



HAL
open science

Characterizing compressibility with Lorenz curves

Claudio Weidmann

► **To cite this version:**

Claudio Weidmann. Characterizing compressibility with Lorenz curves. IEEE International Symposium on Information Theory (ISIT 2011), Jul 2011, Saint-Petersburg, Russia. pp.2228-2232. hal-00688917

HAL Id: hal-00688917

<https://hal.science/hal-00688917>

Submitted on 18 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Characterizing compressibility with Lorenz curves

Claudio Weidmann

ETIS – CNRS UMR 8051 / ENSEA / Univ Cergy-Pontoise

95014 Cergy-Pontoise, France

e-mail: claudio.weidmann@ensea.fr

Abstract—This paper revisits earlier work on rate distortion behavior of sparse sources, namely it highlights the fact that a graphical sparsity characterization proposed in [1] is a *Lorenz curve*, a tool for summarizing income inequality that has been used by economists for over a century. The Lorenz curve associated to a memoryless source can be used to obtain upper bounds on the distortion rate function, thus characterizing source compressibility. It is shown that an order relation on Lorenz curves induces an analogous relation on distortion rate upper bounds. This can be used to characterize the compressibility of certain parametric families of source distributions, for which an order on the parameters induces an order on Lorenz curves.

I. INTRODUCTION

Sparse signal representations are at the core of state-of-the-art lossy source coding and compressed sensing methods. Sparse vectors, in which most components are zero, can be compressed and/or sampled more efficiently. While such strict sparsity is rather straightforward to characterize, several measures have been proposed for the more general case of approximate, non-strict sparsity; see the work by Hurley and Rickard [2], which compares a number of sparsity measures based on intuitive quantifiable criteria.

In this work, we use the term *compressibility* to denote non-strict sparsity, hinting at the ultimate goal, which is the lossy compression of a signal into a digital representation. We will take a probabilistic approach, modeling time-discrete signals as the outputs of memoryless random sources. Thus rate distortion theory is the appropriate tool to study the limits on achievable compression performance. The question that this work addresses is “how does compressibility of a source affect the minimal rate per sample needed to reconstruct it within a given distortion” (i.e. a lossy compression problem), and not “how many samples are needed to guarantee reconstruction” (which would be a compressed sensing type of question).

Section II introduces Lorenz curves and the associated Lorenz order as tools to display and compare the compressibility (sparsity) of sources. Section III recalls two upper bounds on the distortion rate function and shows that they can actually be obtained from a parametric curve that is equivalent to a Lorenz curve. Section IV then uses these findings to show that the Lorenz order induces an order on distortion rate bounds and hence allows to define a compressibility hierarchy in parametric families of source distributions.

II. LORENZ CURVES

In 1905, Max Lorenz proposed a simple graphical tool to display the inequality in the distribution of wealth among a

finite population [3]; this is now known as the Lorenz curve and is still popular among economists. For a population of size n , with wealths x_1, x_2, \dots, x_n arranged from smallest to largest (denoted by $x_{1:n}, x_{2:n}, \dots, x_{n:n}$), the Lorenz curve $L_x(u)$ is obtained by plotting the points

$$(u, L_x(u)) = \left(\frac{k}{n}, \frac{\sum_{i=1}^k x_{i:n}}{\sum_{i=1}^n x_{i:n}} \right), \quad j = 1, 2, \dots, n,$$

as well as the point $(0, 0)$, and interpolating in between. Thus $L(u)$ corresponds to the proportion of total wealth owned by the $u \cdot 100\%$ poorest individuals. Clearly, distribution inequality is larger if the curve is more bent away from the *egalitarian line* running from $(0, 0)$ to $(1, 1)$. By definition, the Lorenz curve allows comparisons among populations of different size. If two populations \mathbf{x} and \mathbf{y} have Lorenz curves L_x and L_y such that $L_x(u) \leq L_y(u), \forall u \in (0, 1)$, then \mathbf{x} shows at least as much inequality as \mathbf{y} . Many single-parameter measures of inequality have been proposed based on the Lorenz curve. Perhaps the most prominent is the *Gini coefficient* G , which equals twice the area between $L(u)$ and the egalitarian line. Thus $G = 0$ for a population in which everyone has the same wealth, and $G = 1$ for a population in which a single individual owns all wealth.

