Rate Distortion Function in the Spin Glass State: a Toy Model

Tatsuto Murayama and Masato Okada
Laboratory for Mathematical Neuroscience
RIKEN Brain Science Institute
Saitama, 351-0198, JAPAN
{murayama,okada}@brain.riken.go.jp

Abstract

We applied statistical mechanics to an inverse problem of linear mapping to investigate the physics of optimal lossy compressions. We used the replica symmetry breaking technique with a toy model to demonstrate Shannon's result. The rate distortion function, which is widely known as the theoretical limit of the compression with a fidelity criterion, is derived. Numerical study shows that sparse constructions of the model provide suboptimal compressions.

1 Introduction

Many information-science studies are very similar to those of statistical physics. Statistical physics and information science may have been expected to be directed towards common objectives since Shannon formulated an information theory based on the concept of entropy. However, envisaging how this actually happened would have been difficult; that the physics of disordered systems, and spin glass theory in particular, at its maturity naturally includes some important aspects of information sciences, thus reuniting the two disciplines. This cross-disciplinary field can thus be expected to develop much further beyond current perspectives in the future [1].

The areas where these relations are particularly strong are Shannon's coding theory [2] and classical spin systems with quenched disorder, which is the replica theory of disordered statistical systems [3]. Triggered by the work of Sourlas [4], these links have recently been examined in the area of matrix-based error corrections [5, 6], network-based compressions [7], and turbo decoding [8]. Recent results of these topics are mostly based on the replica technique. Without exception, their basic characteristics (such as channel capacity, entropy rate, or achievable rate region) are only captured by the concept of a phase transition with a first-order jump between the optimal and the other solutions arising in the scheme.

However, the research in the cross-disciplinary field so far can be categorized as a so-called 'zero-distortion' decoding scheme in terms of information theory: the system requires perfect reproduction of the input alphabets [2]. Here, the same spin glass techniques should be useful to describe the physics of systems with a fidelity criterion; i.e., a certain degree of information distortion is assumed when reproducing the alphabets. This framework is

called the rate distortion theory [9, 10]. Though processing information requires regarding the concept of distortions practically, where input alphabets are mostly represented by continuous variables, statistical physics only employs a few approaches [11, 12].

In this paper, we introduce a prototype that is suitable for cross-disciplinary study. We analyze how information distortion can be described by the concepts of statistical physics. More specifically, we study the inverse problem of a Sourlas-type decoding problem by using the framework of replica symmetry breaking (RSB) of diluted disordered systems [13]. According to our analysis, this simple model provides an optimal compression scheme for an arbitrary fidelity-criterion degree, though the encoding procedure remains an NP-complete problem without any practical encoders.

The paper is organized as follows. In Section 2, we briefly review the concept of the rate distortion theory as well as the main results related to our purpose. In Section 3, we introduce a toy model. In Section 4, we obtain consistent results with information theory. Conclusions are given in the last section. Detailed derivations will be reported elsewhere.

2   Review: Rate Distortion Theory

We briefly recall the definitions of the concepts of the rate distortion theory and state the simplest version of the main result at the end of this section. Let J be a discrete random variable with alphabet $\mathcal{J}$. Assume that we have a source that produces a sequence $J_1, J_2, \cdots, J_M$, where each symbol is randomly drawn from a distribution. We will assume that the alphabet is finit. Throughout this paper we use vector notation to represent sequences for convenience of explanation: $\boldsymbol{J} = (J_1, J_2, \cdots, J_M)^T \in \mathcal{J}^M$. Here, the encoder describes the source sequence $\boldsymbol{J} \in \mathcal{J}^M$ by a codeword $\boldsymbol{\xi} = f(\boldsymbol{J}) \in \mathcal{X}^N$. The decoder represents $\boldsymbol{J}$ by an estimate $\hat{\boldsymbol{J}} = g(\boldsymbol{\xi}) \in \hat{\mathcal{J}}^M$, as illustrated in Figure 1. Note that M represents the length of a source sequence, while N represents the length of a codeword. Here, the rate is defined by $R = N/M$. Note that the relation $N < M$ always holds when a compression is considered; therefore, $R < 1$ also holds.

