IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

On the Statistical Physics of Directed Polymers in a Random Medium and Their Relation to Tree-Structed Lossy Compression

Neri Merhav

CCIT Report #710
December 2008

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT REPORT #710 December 2008

On the Statistical Physics of Directed Polymers in a Random Medium and Their Relation to Tree–Structured Lossy Compression*

Neri Merhav

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
merhav@ee.technion.ac.il

Abstract

Using well–known results from statistical physics, concerning the almost–sure behavior of the free energy of directed polymers in a random medium, we prove that a certain ensemble of tree–structured rate–distortion codes with delayless decoding, asymptotically achieves the rate–distortion function under a certain symmetry condition.

Index Terms: directed polymer, Cayley tree, free energy, partition function, tree–structured code, rate–distortion theory, delay.

*This research was supported by the Israel Science Foundation (ISF), grant no. 208/08.

1 Introduction

The directed polymer in a random medium (DPRM) is a statistical–mechanical model of a disordered physical system, which has attracted considerable attention of physicists for many years (cf. e.g., [1],[2],[3],[4],[5],[6],[8],[12],[13] and references therein). Loosely speaking, in this model, each configuration of the system corresponds to a walk along consecutive bonds of a certain lattice, or a tree, where each such bond is assigned with an independent random variable (energy), and where the total energy of this walk is the sum of energies along the bonds visited. For a given realization of these random energy variables, the probability of each walk is given by the Boltzmann distribution, namely, it is proportional to an exponential function of the negative total energy. The main challenge, as usual in equilibrium statistical physics, is to characterize the asymptotic normalized free energy of a typical realization of the system. For the case where the walks are defined on a tree (from the root to one of the leaves), this problem has a closed–form solution.

The main message in this short paper is the observation that, under a certain symmetry condition, this solution of the DPRM can be harnessed directly to the performance analysis of a certain ensemble of rate–distortion codes which have a tree structure that is motivated by a very simple (actually, trivial) entropy coding as well as delayless decoding. In particular, it turns out that under the above–mentioned symmetry condition, this tree–structured code ensemble asymptotically achieves the rate–distortion function, and not only in expectation, but also almost surely.

The outline of this work is as follows: In Section 2, we establish our notation conventions. In Section 3, we give a brief background in statistical mechanics in general and on the DPRM in particular. Finally, in Section 4, we show how the solution to the DPRM model can be used to prove that our tree structured code ensemble achieves the rate–distortion function.

2 Notation Conventions

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, like X and Y, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values, which will be denoted with the same symbols in the boldface font.

Thus, for example, $\boldsymbol{X}$ will denote a random n-vector $(X_1, \ldots, X_n)$, and $\boldsymbol{x} = (x_1, ..., x_n)$ is a specific vector value in $\mathcal{X}^n$, the n-th Cartesian power of $\mathcal{X}$. Sources and other probability measures that underly sequence generation will be denoted generically by the letters P and Q, and specific letter probabilities will be denoted by the corresponding lower case letters, e.g., $p(x)$, $q(y)$, etc. The expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$.

For two positive sequences $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ means that a_n and b_n are asymptotically of the same exponential order, that is, $\lim_{n\to\infty} \frac{1}{n} \ln \frac{a_n}{b_n} = 0$. Similarly, $a_n \dot{\leq} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \ln \frac{a_n}{b_n} \leq 0$, etc. Information theoretic quantities like entropies and mutual informations will be denoted following the usual conventions of the Information Theory literature.

3 Background

Consider a physical system with n particles, which can be in a variety of microscopic states ('microstates'), defined by combinations of physical quantities associated with these particles, e.g., positions, momenta, angular momenta, spins, etc., of all n particles. For each such microstate of the system, which we shall designate by a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, there is an associated energy, given by an *Hamiltonian* (energy function), $\mathcal{E}(\boldsymbol{x})$. For example, if $x_i = (\boldsymbol{p}_i, \boldsymbol{r}_i)$, where $\boldsymbol{p}_i$ is the momentum vector of particle number i and $\boldsymbol{r}_i$ is its position vector, then classically, $\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} [\frac{\|\boldsymbol{p}_i\|^2}{2m} + mgz_i]$, where m is the mass of each particle, z_i is its height – one of the coordinates of $\boldsymbol{r}_i$, and g is the gravitation constant.

