SHANNON ENTROPY ESTIMATION BASED ON HIGH-RATE QUANTIZATION THEORY

Mattias Nilsson and W. Bastiaan Kleijn

KTH (Royal Institute of Technology)
Department of Signals, Sensors and Systems
SE-100 44 Stockholm, Sweden
{mattias.nilsson, bastiaan.kleijn}@s3.kth.se

ABSTRACT

In this paper we present a method for estimation of the Shannon differential entropy that accounts for embedded manifolds and is based on high-rate quantization theory. It forms an extension of the classical nearest-neighbor entropy estimator and provides simultaneously an explicit estimate of the manifold dimension. A clear advantage of our method over existing methods is that it estimates the more meaningful Shannon entropy rather than the Renyi entropy. Through experiments we confirm the power and the usefulness of our proposed scheme.

1. INTRODUCTION

Accurate estimation of information-theoretical entities such as the entropy and the mutual information is of great importance for the design and operation of applications in areas such as source coding and pattern recognition. In source coding the entropy of a variable provides bounds on the average code length needed for encoding, and in pattern recognition the shared information between features and classes provides insight into the expected classification performance.

The most common methods for entropy estimation are the so-called density *plug-in* estimators. These methods first estimate the underlying probability density function (pdf) or probability mass function (pmf) and then simply insert the density estimate into the information-theoretical expressions and perform the (multi-dimensional) integration. For applications where plug-in entropy estimators have been used see for instance [1, 2, 3]. It is well known that density estimation is a complex and delicate problem, which involves issues such as bin-width selection for histogram methods, and kernel-type and number of components for methods based on mixture models. These problems are avoided by entropy estimators that utilize the data directly for the entropy estimation without the intermediate step of density estimation. We mention in this category are the *m*-spacing estimator (based on order statistics) [4], its multidimensional extension [5], the Renyi[1] entropy estimator based on minimum spanning trees [6], and the *nearest-neighbor* (NN) entropy estimator [7]. For an overview of entropy estimators see [8].

It is common that random data vectors of high dimension are in fact located on a lower-dimensional manifold[2] embedded in the high-dimensional space. This has been observed in various areas such as for instance vision and speech [9]. The possible manifold structure of the data is often overlooked in entropy estimation, resulting in that classical methods, assuming the wrong intrinsic dimension (manifold dimension), are giving erroneous estimates of the entropy. Moreover, many of the *plug-in* methods are not capable of handling dirac functions (in the probability density function) that occur locally due to the manifold structure. Thus, generalizations of the classical entropy estimators that also handle manifolds are of interest.

[1]The Renyi entropy (or α-entropy) given a probability density function $f_X(x)$ is defined as $h_\alpha(X) = \frac{1}{1-\alpha}\log_2\left(\int_{\Omega_X} f_X^\alpha(x)\,dx\right)$ bits.

[2]Locally Euclidean topological space.

Recently, methods for estimating the Renyi-entropy of data located on manifolds have been presented [10, 11]. The method in [10] is based on the construction of so-called geodesic minimal spanning trees obtained from pruning of the ISOMAP produced by the algorithm in [9]. In [11] the same authors reduce the complexity of their previous method by using *k*-nearest neighbor graphs instead of the geodesic minimal spanning trees.

The motivation behind our work is that, in contrast to the methods in [10] and [11], we want an estimator that directly estimates the, by far more common, Shannon differential entropy of random data vectors located on a manifold, rather than the Renyi differential entropy. The Shannon differential entropy can then easily be used to find bounds (e.g., Shannon lower bound) on the average bit rate required for the encoding of a variable at particular distortion. Moreover, the estimator of [10] needs the output of [9], whereas our estimator of the manifold dimension and differential entropy is based on averaging of nearest-neighbor distances between random points in the logarithm domain [7]. This makes the structure of our differential entropy estimator clear, simple and easy to implement. Similar to [10] and [11], we simultaneously estimate both the manifold dimension and the differential entropy of the data located on the manifold.

The remainder of this paper is organized as follows. In section 2 we present our entropy estimator for data vectors located on manifolds. Section 3 shows some results for both toy and real examples, and section 4 is devoted to the conclusions.

