^k \circ \Psi) \asymp \left(\frac{1}{\epsilon} \right)^{\frac{1}{k}}, \quad H(\epsilon, A \circ \Psi) \asymp \log^2 \left(\frac{1}{\epsilon} \right).$$

By Theorems 4.6 and 1.4 we conclude that the sampling entropy of $C^k \circ \Psi, A \circ \Psi$ remains the same as for their regular counterparts, while the regularity may be completely lost.

5. Some questions and comments.

1. Adding finite-dimensional linear combinations. It is easy to show that adding to the functional classes as the above linear combinations of a finite number of fixed Lipschitzian functions, we do not change the order of their Kolmogorov entropy. By Theorem 1.4 the sampling complexity remains the same as for their regular counterparts, while the regularity may be completely lost.

2. Is it possible to take into account a “regular” geometry of the ϵ -nets that appear in the functional classes discussed in this paper, in particular, for analytic functions? It is well known that the dimension bound in the JLL is sharp, up to a factor $\log(\frac{1}{\sigma})$, where σ is the almost-isometry distortion allowed (see [3]). In our applications $\sigma = \frac{1}{2}$ is fixed, but we encounter a gap of order $\log(\frac{1}{\epsilon})$ in the projection dimension for our ϵ -nets, in particular, in the case of analytic functions. Are these two effects related? Compare also bound (3.4) above for finite accuracy measurements.

3. What is the true accuracy rate in sampling of piecewise analytic functions? It is known that inside the continuity intervals the classical reconstruction rates from Fourier coefficients can be recovered [18]. For piecewise smooth functions the same gap in the upper and lower bounds appears as in the piecewise analytic case. However, it does not change the asymptotic recovery rate, because of a much higher complexity of piecewise smooth functions.

4. What is the metric entropy of piecewise regular functions, considered in Besov spaces?

5. What is the “true” order of the allowed measurements errors in the results above? As mentioned in Remark 3.1, we can modify the proofs above to include measurements of a finite accuracy, but the order of the allowed error is $\epsilon^{\frac{3}{2}}$, which may be too high.

6. We showed that under certain conditions the JLL implies a bound on sampling entropy of Q in V (where V is a Hilbert space). What can we say about the contraposition of this statement? For example, assume that for a certain Q in L_p (for $p \neq 2$) we have an optimal sampling entropy estimate (say, using algebraic methods). Assume in addition that the JLL holds for L_p (see the paper of Johnson and Naor [21] on this subject), and apply our method on Q using this JLL (JLL is just a plug-in to the algorithm). Now, if we get a better result for the sampling entropy of Q using this method, then it contradicts one of the assumptions; in other words, it contradicts the existence of JLL on non-Hilbert spaces.

Acknowledgments. The authors would like to thank L. Demanet, R. DeVore, A. Melnikov, G. Schechtman, and R. Vershinin for useful discussions, suggestions, and references, which, in particular, allowed us to correct some steps in the proof and to significantly improve the presentation.

REFERENCES

- [1] B. ADCOCK AND A. C. HANSEN, *Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon*, Appl. Comput. Harmon. Anal., 32 (2012), pp. 357–388.
- [2] N. AILON AND B. CHAZELLE, *Faster dimension reduction*, Commun. ACM, 53 (2010), pp. 97–104.
- [3] N. ALON, *Problems and results in extremal combinatorics*, Discrete Math., 273 (2003), pp. 31–53.
- [4] D. BATENKOV, *Complete algebraic reconstruction of piecewise-smooth functions from Fourier data*, Math. Comp., arXiv:1211.0680, 2012.
- [5] D. BATENKOV AND Y. YOMDIN, *Algebraic Fourier reconstruction of piecewise smooth functions*, Math. Comp., 81 (2012), pp. 277–318.