One of the most fundamental results in statistical physics (based on the law of energy conservation and the basic postulate that all microstates of the same energy level are equiprobable) is that when the system is in thermal equilibrium with its environment, the probability of finding the system in a microstate $\boldsymbol{x}$ is given by the *Boltzmann–Gibbs* distribution

$$P(\boldsymbol{x}) = \frac{e^{-\beta \mathcal{E}(\boldsymbol{x})}}{Z(\beta)} \tag{1}$$

where $\beta = 1/(kT)$, k being Boltzmann's contant and T being temperature, and $Z(\beta)$ is the normalization constant, called the *partition function*, which is given by

$$Z(\beta) = \sum_{\boldsymbol{x}} e^{-\beta \mathcal{E}(\boldsymbol{x})}$$

or

$$Z(\beta) = \int \mathrm{d}\boldsymbol{x} e^{-\beta \mathcal{E}(\boldsymbol{x})},$$

depending on whether $\boldsymbol{x}$ is discrete or continuous. The role of the partition function is by far deeper than just being a normalization factor, as it is actually the key quantity from which many macroscopic physical quantities can be derived, for example, the free energy[1] is $F(\beta) = -\frac{1}{\beta} \ln Z(\beta)$, the average internal energy (i.e., the expectation of $\mathcal{E}(\boldsymbol{x})$ where $\boldsymbol{x}$ drawn is according (1)) is given by $\bar{E} \stackrel{\Delta}{=} \boldsymbol{E}\{\mathcal{E}(\boldsymbol{X})\} = -(\mathrm{d}/\mathrm{d}\beta) \ln Z(\beta)$, the heat capacity is obtained from the second derivative, etc. One of the ways to obtain eq. (1), is as the maximum entropy distribution under an average energy constraint (owing to the second law of thermodynamics), where β plays the role of a Lagrange multiplier that controls the average energy.

Quite often, real–world physical systems of many particles, such as magnetic materials and solid–state devices, are subjected to effects of impurity (e.g., defects) that may appear as amorphic structures and disorder. To model such disorder, it is customary to let the Hamiltonian, $\mathcal{E}(\boldsymbol{x})$, depend also on certain random parameters and to examine the behavior of systems pertaining to typical realizations of these random parameters. There are many models of this kind in the physics literature. One of them is the DPRM, which is defined on a certain graph, such as a hypercubic lattice, or a tree. We henceforth focus on the latter and describe it more formally than in the Introduction.

Consider a *Cayley tree*, namely, a full balanced tree with branching ratio d and depth n (cf. Fig. 1, where $d = 2$ and $n = 3$). Let us index the branches by a pair of integers (i, j), where $1 \leq i \leq n$ describes the generation (with $i = 1$ corresponding to the d branches that emanate from the root), and $0 \leq j \leq d^i - 1$ enumerates the branches of the i–th generation, say, from left to right (see Fig. 1). For each branch (i, j), $1 \leq j \leq d^i$, $1 \leq i \leq n$, we randomly draw an independent random variable $\varepsilon_{i,j}$ according to a fixed probability function $q(\varepsilon)$ (i.e., a probability mass function in the discrete case, or probability density function in the continuous case).

A *walk* $\boldsymbol{w}$, from the root of the tree to one of its leaves, is described by a finite sequence $\{(i, j_i)\}_{i=1}^{n}$, where $0 \leq j_1 \leq d - 1$ and $dj_i \leq j_{i+1} \leq dj_i + d - 1$, $i = 1, 2, \ldots, (n-1)$.[2] For a given

[1] The free energy means the maximum work that the system can carry out in any process of fixed temperature. The maximum is obtained when the process is reversible (slow, quasi–static changes in the system).

[2] In fact, for a given n, the number j_n alone dictates the entire walk.