2. ENTROPY ESTIMATION

In this work we are interested in estimating the differential entropy of a set of random data vectors that lie on a $\tilde{d}$-dimensional manifold embedded in $\mathbb{R}^d$ ($\tilde{d} \leq d$). Let X be a continuous stochastic variable (s.v.) in $\mathbb{R}^d$ with a pdf $f_X(x)$. Then the (Shannon) differential entropy of X is defined as

$$h(X) = -\int_{\mathbb{R}^d} f_X(x)\log_2\left(f_X(x)\right)dx. \tag{1}$$

In the following we assume that we have access to a set of N independent identically distributed (i.i.d.) vectors $\{x_n\}_{n=1}^N$.

2.1 Entropy estimator based on high-rate quantization theory

Our derivation of the nearest-neighbor based entropy estimator relies on high-rate theory for constrained resolution vector quantization (VQ). In the design of a constrained resolution VQ operating at high-rate we seek the distribution of quantization points (centroids), $g_C(x)$, that minimizes the total average distortion (per dimension), D_r, given a fixed number of centroids ($\int_{\mathbb{R}^d} g_X(x)\,dx = N$). The average distortion for a nearest-neighbor quantizer at a given power distortion r is defined as

$$D_r = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\tilde{d}}\left((x_n - x_{nn})^T(x_n - x_{nn})\right)^{\frac{r}{2}}$$
$$\equiv \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\tilde{d}}\parallel e_n \parallel_r, \qquad (2)$$

where x_{nn} denotes the nearest-neighbor quantization point to x_n. It is important to note that, in (2), we use the data vectors x both as observations and quantization points. Furthermore, using the centroid distribution $g_C(x)$, the average distortion can be approximated as [12]

$$D_r \approx C(r,\tilde{d},\mathscr{G}_{opt})\int_{\mathbb{R}^d} f_X(x) g_C(x)^{-\frac{r}{\tilde{d}}}\,dx, \qquad (3)$$

where $C(r,\tilde{d},\mathscr{G}_{opt})$ represents the coefficient of quantization, which, for a cell of volume V with optimal cell shape $\mathscr{G}_{opt}$, is defined as

$$C(r,\tilde{d},\mathscr{G}_{opt}) = \frac{1}{\tilde{d}}\frac{1}{V^{\frac{\tilde{d}+r}{\tilde{d}}}}\int_V \parallel e \parallel_r\, dx. \qquad (4)$$

Using variational calculus, the minimization of the average distortion in (3) under the constrained resolution constraint $\int_{\mathbb{R}^d} g_X(x)\,dx = N$ is straightforward, and the resulting optimal centroid distribution becomes [13]

$$g_C(x) = N\frac{f_X(x)^{\frac{\tilde{d}}{\tilde{d}+r}}}{\int_{\mathbb{R}^d} f_X(x)^{\frac{\tilde{d}}{\tilde{d}+r}}\,dx}, \qquad (5)$$

which when inserted into (3) yields

$$D_r \approx C(r,\tilde{d},\mathscr{G}_{opt})N^{-\frac{r}{\tilde{d}}}\left(\int_{\mathbb{R}^d} f_X(x)^{\frac{\tilde{d}}{\tilde{d}+r}}\,dx\right)^{\frac{\tilde{d}+r}{\tilde{d}}}. \qquad (6)$$

Taking the logarithm of both sides of (6) and rearranging the terms we identify the Renyi (α-) entropy to be

$$h_\alpha(X) = \frac{1}{1-\alpha}\log_2\left(\int_{\Omega_X} f_X^\alpha(x)\,dx\right) \qquad (7)$$
$$\approx \log_2(N) + \frac{1}{1-\alpha}\log_2\left(\frac{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\tilde{d}}\parallel e_n\parallel_r}{C(r,\tilde{d},\mathscr{G}_{opt})}\right),$$

where $\alpha = \tilde{d}/(\tilde{d}+r)$, and where we have inserted the average nearest-neighbor expression in (2) for the average distortion D_r. The Renyi entropy converges to the Shannon entropy as α goes to one, i.e., $h(X) = \lim_{\alpha\to 1} h_\alpha(X)$. Generally, we are interested in the Shannon entropy rather than the Renyi entropy and, therefore, we develop (7) under the assumption that the distortion power r is small relative to $\tilde{d}$ (which in turn implies that α is close to one). For small r, the logarithm of the average distortion can be approximated by average of the logarithm of the nearest-neighbor distances. When applied to (7) gives an expression of the Shannon differential entropy

$$h(X) \approx \tilde{d}\frac{1}{N}\sum_{n=1}^{N}\log_2(\parallel e_n\parallel_1)$$
$$+ \log_2(N) + \log_2\left(\left(\tilde{d}\cdot C(r,\tilde{d},\mathscr{G}_{opt})\right)^{-\frac{\tilde{d}}{r}}\right), \qquad (8)$$