Definition 2.1 *A **distortion function** is a mapping*

$$d : \mathcal{J} \times \hat{\mathcal{J}} \to \boldsymbol{R}^+ \tag{1}$$

from the set of source alphabet-reproduction alphabet pairs into the set of non-negative real numbers.

Intuitively, the distortion $d(J, \hat{J})$ is a measure of the cost of representing the symbol J by the symbol $\hat{J}$. This definition is quite general. In most cases, however, the reproduction alphabet $\hat{\mathcal{J}}$ is the same as the source alphabet $\mathcal{J}$. Hereafter, we set $\hat{\mathcal{J}} = \mathcal{J}$ and the following distortion measure is adopted as the fidelity criterion:

Definition 2.2 *The **Hamming distortion** is given by*

$$d(J, \hat{J}) = \begin{cases} 0 & \text{if } J = \hat{J} \\ 1 & \text{if } J \neq \hat{J} \end{cases}, \tag{2}$$

,

which results in a probable error distortion, since the relation $\mathrm{E}[d(J, \hat{J})] = \mathrm{P}[J \neq \hat{J}]$ holds, where $\mathrm{E}[\cdot]$ represents the expectation and $\mathrm{P}[\cdot]$ the probability of its argument. The distortion measure is so far defined on a symbol-by-symbol basis. We extend the definition to sequences:

Definition 2.3 *The distortion between sequences $\boldsymbol{J}, \hat{\boldsymbol{J}} \in \mathcal{J}^M$ is defined by*

$$d(\boldsymbol{J}, \hat{\boldsymbol{J}}) = \frac{1}{M} \sum_{j=1}^{M} d(J_j, \hat{J}_j) \, . \tag{3}$$

Therefore, the distortion for a sequence is the average distortion per symbol of the elements of the sequence.

Definition 2.4 *The distortion associated with the code is defined as*

$$D = \mathrm{E}[d(\boldsymbol{J}, \hat{\boldsymbol{J}})] \, , \tag{4}$$

where the expectation is with respect to the probability distribution on $\mathcal{J}$.

A rate distortion pair (R, D) should be **achiebable** if a sequence of rate distortion codes (f, g) exist with $\mathrm{E}[d(\boldsymbol{J}, \hat{\boldsymbol{J}})] \leq D$ in the limit $M \to \infty$. Moreover, the closure of the set of achievable rate distortion pairs is called the **rate distortion region** for a source. Finally, we can define a function to describe the boundary:

Definition 2.5 *The **rate distortion function** $R(D)$ is the infimum of rates R, so that (R, D) is in the rate distortion region of the source for a given distortion D.*

As in [7], we restrict ourselves to a binary source J with a Hamming distortion measure for simplicity. We assume that binary alphabets are drawn randomly, i.e., the source is not biased to rule out the possiblity of compression due to redundancy. We now find the description rate $R(D)$ required to describe the source with an expected proportion of errors less than or equal to D. In this simplified case, according to Shannon, the boundary can be written as follows.

Theorem 2.1 *The rate distortion function for a binary source with Hamming distortion is given by*

$$R(D) = \begin{cases} 1 - H(D) & 0 \leq D \leq \frac{1}{2} \\ 0 & \frac{1}{2} < D \end{cases}, \tag{5}$$

where $H(\cdot)$ represents the binary entropy function.
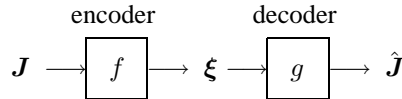
Figure 1: Rate distortion encoder and decoder

3   General Scenario

In this section, we introduce a toy model for lossy compression. We use the inverse problem of Sourlas-type decoding to realize the optimal encoding scheme [4]. As in the previous section, we assume that binary alphabets are drawn randomly from a non-biased source and that the Hamming distortion measure is selected for the fidelity criterion.