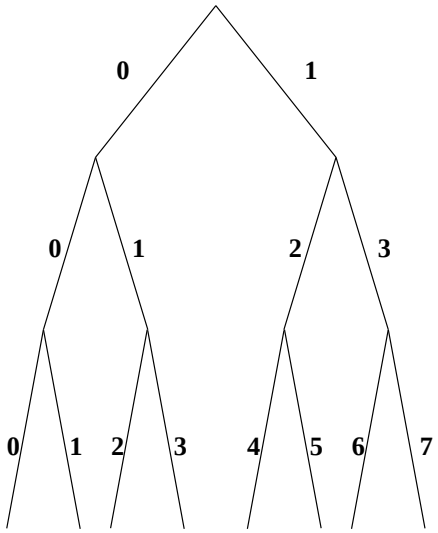

Figure 1: A Cayley tree with branching factor $d = 2$ and depth $n = 3$.

realization of the RV's $\{\varepsilon_{i,j} : \ i = 1, 2, \ldots, n, \ j = 0, 1, \ldots, d^i - 1\}$, we define the Hamiltonian associated with $\boldsymbol{w}$ as $\mathcal{E}(\boldsymbol{w}) = \sum_{i=1}^{n} \varepsilon_{i,j_i}$, and then the partition function as:

$$Z_n(\beta) = \sum_{\boldsymbol{w}} \exp\{-\beta \mathcal{E}(\boldsymbol{w})\}. \tag{2}$$

Of course, since $\{\varepsilon_{i,j}\}$ are RV's, then so is $Z_n(\beta)$. The primary question addressed by physicists, in this context, concerns the (typical) behavior of the RV

$$f_n(\beta) \stackrel{\Delta}{=} \frac{1}{n\beta} \ln Z_n(\beta) \tag{3}$$

for n large, which is (up to the minus sign), exactly the normalized free energy per step. It turns out (as proved e.g., in [2],[6]) that $f_n(\beta)$ has a *self–averaging* property, in the terminology of physicists, in other words, the sequence of random variables $\{f_n(\beta)\}_{n \geq 1}$ converges in probability (and in fact, almost surely as is shown in [2]) to a deterministic constant $f(\beta)$, which is given by

$$f(\beta) = \begin{cases} \phi(\beta) & \beta \leq \beta_c \\ \phi(\beta_c) & \beta > \beta_c \end{cases} \tag{4}$$

where

$$\phi(\beta) \stackrel{\Delta}{=} \frac{\ln[d \cdot \boldsymbol{E}\{e^{-\beta\varepsilon}\}]}{\beta} \tag{5}$$

where the expectation, which is assumed finite, is taken w.r.t. $q(\varepsilon)$, and where β_c is the value of β for which $\phi(\beta)$ is minimum, or equivalently, the solution to the equation $\phi'(\beta) = 0$, where ϕ' is the derivative of ϕ.

As can be seen, $\beta = \beta_c$ is a point at which the asymptotic normalized free energy per step changes its behavior: Although $f(\beta)$ and its first derivative are continuous functions for all β, the second derivative is discontinuous at $\beta = \beta_c$. In the terminology of physicists, this is referred to as a *second order phase transition*. Observe that while one might expect that the sequence $f_n(\beta)$ would converge to the same limit as $\boldsymbol{E}\{f_n(\beta)\} = \frac{1}{n\beta}\boldsymbol{E}\{\ln Z_n(\beta)\}$, i.e., the so called *quenched average*, the high temperature phase result (4) corresponds to $\frac{1}{n\beta}\ln[\boldsymbol{E}\{Z_n(\beta)\}]$, which is called the *annealed average*. This means that Jensen's inequality is essentially tight at this range of β. However, these two averages depart from each other at the low temperature phase, $\beta > \beta_c$. As can be observed, in this phase, the asymptotic normalized free energy no longer depends on β, and it is referred to the *glassy phase* or the *frozen phase*, which is characterized by zero thermodynamical entropy, in other words, the partition function is dominated by a sub–exponential number of configurations possessing the ground–state energy (cf. e.g., [11, Chap. 5]). For reasons that will become apparent shortly, this frozen phase is the relevant phase for our source coding problem.