where we twice have used that $\log(x)\approx x-1$ when x is close to one. The last term of equation (8) can be seen as compensation for the implicit hyper-cube quantizer cell shape that is built into the differential entropy[3] and the actual cell shape imposed by the quantizer used. In the following we will use random codebooks, i.e., codebooks (set of quantization points) generated by random selection of a number of points from the data set. Worth noting here is that, since the power-distortion r is very small, the optimal quantization point distribution for the constrained resolution quantizer coincides with the distribution of the data, i.e., $g_C(x)\approx N\cdot f_X(x)$. Inserting the coefficient of quantization for random codebooks (and high-rate), derived by Zador [12], into (8) results in

$$h(X) \approx \tilde{d}\frac{1}{N}\sum_{n=1}^{N}\log_2(\parallel e_n\parallel_1)$$
$$+ \log_2(N) + \log_2\left(\left(\Gamma(1+\frac{r}{\tilde{d}})\right)^{-\frac{\tilde{d}}{r}} V_{\tilde{d}}\right), \qquad (9)$$

where $\Gamma(\cdot)$ represents the gamma function, and where $V_{\tilde{d}}$ denotes the volume of a $\tilde{d}$-dimensional sphere of unit radius

$$V_{\tilde{d}} = \frac{\pi^{\tilde{d}/2}}{(\tilde{d}/2)!}. \qquad (10)$$

In our work we allow the manifold dimension $\tilde{d}$ to be fractional and therefore, since the factorial operation is only defined for integers, we use [14]

$$\log_2(V_{\tilde{d}}) = -\frac{\tilde{d}}{2}\log_2\left(\frac{\tilde{d}}{2\pi e}\right) - \frac{1}{2}\log_2(\tilde{d}\pi) - \varepsilon, \qquad (11)$$

where the error-term ε is bounded as $0 \le \varepsilon \le \frac{\log_2(e)}{6\tilde{d}}$.

Taking the limit of (9) as r goes to zero we arrive at a final expression for the Shannon differential entropy (in bits)

$$\hat{h}(X) = \tilde{d}\frac{1}{N}\sum_{n=1}^{N}\log_2(\parallel e_n\parallel_1) + \log_2(N)$$
$$+\frac{\gamma}{\log_e(2)} - \frac{\tilde{d}}{2}\log_2\left(\frac{\tilde{d}}{2\pi e}\right) - \frac{1}{2}\log_2(\tilde{d}\pi), \qquad (12)$$

where $\gamma \approx 0.5772$ is the Euler constant, and where we have set the ε-term in (11) to zero. The expression for the entropy estimator in (12) is similar to the estimator in [7] with the difference that (12) does not assume an integer dimension.

2.2 Joint estimation of manifold dimension and entropy

Estimation of the manifold dimension and the differential entropy can be done simultaneously using random codebooks. Starting with the final expression for the Shannon differential entropy (12) we note that the differential entropy itself can be seen as a constant that we can combine with the last three terms on the right-hand side of (12) into a new constant κ, i.e.,

$$\kappa = \frac{1}{\tilde{d}}\left(\hat{h}(X) - \frac{\gamma}{\log_e(2)} + \frac{\tilde{d}}{2}\log_2\left(\frac{\tilde{d}}{2\pi e}\right) + \frac{1}{2}\log_2(\tilde{d}\pi)\right). \qquad (13)$$

Equation 12 can then be rewritten as

$$\frac{1}{\bar{N}}\sum_{n=1}^{\bar{N}}\log_2\left(\parallel e_{n|M_k}\parallel_1\right) = -\frac{1}{\tilde{d}}\log_2(M_k) + \kappa, \qquad (14)$$

where $\bar{N}$ denotes the number of observations used for the averaging and $\parallel e_{n|M_k}\parallel$ represents the distance between a random observation x_n and its nearest neighbor in a random codebook of size M_k. The random codebook is formed by randomly selected vectors from the observation set. Using the average of the logarithm of nearest-neighbor distance for K different codebook sizes M_k,

[3]The expression $h(X)\approx H(X) + d\log_2(\Delta)$ relates the differential entropy of X to the entropy $H(X)$ through uniform scalar quantizers (dimension d). The approximation is only valid when the quantization stepsize Δ is small compared to the smoothness of the pdf of X.