Coming back to signal representations, we may consider a vector of sample or coefficient magnitudes and say that it is sparser (in non-strict sense) if its Lorenz curve is below that of an alternative representation, since fewer samples will be needed to achieve the same reconstruction fidelity.¹ Hence the Gini coefficient can be used as a measure of sparsity; in fact, [2] found that it is one of the few sparsity measures that satisfies a range of intuitive properties that one should expect from such measures. However, the focus of this work is not on single-parameter sparsity measures, but on using the Lorenz curve as a tool to characterize the compressibility (non-strict sparsity) over the entire range of source coding rates. This motivates using the entire curve, rather than a single-parameter measure like the source entropy, which would yield only high-rate characterizations.

In the following, we will model the signal representation by a real-valued memoryless source X with density $f_X(x)$ and distribution function $F_X(x)$. We may compute the expected Lorenz curve for a block \mathbf{X} of samples generated by the source

¹Here we assume e.g. squared error distortion measure and nonlinear approximation by the k largest samples (from equal-sized vectors). Increased sparsity thus corresponds to better energy compaction.

X . By letting the block size grow to infinity, one arrives at the standard definition of the Lorenz curve of a random variable X [4, Sec. 17.C]. Let \mathcal{L} be the class of distributions of non-negative random variables with positive finite expectation (we will slightly abuse notation and use both $X \in \mathcal{L}$ and $F_X \in \mathcal{L}$). The Lorenz curve corresponding to the distribution F in \mathcal{L} is described by the set of points

$$\left\{ \left(F(x), F^{(1)}(x) \right) \right\}, \quad x \geq 0, \quad (1)$$

where $F^{(1)}$ is the first moment distribution

$$F^{(1)}(x) = \frac{\int_0^x z dF(z)}{\int_0^\infty z dF(z)}. \quad (2)$$

Defining the *right continuous inverse distribution function* or *quantile function* by

$$F^{-1}(y) = \sup\{x : F(x) \leq y\}, \quad (3)$$

one obtains the explicit expression for the Lorenz curve of a random variable X with distribution $F \in \mathcal{L}$,

$$L(u) = \frac{\int_0^u F^{-1}(y) dy}{\int_0^1 F^{-1}(y) dy} = \frac{\int_0^u F^{-1}(y) dy}{EX}, \quad 0 \leq u \leq 1. \quad (4)$$

Using definition (4) it is possible to show that $L(u)$ is a continuous, non-decreasing convex function on $[0, 1]$, which is differentiable almost everywhere, and with $L(0) = 0$ and $L(1) = 1$. The Lorenz curve is scale-free by definition, i.e. $L_{aX}(u) = L_X(u)$ for $a > 0$. The parametric representation (1) is often more convenient if no closed-form expression is available for the quantile function. Notice that a Lorenz curve can obviously be defined also for higher order moments in (2), but for technical reasons that will become clear later, we prefer to consider distributions of the powers $Y = X^r$ of the original random variable (allowing e.g. to write $X^r \in \mathcal{L}$ without having to explicitly define the class of distributions with finite r -th moment).

By considering the Lorenz curves associated to random variables it is possible to define a partial order in the class \mathcal{L} (see e.g. [4, Sec. 17.C]).

Definition 1 (Lorenz order): Let $F_X, F_Y \in \mathcal{L}$, with corresponding Lorenz curves L_X and L_Y . Then X is less than Y in the *Lorenz order*, denoted $X \leq_L Y$ or $F_X \leq_L F_Y$, if $L_X(u) \geq L_Y(u)$ for all $u \in [0, 1]$.

Thus $X \leq_L Y$ means that X is more equally distributed, or less compressible (sparse) than Y , that is, L_X is closer to the egalitarian line than L_Y .