The asymptotic free energy formula (4) has been proved in the physics literature at least in four different ways: The first [2] is based on martingales, the second is based on non–integer moments of the partition function [6],[8], the third is based on a recursion of a certain generating function of the partition function as well as on traveling waves [5],[7], and the fourth method is the so–called *replica method* [5], which, although not rigorous, is very useful in statistical mechanics.

4 Main Result

We now turn to our lossy source coding problem, where some of the notation that will be used will be deliberately identical to that of Section 3. Consider a discrete memoryless source (DMS) P that generates symbols $X_1, X_2, \ldots$ of a finite[3] alphabet $\mathcal{X}$. Let $\mathcal{Y}$ denote a finite reproduction alphabet and let $\rho : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a given distortion function.

Consider next a tree–structured ensemble of codes for encoding source n–tuples, $\boldsymbol{x} = (x_1, \ldots, x_n)$, which is defined as follows: Given a coding rate R (in nats/source–symbol), which is assumed to be the natural logarithm of some positive integer d, and given a probability distribution on the reproduction alphabet, $Q = \{q(y),\ y \in \mathcal{Y}\}$, let us draw $d = e^R$ independent copies of Y under Q,

[3]Finite alphabet assumptions are made mostly for simplicity. It is expected that our derivations continue to hold in the continuous case as well under suitable regularity conditions.

and denote them by $Y_1, Y_2, \ldots, Y_d$. We shall refer to the randomly chosen set, $\mathcal{C}_1 = \{Y_1, Y_2, \ldots, Y_d\}$ as our 'codebook' for the first source symbol, X_1. Next, for each $1 \le j_1 \le d$, we randomly select another such codebook under Q, $\mathcal{C}_{2,j_1} = \{Y_{j_1,1}, Y_{j_1,2}, \ldots, Y_{j_1,d}\}$, for the second symbol, X_2. Then, for each $1 \le j_1 \le d$ and $1 \le j_2 \le d$, we again draw under Q yet another codebook $\mathcal{C}_{3,j_1,j_2} = \{Y_{j_1,j_2,1}, Y_{j_1,j_2,2}, \ldots, Y_{j_1,j_2,d}\}$, for X_3, and so on. In general, for each $t \le n$, we randomly draw d^{t-1} codebooks under Q, which are indexed by $(j_1, j_2, \ldots, j_{t-1})$, $1 \le j_k \le d$, $1 \le k \le t-1$.

Once the above described random code selection process is complete, the resulting set of codebooks $\{\mathcal{C}_1, \mathcal{C}_{t,j_1,\ldots,j_{t-1}},\ 2 \le t \le n,\ 1 \le j_k \le d,\ 1 \le k \le t-1\}$ is revealed to both the encoder and decoder, and the encoding–decoding system works as follows:

- *Encoding:* Given a source n–tuple X^n, find a vector of indices $(j_1^*, j_2^*, \ldots, j_n^*)$ that minimizes the overall distortion $\sum_{t=1}^n \rho(X_t, Y_{j_1,\ldots,j_t})$. Represent each component j_t^* by $R = \ln d$ nats (that is, $\log_2 d$ bits), thus a total of nR nats.

- *Decoding:* At each time t $(1 \le t \le n)$, after having decoded $(j_1^*, \ldots, j_t^*)$, output the reproduction symbol $Y_{j_1^*,\ldots,j_t^*}$.

A few comments are in order at this point: First, as we see, the codebook generation process is branching hierarchically by a factor of d at each step, hence it is convenient to think of the code as having the structure of a Cayley tree, as in Section 3. The encoder seeks the best walk on that tree in the sense of minimum distortion. Note also that the process of converting the optimum walk $\boldsymbol{w}^* = (j_1^*, j_2^*, \ldots, j_n^*)$ into a compressed bitstream is extremely simple: We just convert each $j_t \in \{1, \ldots, d\}$ into its binary representation using $\log_2 d$ bits without any attempt at compression. In other words, the entropy coding part is trivial in the sense that it uses neither the memory that may be present in the sequence $(j_1^*, j_2^*, \ldots, j_n^*)$, nor the possible skewdness of the distributions of these symbols. Finally, the decoding process is a purely sequential delayless process: At time t, the decoder outputs the t-th reproduction symbol. This is in contrast to the decoder of a general block code, which has to wait until the entire bit string of length nR has been received before it can start to decode. Thus, at least the decoding delay is saved this way.