($k=[1,...,K]$), we can easily find the least-squares estimate of the manifold dimension $\tilde{d}$ and the constant κ, i.e.,

$$\Theta = \begin{pmatrix} -\frac{1}{\tilde{d}} \\ \kappa \end{pmatrix} = \left(\bar{M}^T\bar{M}\right)^{-1}\bar{M}^T\bar{e}, \tag{15}$$

where,

$$\bar{M} = \begin{bmatrix} \log_2(M_1) & \dots & \log_2(M_K) \\ 1 & \dots & 1 \end{bmatrix}^T, \tag{16}$$

$$\begin{aligned}\bar{e} = \Bigg[& \frac{1}{\bar{N}}\sum_{n=1}^{\bar{N}} \log_2\left(\| e_{n|M_1} \|_1\right), \dots \\ & , \frac{1}{\bar{N}}\sum_{n=1}^{\bar{N}} \log_2\left(\| e_{n|M_K} \|_1\right)\Bigg]^T.\end{aligned} \tag{17}$$

Given the estimates of $\tilde{d}$ and κ the estimate of the differential entropy is easily obtained from (13).

3. EXPERIMENTS AND RESULTS

In this section we consider two different experiments where we apply our method for the simultaneous estimation of the manifold dimension and the Shannon differential entropy. The first experiment is purely a toy experiment, where both the true manifold dimension and the differential entropy are known. In the second, real-world, experiment we consider a problem that is relevant for the compression of speech. In particular we investigate the manifold dimensions and differential entropies of narrowband speech spectral envelope (of 20 ms non-overlapping segments of speech sampled at 7 kHz).

3.1 Toy experiment: Swiss roll

Consider that we have access to N random vectors in $\mathbb{R}^3$ that are located on a two-dimensional manifold embedded in this three-dimensional space. In our setup we generate the data points for the "Swiss roll" manifold by creating the coordinate vector $\bar{x} = [\varphi\cos(\varphi), \varphi\sin(\varphi), \rho]$, where φ and ρ are uniformly distributed random numbers within the intervals $[2\pi,\dots,4\pi]$ and $[0,\dots,10]$, respectively. The true differential entropy of the two-dimensional manifold is 9.22 bits (with two digits precision). Figure 1 depicts the structure of the Swiss roll manifold. For the experiment we

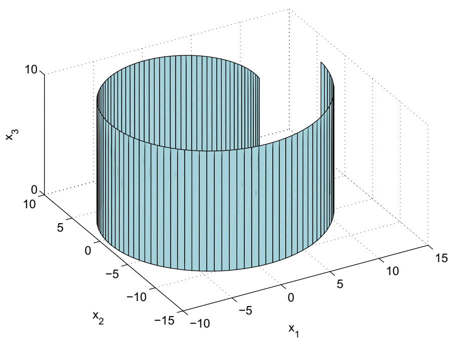

Figure 1: Shows the structure of the manifold used for the toy experiment. The true differential entropy for the uniformly distributed data over the manifold was calculated to be 9.22 bits.

set the number of available observations to $N=1000$, the number of nearest-neighbors for averaging $\bar{N}=250$, and $K=50$ random codebooks ranging from $M_1=2^7$ to $M_{50}=2^9$ in size. Figure 2 shows the average distortion $\frac{1}{\bar{N}}\sum_{n=1}^{\bar{N}}\log_2\left(\| e_{n|M_k} \|\right)$ as a function of the codebook size $\log_2(M_k)$ for one run of the Swiss roll experiment. The line in Figure 2 displays the least-squares fit using (15)

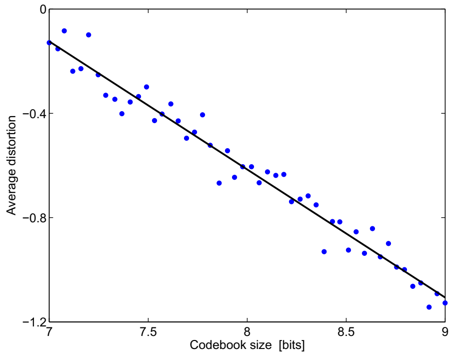

Figure 2: The points in the graph show the average distortion $\frac{1}{\bar{N}}\sum_{n=1}^{\bar{N}}\log_2\left(\| e_{n|M_k} \|\right)$ for each corresponding codebook size $\log_2(M_k)$, and the line shows the least-squares fit to these points. The slope of the line is inversely proportional to the manifold dimension (with a minus sign), and the differential entropy can be computed from the intersection of the line and the average distortion axis at the codebook size of zero bits.