III. DISTORTION RATE BOUNDS

This section briefly recalls two upper bounds on the mean-squared error (MSE) distortion rate function of continuous random variables, which first appeared in [5] and are discussed in more detail in [6]. Although from here on we focus on MSE, all results in this work can be extended to absolute r -th power distortion measures $d(x, \hat{x}) = |x - \hat{x}|^r$.

The (information) rate distortion function [7]

$$R(D) = \min_{f(\hat{x}|x): E d(X, \hat{X}) \leq D} I(X; \hat{X}) \quad (5)$$

is the minimal rate at which a source X can be encoded with distortion $E d(X, \hat{X}) \leq D$. Alternatively, it may be expressed as a distortion rate function $D(R)$. Two important properties for MSE distortion, $d(x, \hat{x}) = (x - \hat{x})^2$, are the *Gaussian upper bound*,

$$D(R) \leq \sigma^2 e^{-2R}, \quad (6)$$

and the *Shannon lower bound* (SLB),

$$R(D) \geq R_{\text{SLB}}(D) = h(X) - \frac{1}{2} \log(2\pi e D). \quad (7)$$

Note that in this work all rates are expressed in nats and all logarithms are natural, unless otherwise stated.

The following bounds are obtained by classifying the magnitudes of the source samples using a threshold t and applying the Gaussian upper bound (6) to each of the two classes. They are upper bounds on the operational rate distortion function of *magnitude classifying quantization* (MCQ), which sends the classification as side information and uses it to switch between two codebooks. The samples with magnitude above threshold t are called *significant* and are characterized by two incomplete moments, namely the probability

$$\mu(t) = \int_{-\infty}^{-t} f(x) dx + \int_t^{\infty} f(x) dx \quad (8)$$

and the second moment

$$A(t) = \int_{-\infty}^{-t} x^2 f(x) dx + \int_t^{\infty} x^2 f(x) dx, \quad (9)$$

where $A(0) = \sigma^2$ is the source variance (we assume $EX = 0$ without loss of generality). From these we compute the conditional second moment of the significant samples,

$$\sigma_1^2(t) = E[X^2 | |X| \geq t] = \frac{A(t)}{\mu(t)},$$

as well as that of the *insignificant* samples,

$$\sigma_0^2(t) = E[X^2 | |X| < t] = \frac{\sigma^2 - A(t)}{1 - \mu(t)}.$$

The classification decision is sent as side information to the decoder, using $h_b(\mu)$ nats per sample ($h_b(p) = -p \log p - (1-p) \log(1-p)$ is the binary entropy function). The encoder can now use two separate Gaussian codebooks, one for the insignificant samples with rate R_0 and one for the significant samples with rate R_1 . The average rate per sample becomes

$$R = h_b(\mu(t)) + \mu(t)R_1 + (1 - \mu(t))R_0. \quad (10)$$

By standard rate allocation (reverse water-filling) over the two codebooks we obtain an upper bound.

Theorem 1: (High-Rate Upper Bound) For all

$$R \geq R_{\min}(t) = h_b(\mu(t)) + \frac{1}{2} \mu(t) \log \frac{\sigma_1^2(t)}{\sigma_0^2(t)}, \quad (11)$$

the MSE distortion rate function of a memoryless source is upper-bounded by

$$D(R) \leq B_{hr}(t, R) = c(t) \sigma^2 e^{-2R}, \quad (12)$$

where

$$c(t) = \exp\left(2h_b(\mu(t)) + (1 - \mu(t)) \log \frac{\sigma_0^2(t)}{\sigma^2} + \mu(t) \log \frac{\sigma_1^2(t)}{\sigma^2}\right). \quad (13)$$

The best asymptotic upper bound is obtained by finding the $t^* \geq 0$ that minimizes $c(t)$. Since $\lim_{t \rightarrow 0^+} c(t) = 1$, the Gaussian upper bound is always a member of this family.