In order to analyze the rate–distortion performance of this ensemble of codes, using the results of Section 3, we now make the following assumption:

The random coding distribution Q be such that the distribtion of the RV $\rho(x, Y)$ is the same for all $x \in \mathcal{X}$.

It turns out that this assumption is fulfilled quite often – it is the case whenever the random coding distribution together with distortion function exhibit a sufficiently high degree of symmetry. For example, if Q is the uniform distribution over $\mathcal{Y}$ and the rows of the distortion matrix $\{\rho(x, y)\}$ are permutations of each other, which is in turn the case, for example, when $\mathcal{X} = \mathcal{Y}$ is a group and $\rho(x, y) = \gamma(x - y)$ is a difference distortion function w.r.t. the group difference operation. Somewhat more generally, this assumption still holds when the different rows of the distortion matrix are formed by permutations of each other subject to the following rule: $\rho(x, y)$ can be swapped with $\rho(x, y')$ provided that $q(y') = q(y)$.

It should be pointed out that if the optimum random coding distribution Q^*, namely, the one corresponding to the output of the test channel that achieves the rate–distortion function of X, happens to satisfy the above symmetry assumption, then as we show below, the rate–distortion performance of the above descrirbed code ensemble achieves the rate–distortion function. Moreover, this will turn out to be the case, not only in expectation, but also with probability one.

We now turn to our analysis which makes heavy use of the results of Section 3. For a given $\boldsymbol{x}$ and a given realization of the set of codebooks, define the partition function in analogy to that of the DPRM:

$$Z_n(\beta) = \sum_{\boldsymbol{w}} \exp\{-\beta \sum_{t=1}^{n} \rho(x_t, Y_{j_1,\ldots,j_t})\}, \tag{6}$$

where the summation extends over all d^n possible walks, $\boldsymbol{w} = (j_1, \ldots, j_n)$, along the Cayley tree, as defined in Section 3. Clearly, considering our symmetry assumption, this falls exactly under the umbrella of the DPRM, with the distortions $\{\rho(x_t, Y_{j_1,\ldots,j_t})\}$ playing the role of the branch energies $\{\varepsilon_{i.j}\}$. Therefore, $\frac{1}{n\beta} \ln Z_n(\beta)$ converges almost surely, as n grows without bound, to $f(\beta)$, now defined as

$$f(\beta) = \begin{cases} \phi(\beta) & \beta \leq \beta_c \\ \phi(\beta_c) & \beta > \beta_c \end{cases} \tag{7}$$

where

$$\begin{aligned}\phi(\beta) &\triangleq \frac{\ln[d \cdot \boldsymbol{E}\{e^{-\beta\rho(x,Y)}\}]}{\beta} \\ &= \frac{\ln[e^R \cdot \boldsymbol{E}\{e^{-\beta\rho(x,Y)}\}]}{\beta} \\ &= \frac{R + \ln[\boldsymbol{E}\{e^{-\beta\rho(x,Y)}\}]}{\beta}. \end{aligned} \tag{8}$$

Thus, for every $(x_1, x_2, \ldots)$, the distortion is given by

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \rho(x_t, Y_{j_1^*,\ldots,j_t^*}) &\triangleq \limsup_{n\to\infty} \frac{1}{n} \min_{\boldsymbol{w}} \left[\sum_{t=1}^{n} \rho(x_t, Y_{j_1,\ldots,j_t})\right] \\ &= \limsup_{n\to\infty} \lim_{\beta\to\infty} \left[-\frac{\ln Z_n(\beta)}{n\beta}\right] \\ &= \lim_{\beta\to\infty} \limsup_{n\to\infty} \left[-\frac{\ln Z_n(\beta)}{n\beta}\right] \\ &\overset{\text{a.s.}}{=} -\lim_{\beta\to\infty} f(\beta) \qquad (9) \\ &= -\phi(\beta_c) \\ &= \max_{\beta\geq 0} \left[-\frac{\ln[\boldsymbol{E}\{e^{-\beta\rho(x,Y)}\}] + R}{\beta}\right] \\ &\triangleq D_0(R), \qquad (10)\end{aligned}$$