resulting in $\tilde{d}=2.04$ as an estimate of the manifold dimension and $\hat{h}(X)=9.36$ bits for the differential entropy, with the true values being 2 and 9.22 bits, respectively. Repeating the experiment 100 times we get with 99% confidence that the manifold dimension is 1.96 ± 0.02 and the differential entropy is 9.38 ± 0.01. The bias that we observe in this example comes to a large extent from the fact that we have neglected the ε-term in (11). However, this effect becomes increasingly less prominent with increasing manifold dimensionality. In the example above, the ε-term is approximately equal to 0.12, which would then result in a differential entropy estimate of 9.26 ± 0.01 bits. Experiments with larger codebooks (than 9 bits) indicate that the remaining small bias in the estimates of manifold dimension and differential entropy can most likely be related to the validity of the high-rate assumption.

3.2 Real experiment: cepstral coefficients of speech

In this experiment we investigate the manifold dimensions and differential entropies of narrowband speech linear prediction coefficients (LPC) represented by cepstral coefficients (CC). One reasons for using the cepstral coefficients is that the squared error between cepstral vectors approximates the spectral distortion [15], which provides an indication of the distortion we perceive. We used speech from the NTT-AT database, which comprises speech from a wide range of different languages and speakers, for the generation of the data set. The speech was first bandpass filtered (pass-band between 300 and 3400 Hz) and then downsampled to a sampling frequency of 7 kHz. Tenth-order linear-prediction analysis was performed on hamming windowed 20 ms segments with no overlap. A simple voice activity detector was used to discard non-speech frames, and finally the LPCs were transformed into CCs. The total data set consisted of $2.75\cdot10^5$ vectors.

Let X denote the s.v. of narrow-band CCs. We then estimate the differential entropy $h(X)$ and manifold dimensionality from the data set. For the experiment we set the numbers of nearest-neighbors for the averaging to $\bar{N}=1000$, and the $K=15$ random codebooks we use ranges in size from $M_1=2^{12}$ to $M_{15}=2^{17}$. Figure 3 shows one typical run of the experiment, which clearly confirms the linear relationship assumed in (14). We repeat the experiment 10 times, which yields, with 95% confidence, an estimated manifold dimension of 8.12 ± 0.12 and differential entropy of 1.91 ± 0.26 bits. Thus, the narrowband speech CCs are located on an approximately 8-dimensional manifold embedded in the 10-dimensional Euclidean space. The differential entropy as such does not provide any insight into the lossy compression. To quantify the result we relate the differential entropy to a lower bound on the average bit rate for a given average distortion D. The Shannon lower bound (SLB) [16] is a lower bound on the minimum rate possible to encode a

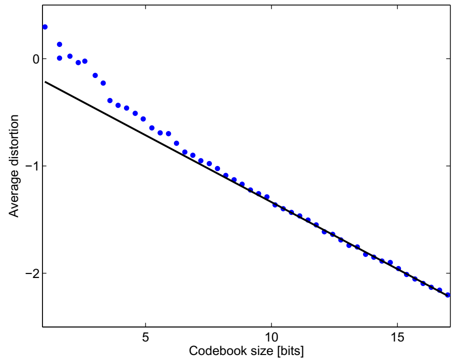

Figure 3: The points represent pairs of average distortion (with $\bar{N} = 1000$) and codebook size for the 10-dimensional cepstral coefficient vectors of narrowband speech. The line shows the least-squares fit to the points corresponding to codebook sizes of 12 bits and larger, which results in $\tilde{d} = 8.14$ as an estimate of the manifold dimension and $\hat{h}(X) = 1.90$ bits for the differential entropy.

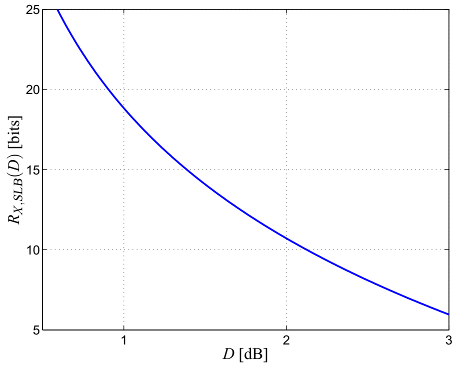

Figure 4: Shannon lower bound of the narrowband rate, in bits per frame, as a function of the root-mean-square log spectral distortion.