Exploiting the trivial fact that (12) also upper bounds the Shannon lower bound, we obtain

Corollary 2: Let $\mu^* = \mu(t^*)$ and $A^* = A(t^*)$ yield the tightest bound in Theorem 1. Define the pmf's

$$\boldsymbol{\mu}^* = [\mu^*, 1 - \mu^*], \quad \boldsymbol{a}^* = \left[\frac{A^*}{\sigma^2}, 1 - \frac{A^*}{\sigma^2}\right].$$

Then the differential entropy $h(X)$ is upper-bounded by

$$h(X) \leq \frac{1}{2} \ln(2\pi e \sigma^2) + h_b(\boldsymbol{\mu}^*) - \frac{1}{2} D(\boldsymbol{\mu}^* \|\boldsymbol{a}^*), \quad (14)$$

where $D(\cdot \|\cdot)$ is the divergence or Kullback-Leibler distance between the pmf's.

A low-rate bound is obtained by upper-bounding only the significant samples, while the other samples are quantized to zero, thus yielding a distortion floor.

Theorem 3: (Low-Rate Upper Bound) The MSE distortion rate function of a memoryless source is upper-bounded by

$$D(R) \leq B_{lr}(t, R), \quad \text{for } t \geq 0 \text{ and } R \geq 0 \quad (15)$$

where

$$B_{lr}(t, R) = A(t) \exp\left(-2 \frac{R - h_b(\mu(t))}{\mu(t)}\right) + \sigma^2 - A(t).$$

For a given threshold $t \geq 0$, this bound can be optimized to yield

$$D(R^*(t)) \leq B_{lr}(t, R^*(t)), \quad (16)$$

with the locally optimal rate $R^*(t)$ given by

$$R^*(t) = h_b(\mu(t)) - \frac{1}{2} \mu(t) \left[2h'_b(\mu(t)) + \gamma(t) + W_{-1}\left(-\gamma(t) e^{-2h'_b(\mu(t)) - \gamma(t)}\right) \right], \quad (17)$$

where γ is the reciprocal normalized second tail moment $\gamma(t) = \frac{\mu(t)}{A(t)} t^2 = \frac{t^2}{\mathbb{E}[X^2 | |X| \geq t]}$ and W_{-1} is the second real branch of Lambert's W function, taking values on $(-\infty, -1]$. ($W(x)$ solves $W(x)e^{W(x)} = x$.)

Proofs and a discussion of the situation when (17) has no solution appear in [5] and [6]. Furthermore, a corollary shows that the low-rate and high-rate bounds coincide in the minimum of the latter, that is, as expected there is a continuous transition between the two bounds. A parametric upper bound for all rates is thus

$$D(R) \leq B(R) = \begin{cases} B_{lr}(t, R), & t > t^*, \\ R = R^*(t) < R_{\min}(t^*) \\ B_{hr}(t^*, R), & R \geq R_{\min}(t^*), \end{cases} \quad (18)$$

where t^* is found by optimizing the high-rate bound (12).

Results by Sakrison [8] and Gish-Pierce [9] imply that the operational distortion rate function $\delta(R)$ of a magnitude

classifier followed by a Gaussian scalar quantizer (adapted to the class variance) will be at most a factor of $\pi e/6$ (1.53 dB) above the bound (18). Actually, this gap is even smaller at low rates, since the distortion $D_0(0) = \sigma_0^2$ is trivially achieved for the insignificant samples.

The incomplete moments (8) and (9) can be displayed with a *moment profile* curve,

$$\{(\mu(t), A(t))\}, \quad t \geq 0. \quad (19)$$

In [5], we pointed out that this curve is a good tool to characterize source sparsity and showed that it contains all information needed to compute the bounds (18), since $\frac{dA}{d\mu} = t^2$.

A simple inspection of (1) and (19) reveals that by normalizing $A(t)$ with $A(0) = \sigma^2$, one obtains a *reverse* Lorenz curve for the random variable X^2 , defined as $\bar{L}_{X^2}(u) = 1 - L_{X^2}(1 - u)$. That is, all properties for Lorenz curves hold for the normalized moment profile, with the appropriate direction reversals in the inequalities.