where the interchange of limits at the third equality is justified by the following chain:

$$\begin{aligned}\limsup_{n\to\infty} \lim_{\beta\to\infty} \left[-\frac{\ln Z_n(\beta)}{n\beta}\right] &\leq \limsup_{n\to\infty} \lim_{\beta\to\infty} \left[-\frac{\ln \exp\{-\beta \sum_{t=1}^{n} \rho(x_t, Y_{j_1^*,\ldots,j_t^*})\}}{\beta n}\right] \\ &= \limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \rho(x_t, Y_{j_1^*,\ldots,j_t^*}) \\ &= \lim_{\beta\to\infty} \limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \rho(x_t, Y_{j_1^*,\ldots,j_t^*}) \\ &= \lim_{\beta\to\infty} \limsup_{n\to\infty} \left[-\frac{\ln \exp\{-\beta \sum_{t=1}^{n} \rho(x_t, Y_{j_1^*,\ldots,j_t^*})\}}{\beta n}\right] \\ &\leq \lim_{\beta\to\infty} \limsup_{n\to\infty} \left[-\frac{\ln[d^{-n} \sum_{\boldsymbol{w}} \exp\{-\beta \sum_{t=1}^{n} \rho(x_t, Y_{j_1,\ldots,j_t})\}}{\beta n}\right] \\ &= \lim_{\beta\to\infty} \limsup_{n\to\infty} \left[-\frac{\ln Z_n(\beta)}{\beta n}\right] + \lim_{\beta\to\infty} \frac{\ln d}{\beta} \\ &= \lim_{\beta\to\infty} \limsup_{n\to\infty} \left[-\frac{\ln Z_n(\beta)}{\beta n}\right] \qquad (11)\end{aligned}$$

and the opposite inequality (and hence equality) is obtained in exactly the same way, with the only exception that the limits over β and over n exchange their roles at all places.

We have shown then that the almost–sure distortion performance is uniformly given by $D_0(R)$ for every individual source sequence $x_1, x_2, \ldots$. Now, let us suppose that Q is chosen to be the output distribution Q^* induced by the source P and the test channel $\mathcal{X} \to \mathcal{Y}$ that achieves the rate–distortion function, and that the symmetry assumption continues to hold for $Q^* = \{q^*(y),\ y \in \mathcal{Y}\}$. Then, we claim that $D_0(R)$, defined with $Q = Q^*$, coincides with the distortion–rate function of the source, $D(R)$.

To see why this is true, recall that the rate–distortion function $R(D)$ has the following representation (see, e.g., [9, p. 90, Corollary 4.2.3],[14],[10]):

$$R(D) = -\min_{\beta \geq 0} \min_{Q} \left\{ \beta D + \sum_{x \in \mathcal{X}} p(x) \ln \left[\sum_{y \in \mathcal{Y}} q(y) e^{-\beta \rho(x,y)} \right] \right\} \tag{12}$$

which, due to convexity in β and concavity in Q, is equaivalent to

$$\begin{aligned} R(D) &= -\min_{Q} \min_{\beta \geq 0} \left\{ \beta D + \sum_{x \in \mathcal{X}} p(x) \ln \left[\sum_{y \in \mathcal{Y}} q(y) e^{-\beta \rho(x,y)} \right] \right\} \\ &= -\min_{\beta \geq 0} \left\{ \beta D + \sum_{x \in \mathcal{X}} p(x) \ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta \rho(x,y)} \right] \right\}, \end{aligned} \tag{13}$$

and which, under the symmetry assumption, tells us that for every point (D, R) on the rate–distortion curve, we have:

$$R = -\min_{\beta \geq 0} \left\{ \beta D + \ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta \rho(x,y)} \right] \right\}. \tag{14}$$