variable at a given distortion and can be defined as

$$R_{X,SLB}(D) = h(X) - \sup_{\{f_W(w): \int_{\Omega_W} f_W(w)\delta(w)dw \leq \left(\frac{D\log_e(10)}{10\sqrt{2}}\right)^2\}} h(W), \tag{18}$$

where $W = X - \hat{X}$ denotes the quantization noise of X ($\hat{X}$ denotes the reconstruction point), $\delta(w)$ is the single-letter squared error distortion, and both the rate and the distortion are defined on a per-frame basis. Since X is represented by cepstral coefficients, D approximates the root-mean-squared (RMS) log spectral distortion. The supremum of the differential entropy in (18) results in a multivariate Gaussian distribution (with a diagonal covariance matrix) of the quantization noise, and the Shannon lower bound becomes

$$R_{X,SLB}(D) = h(X) - \frac{\tilde{d}}{2}\log_2\left(\frac{2\pi e}{\tilde{d}}\left(\frac{D\log_e(10)}{10\sqrt{2}}\right)^2\right), \tag{19}$$

where $\tilde{d}$ is the manifold dimension of X. Inserting our estimates of the manifold dimension and differential entropy into (19) we can compute the lower bound on the minimum average bit rate required to achieve an average distortion D. This is displayed in Figure 4 for the typical distortion range of interest. Generally, it is considered that we get non-audible distortion if the RMS log-spectral distortion is less than 1 dB[4]. From Figure 4 we extract that on average at least 19 bits (per 20 ms block) are required to achieve 1 dB average distortion. This estimate seems reasonable considering that systems exist that reach this distortion level at a rates between 20-22 bits per block [17, 18].

4. CONCLUSIONS

In this work we first derived the nearest-neighbor entropy quantizer based on theory for high-rate constrained-resolution vector quantization. We have extended the entropy estimator to jointly estimate the dimensionality and differential entropy of random data located on embedded manifolds. In contrast to existing algorithms [10, 11] we estimate the more meaningful Shannon differential entropy instead of the Renyi entropy. In an experiment on spectrum quantization we showed a practical use of our method, where we found a lower bound on the minimum bit rate required for transparent quantization (non-audible distortions) equal to 19 bits. This estimate is consistent with bit rates for transparent quantization reported by other researchers [17, 18].

[4]Additional constraints on the maximum allowed percentage of outliers are not considered in this work.

REFERENCES

[1] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.

[2] J. A. Bilmes, "Maximum mutual information based reduction strategies for cross-correlation based joint distributional modeling," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, 1998, pp. 469–472.

[3] M. Nilsson, H. Gustafsson, S. V. Andersen, and W. B. Kleijn, "Gaussian mixture model based mutual information estimation between frequency bands in speech," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, 2002, pp. 525–528.

[4] O. Vasicek, "A test for normality based on sample entropy," *J. Royal Statistical Soc., Ser. B*, vol. 38, no. 1, pp. 54–59, 1976.

[5] E. Miller, "A new class of entropy estimators for multi-dimensional densities," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, 2003, pp. 297–300.

[6] A. O. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Proc. Magazine*, vol. 19, no. 5, pp. 85–95, 2002.

[7] L. F. Kozachenko and N. N. Leonenko, "Sample estimate of entropy of a random vector," *Problems of Information Transmission*, vol. 23, no. 1, pp. 95–101, 1987.

[8] J. Beirlant, E. Dudewicz, L. Györfi, and E. van der Meulen, "Nonparametric entropy estimation: an overview," *Intern. K. Math. Stat. Sci.*, vol. 6, no. 1, pp. 17–39, 1997.

[9] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 1, pp. 2319–2323, 2000.

[10] J. Costa and A. Hero, "Manifold learning with geodesic minimal spanning trees," *Preprint*, 2003.

[11] J. Costa and A. Hero, "Manifold learning using euclidean k-nearest neighbor graphs," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, 2004.

[12] P. L. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 139–149, 1982.

[13] R. M. Gray, *Source Coding Theory*, Kluwer Academic Publishers, 1990.

[14] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, 1993.

[15] A. H. Gray and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 24, no. 5, pp. 380–389, 1976.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[17] P. Hedelin, "Single stage spectral quantization at 20 bits," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, 1994, pp. 525–528.

[18] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. Speech and Audio Proc.*, vol. 8, no. 4, pp. 385–401, 2000.