IV. LORENZ ORDER INDUCES A HIERARCHY OF BOUNDS

The above observation immediately leads to the main result of this work.

Theorem 4: Let X, Y be two continuous random variables with $\mathbb{E}X^2 = \mathbb{E}Y^2 < \infty$. If X^2 is less than Y^2 in Lorenz order, then the associated MSE distortion rate bound $B_X(R)$ will lie above $B_Y(R)$ for all rates, that is, $X^2 \leq_L Y^2$ implies

$$B_X(R) \geq B_Y(R), \quad R \geq 0.$$

By the above remarks on the operational significance of the bound (18), it follows that $X^2 \leq_L Y^2$ implies that when using a threshold-based MCQ scheme, Y is more compressible than X . (To show that Y is more compressible with an *optimal* scheme would require also ordered lower bounds on $D(R)$.)

Proof: Since the normalized moment profile $\{(\mu(t), a(t))\}_{t \geq 0}$ is actually identical to the reverse Lorenz curve of X^2 , defined as $\bar{L}_{X^2}(u) = 1 - L_{X^2}(1 - u)$, it follows from $X^2 \leq_L Y^2$ that $L_{X^2}(u) \geq L_{Y^2}(u)$ and therefore

$$a_X(\mu_X^{-1}(u)) \leq a_Y(\mu_Y^{-1}(u)), \quad \forall u \in [0, 1]. \quad (20)$$

Here the threshold $\mu^{-1}(u)$ is defined via (3) as $\mu^{-1}(u) = F^{-1}(1 - u)$. Inequality (20) may be written as $a_Y(\mu_Y^{-1}(u)) = a_X(\mu_X^{-1}(u)) + \Delta a$, with $\Delta a \geq 0$. Hence to prove the theorem it is sufficient to show that, for a fixed $u = \mu(t)$, the bounds B_{lr} and B_{hr} do not increase with Δa , that is, $\frac{\partial B}{\partial a} \leq 0$. Fixing u avoids having to deal with the optimization of the bounds.

For the low-rate bound, we have

$$\frac{B_{lr}}{\sigma^2} = a \exp\left(-2 \frac{R - h_b(\mu)}{\mu}\right) + 1 - a,$$

from which

$$\frac{\partial}{\partial a} \left(\frac{B_{lr}}{\sigma^2} \right) = \exp\left(-2 \frac{R - h_b(\mu)}{\mu}\right) - 1,$$

which is negative for all $R > h_b(\mu)$, that is the rates for which the bound actually improves over the trivial bound $D(R) \leq \sigma^2$

(otherwise we would choose the Gaussian upper bound and have $B_X(R) = B_Y(R)$).

For the high-rate bound, we only need to look at the divergence term in (14), since it determines (12) through $c(t)$ (recall that $\mu(t)$ is fixed). The derivative of the divergence needs to be positive for the bound to decrease with Δa . We have

$$\begin{aligned} \frac{\partial}{\partial a} D(\boldsymbol{\mu} \parallel \mathbf{a}) &= \frac{\partial}{\partial a} \left(\mu \log \frac{\mu}{a} + (1 - \mu) \log \frac{1 - \mu}{1 - a} \right) \\ &= -\frac{\mu}{a} + \frac{1 - \mu}{1 - a} = \frac{a - \mu}{a(1 - a)} \geq 0, \end{aligned}$$

where the inequality $a - \mu \geq 0$ follows from the concavity of the reverse Lorenz curve. ■

An interesting application of this result draws on the existence of parametric families of distributions for which an order on the parameters induces Lorenz order, see e.g. [4, Sec. 17.D] and [10]. Through Theorem 4, the order on the parameters therefore induces an order on the source compressibility in the operational sense outlined above. This could be used for example to optimize a transform which outputs coefficients that can be modeled as coming from a certain family of distributions (with constant variance).