- [6] P. BINEV, A. COHEN, W. DAHMEN, R. DEVORE, G. PETROVA, AND P. WOJTASZCZYK, *Covergence rates for greedy algorithms in reduced basis method*, SIAM J. Math. Anal., 43 (2011), pp. 1457–1472.
- [7] T. BLU, P. L. DRAGOTTI, M. VETTERLI, P. MARZILIANO, AND L. COULOUT, *Sparse sampling of signal innovations*, IEEE Signal Process. Mag., 25 (2008), pp. 31–40.
- [8] R. BARANIUK, V. CEVHER, M. F. DUARTE, AND C. HEGDE, *Model-based compressive sensing*, IEEE Trans. Inform. Theory, 56 (2010), pp. 1982–2001.
- [9] R. BARANIUK, M. DAVENPORT, R. DEVORE, AND M. WAKIN, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx., 28 (2008), pp. 253–263.
- [10] E. J. CANDÈS, *An introduction to compressive sensing*, IEEE Signal Process. Mag., 25 (2008), pp. 21–30.
- [11] B. CARL, *Entropy numbers, s -numbers, and eigenvalue problems*, J. Funct. Anal., 41 (1981), pp. 290–306.
- [12] A. COHEN, W. DAHMEN, AND R. A. DEVORE, *Compressed sensing and best k -term approximation*, J. Amer. Math. Soc., 22 (2009), pp. 211–231.
- [13] R. A. DEVORE, R. HOWARD, AND C. MICHELLI, *Optimal nonlinear approximation*, Manuscripta Math., 63 (1989), pp. 469–478.
- [14] R. A. DEVORE, G. KYRIAZIS, D. LEVIATAN, AND V. M. TIKHOMIROV, *Wavelet compression and nonlinear n -widths*, Adv. Comput. Math., 1 (1993), pp. 197–214.
- [15] M. F. DUARTE AND Y. C. ELДАР, *Structured compressed sensing: From theory to applications* IEEE Trans. Signal Process., 59 (2011), pp. 4053–4085.
- [16] B. ETTINGER, N. SARIG, AND Y. YOMDIN, *Linear versus non-linear acquisition of step-functions*, J. Geom. Anal., 18 (2008), pp. 369–399.
- [17] K. S. ECKHOFF, *On a high order numerical method for functions with singularities*, Math. Comput., 67 (1998), pp. 1063–1088.
- [18] A. GELB AND E. TADMOR, *Detection of edges in spectral data*, Appl. Comput. Harmon. Anal., 7 (1999), pp. 101–135.
- [19] T. HRÝČAK AND K. GROCHENIG, *Pseudospectral Fourier reconstruction with the modified inverse polynomial reconstruction method*, J. Comput. Phys., 229 (2010), pp. 933–946.
- [20] W. B. JOHNSON AND J. LINDENSTRAUSS, *Extensions of Lipschitz mappings into a Hilbert space*, Contemp. Math, 26 (1984), pp. 189–206.
- [21] W. B. JOHNSON AND A. NAOR, *The Johnson-Lindenstrauss lemma almost characterizes Hilbert space, but not quite*, Discrete Comput. Geom., 43 (2010), pp. 542–553.
- [22] B. S. KASHIN AND V. N. TEMLYAKOV, *A remark on compressed sensing*, Math. Notes, 82 (2007), pp. 748–755.
- [23] A. N. KOLMOGOROV, *Estimates of the minimal number of elements of ε -nets in various functional spaces and their application to the problem of the representation of functions of several variables with superpositions of functions of lesser number of variables*, Uspekhi Mat. Nauk, 10 (1955), pp. 192–193.
- [24] A. N. KOLMOGOROV, *Asymptotic characteristics of certain totally bounded metric spaces*, Dokla. Akad. Nauk SSSR, 108 (1956), pp. 585–589.
- [25] A. N. KOLMOGOROV AND V. M. TIHOMIROV, *e -entropy and e -capacity of sets in functional space*, Amer. Math. Soc. Transl., 17 (1961) pp. 277–364.
- [26] G. KVERNADZE, *Approximation of the discontinuities of a function by its classical orthogonal polynomial Fourier coefficients*, Math. Comp, 79 (2010), pp. 2265–2285.
- [27] Y. LIPMAN AND D. LEVIN, *Approximating piecewise-smooth functions*, IMA J. Numer. Anal., 30 (2010), pp. 1159–1183.
- [28] T. PETER AND G. PLONKA, *A generalized Prony method for reconstruction of sparse sums of eigenfunctions of linear operators*, Inverse Proble., 29 (2013), 025001.
- [29] V. N. TEMLYAKOV, *Nonlinear Kolmogorov widths*, Math. Notes, 63 (1998), pp. 785–795.
- [30] R. VERSHYNIN, *Sampling and high-dimensional convex geometry*, in Proceedings of SampTA 2013, Bremen, Germany, 2013.
- [31] A. G. VITUSHKIN, *On representation of functions by means of superpositions and related topics*, Enseign. Math. (2), 23 (1977), pp. 255–320.
- [32] Y. YOMDIN, *Semialgebraic complexity of functions*, J. Complexity, 21 (2005), pp. 111–148.
- [33] V. ZAKHARYUTA, *On asymptotics of entropy of a class of analytic functions*, Funct. Approx. Comment. Math., 44 (2011), pp. 307–315.