Let β^* achieve this minimum, i.e.,

$$-R = \beta^* D + \ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta^* \rho(x,y)} \right], \tag{15}$$

or, equivalently,

$$D(R) = -\frac{\ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta^* \rho(x,y)} \right] + R}{\beta^*} \tag{16}$$

Thus, clearly,

$$D(R) \leq \max_{\beta \geq 0} \left\{ -\frac{\ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta \rho(x,y)} \right] + R}{\beta} \right\} = D_0(R), \tag{17}$$

and so, it remains to show also the converse inequality, $D(R) \geq D_0(R)$. To this end, observe that eq. (14) implies that for every point (D, R) on the rate–distortion function:

$$-R \leq \beta D + \ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta \rho(x,y)} \right], \tag{18}$$

holds for all $\beta \geq 0$ (with equality for $\beta = \beta^*$). Equivalently, for all $\beta \geq 0$:

$$D \geq -\frac{\ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta \rho(x,y)} \right] + R}{\beta}, \tag{19}$$

and so,

$$D(R) \geq \max_{\beta \geq 0} \left\{ -\frac{\ln \left[\sum_{y \in \mathcal{Y}} q^*(y) e^{-\beta \rho(x,y)} \right] + R}{\beta} \right\} = D_0(R), \tag{20}$$

thus proving that $D_0(R) = D(R)$.

5 Conclusion

In this paper we have shown that for a certain ensemble of tree structured codes, having a very simple entropy coder and a zero–delay decoder, the rate–distortion bound is achieved under a certain symmetry assumption.

Perhaps even more importantly, this paper makes an attempt to demonstrate that occasionally, analysis tools, and even explicit results, may be existent in other disciplines, like statistical physics, and they can be useful, sometimes even almost verbatim.

References

[1] S. M. Bhattacharjee, "Directed polymer in a random medium – an introduction," arXiv:cond-mat/0402117v1 [cond-mat.stat-mech] 4 Feb 2004.

[2] E. Buffet, A. Patrick, and J. V. Pulé, "Directed polymers on trees: a martingale approach," *J. Phys. A: Math. Gen.*, vol. 26, pp. 1823–1834, 1993.

[3] F. Comets, T. Shiga, and N. Yoshida, "Probabilistic analysis of directed polymers in a random environment: a review," *Adv. Stud. in Pure Math.*, 2004.

[4] J. Cook and B. Derrida, "Finite–size effects in random energy models and in the problem of polymers in a random medium," *Journal of Statistical Physics*, vol. 63, nos. 3/4, 1991.

[5] B. Derrida, "Directed polymers in a random medium," *Physica A*, vol. 163, pp. 71–84, February 1990.

[6] B. Derrida, M. R. Evans, and E. R. Speer, "Mean field theory of directed polymers with random complex weights," *Commun. Math. Phys.*, vol. 156, pp. 221–244, 1993.

[7] B. Derrida and H. Spohn, "Polymers on disordered trees, spin glasses, and traveling waves," *Journal of Statistical Physics*, vol. 51, nos. 5/6, pp. 817–840, 1988.

[8] M. R. Evans and B. Derrida, "Improved bounds for the transition temperature of directed polymers in a finite–dimensional random medium," *Journal of Statistical Physics*, vol. 69, nos. 1/2, pp. 427–437, 1992.

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

[10] N. Merhav, "An identity of Chernoff bounds with an interpretation in statistical physics and applications in information theory," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3710–3721, August 2008.

[11] M. Mézard and A. Montanari, *Information, Physics and Computation*, draft, November 9, 2007. Available on–line at:
http://www.stanford.edu/∼montanar/BOOK/book.html.

[12] C. Monthus and T. Garel, "On the critical weight statistics of the random energy model and the directed polymer on the Cayley tree," arXiv:cond-mat/070301v1 [cond-mat.dis-nn], March 1, 2007.

[13] A. Procacci, B. N. B. de Lima, and B. Scoppola, "A remark on high temperature polymer expansion for lattice systems with infinite range pair interactions," *Lett. in Math. Phys.*, vol. 45, pp. 303–322, 1998.

[14] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1939–1952, November 1994.