As an example, we consider a distribution with density of the form

$$f_{\alpha, \xi, \theta}(x) = \frac{\alpha x^{\alpha\theta - 1}}{B(\xi, \theta) [1 + x^\alpha]^{\xi + \theta}}, \quad x \geq 0, \quad \alpha, \xi, \theta \geq 0, \quad (21)$$

which is known as *generalized F distribution*, *generalized beta distribution of the second kind* or *Feller-Pareto distribution* (see [4, Sec. 17.D] or [10]; $B(\cdot, \cdot)$ is the Beta function). The appeal of this family of distributions is its ability to model heavy tails, which are typical for compressible sources such as e.g. wavelet coefficients (and the fact that incomplete moments of the distribution can be expressed with hypergeometric functions, thus avoiding the need for explicit numeric integration).

Theorem 5 (Kleiber 1999, [10]): Let X_1, X_2 be in \mathcal{L} , $X_i \sim f_{\alpha_i, \xi_i, \theta_i}$, $i = 1, 2$. Then $\alpha_1 \leq \alpha_2$, $\alpha_1 \theta_1 \leq \alpha_2 \theta_2$, and $\alpha_1 \xi_1 \leq \alpha_2 \xi_2$ imply $X_1 \geq_L X_2$.

At this point, the reason for working with the standard (first moment) Lorenz curve for $Y = X^2$, instead of the second moment Lorenz curve for X , becomes apparent: we will apply Theorem 5 on the distribution of X^2 and then map the resulting parametric order “down” to the corresponding source distributions (since $X_1 \leq_L X_2$ does not necessarily imply $X_1^2 \leq_L X_2^2$). In the example at hand, we let $Y \sim f_Y = f_{\alpha_Y=1, \xi_Y=1.1, \theta_Y}$, establishing that $\theta_{Y_1} \leq \theta_{Y_2}$ implies $Y_1 \geq_L Y_2$ (the source becomes less compressible with increasing θ). The density f_Y determines only the density of $|X|$; a common choice for modeling the source would then be the symmetric density $f_X(x) = |x|f_Y(x^2)$. Here it turns out to be a generalized F distribution with parameters $\alpha_X = 2$, $\xi_X = \xi_Y = 1.1$ and $\theta_X = \theta_Y$. Figure 1 shows three (one-sided) densities from this family for different values of θ . Their reverse Lorenz curves in Figure 2 show the expected order. Finally, Figure 3 plots the optimized $D(R)$ bounds,

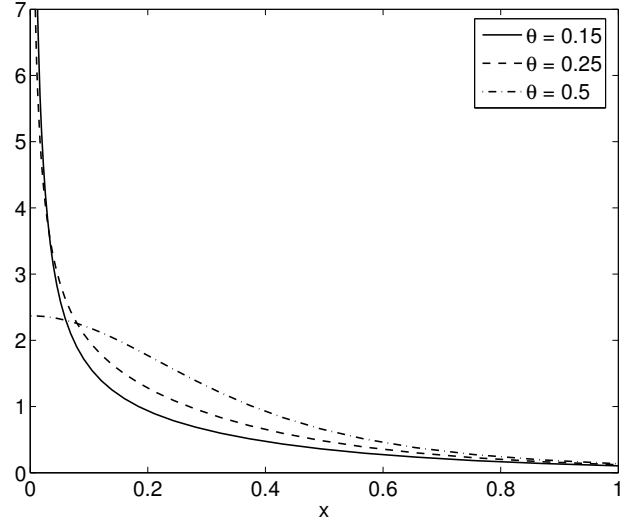


Fig. 1. Densities $f_{\alpha, \xi, \theta}(x)$ of the generalized F distribution (21) with $\alpha = 2$, $\xi = 1.1$ and $\theta = 0.15, 0.25, 0.5$. The actual source density is symmetric with $f_X(x) = 0.5f_{\alpha, \xi, \theta}(|x|)$. All densities scaled to unit variance.

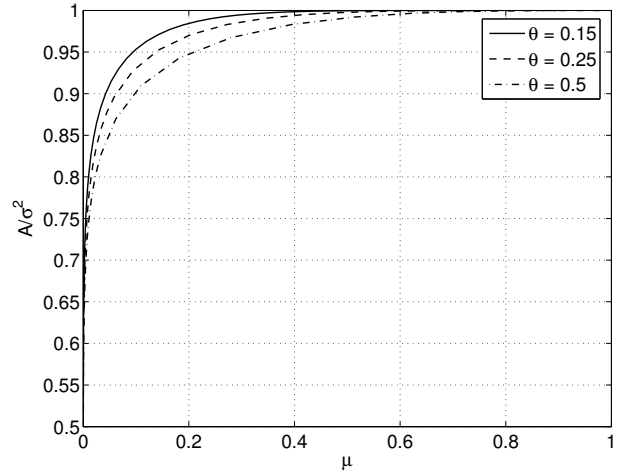


Fig. 2. Moment profiles (reverse Lorenz curves \bar{L}_{X^2}) for the source densities in Figure 1.

confirming the order implied by Theorem 4. The interest of this approach based on Lorenz order becomes apparent when one considers that a visual inspection of Figure 1 might lead to a correct guess about the order of high-rate $D(R)$ curves (which is determined by source entropy alone), but that predicting behavior at low rates seems much harder.

V. CONCLUSION

Lorenz curves and the Lorenz order were shown to be useful tools to characterize the compressibility and the distortion rate behavior of parametric families of source densities. An interesting open problem is the extension to the multivariate case; there are a number of proposals for multidimensional Lorenz curves, but their properties do not always match intuition about measures of distribution inequality (sparsity), see e.g. [4, Sec. 17.C]. A multivariate extension geared towards

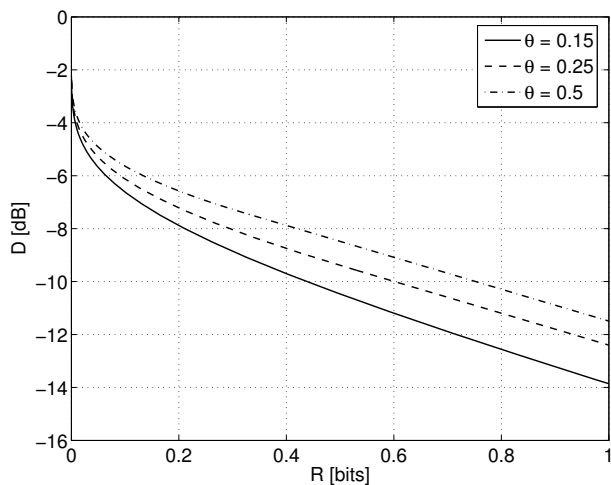


Fig. 3. Upper bounds on distortion rate for the source densities in Figure 1.

measuring compressibility, with lossy source coding in mind, would certainly need to have properties that are different from those of interest to economists.

REFERENCES

- [1] C. Weidmann, "Oligoquantization in low-rate lossy source coding," Ph.D. dissertation, EPFL, Lausanne, Switzerland, July 2000.
- [2] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.
- [3] M. O. Lorenz, "Methods of measuring the concentration of wealth," *Publications of the American Statistical Association*, vol. 9, no. 70, pp. 209–219, Jun. 1905.
- [4] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of majorization and its applications*, 2nd ed. Springer, 2011.
- [5] C. Weidmann and M. Vetterli, "Rate distortion behavior of threshold-based nonlinear approximations," in *Proc. Data Compression Conference (DCC)*, Mar. 2000, pp. 333–342.
- [6] —, "Rate distortion behavior of sparse sources," Dec. 2008, submitted to *IEEE Trans. Inform. Theory*, in revision.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [8] D. J. Sakrison, "Worst sources and robust codes for difference distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 301–309, May 1975.
- [9] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sep. 1968.
- [10] C. Kleiber, "On the Lorenz order within parametric families of income distributions," *Sankhyā B*, vol. 61, pp. 514–517, 1999.