We take the Boolean representation of the binary alphabet $\mathcal{J}$, i.e., we set $\mathcal{J} = \{0, 1\}$. We also set $\mathcal{X} = \{0, 1\}$ to represent the codewords throughout the rest of this paper.

Let $\boldsymbol{J}$ be an M-bit source sequence, $\boldsymbol{\xi}$ an N-bit codeword, and $\hat{\boldsymbol{J}}$ an M-bit reproduction sequence. Here, the encoding problem can be written as follows. Given a distortion D and a randomly-constructed Boolean matrix A of dimensionality $M \times N$, we find the N-bit codeword sequence $\boldsymbol{\xi}$, which satisfies

$$\hat{\boldsymbol{J}} = A\boldsymbol{\xi} \qquad (\text{mod } 2) \, , \tag{6}$$

where the fidelity criterion

$$D = \mathrm{E}[d(\boldsymbol{J}, \hat{\boldsymbol{J}})] \tag{7}$$

holds, according to every M-bit source sequence $\boldsymbol{J}$. Note that we applied modulo 2 arithmetics for the additive operations in (6). In our framework, decoding will just be a linear mapping $\hat{\boldsymbol{J}} = A\boldsymbol{\xi}$, while encoding remains a NP-complete problem.

Kabashima and Saad recently expanded on the work of Sourlas, which focused on the zero-rate limit, to an arbitrary-rate case [5]. We follow their construction of the matrix A, so we can treat practical cases. Let the Boolean matrix A be characterized by K ones per row and C per column. The finite, and usually small, numbers K and C define a particular code. The rate of our codes can be set to an arbitrary value by selecting the combination of K and C. We also use K and C as control parameters to define the rate $R = K/C$. If the value of K is small, i.e., the relation $K \ll N$ holds, the Boolean matrix A results in a very sparse matrix. By contrast, when we consider densely constructed cases, K must be extensively big and have a value of $\mathcal{O}(N)$. We can also assume that K is not $\mathcal{O}(1)$ but $K \ll N$ holds. The codes within *any* parameter region, including the sparsely-constructed cases, will result in optimal codes as we will conclude in the following section. This is one new finding of our analysis using statistical physics.

4 Physics of the model: One-step RSB Scheme

The similarity between codes of this type and Ising spin systems was first pointed out by Sourlas, who formulated the mapping of a code onto an Ising spin system Hamiltonian in the context of error correction [4]. To facilitate the current investigation, we first map the problem to that of an Ising model with finite connectivity following Sourlasfmethod. We use the Ising representation $\{1, -1\}$ of the alphabet $\mathcal{J}$ and $\mathcal{X}$ rather than the Boolean one $\{0, 1\}$; the elements of the source $\boldsymbol{J}$ and the codeword sequences $\boldsymbol{\xi}$ are rewritten in Ising values by mapping only, and the reproduction sequence $\hat{\boldsymbol{J}}$ is generated by taking products of the relevant binary codeword sequence elements in the Ising representation $\hat{J}_{\langle i_1, i_2, \cdots, i_K \rangle} = \xi_{i_1} \xi_{i_2} \cdots \xi_{i_K}$, where the indices $i_1, i_2, \cdots, i_K$ correspond to the ones per row A, producing a Ising version of $\hat{\boldsymbol{J}}$. Note that the additive operation in the Boolean representation is translated into the multiplication in the Ising one. Hereafter, we set $J_j, \hat{J}_j, \xi_i = \pm 1$ while we do not change the notations for simplicity. As we use statistical-mechanics techniques, we consider the source and codeword-sequence dimensionality (M and N, respectively) to be infinite, keeping the rate $R = N/M$ finite. To explore the system's capabilities, we examine the Hamiltonian:

$$\mathcal{H}(\boldsymbol{S}) = - \sum_{\langle i_1, \cdots, i_K \rangle} \mathcal{A}_{\langle i_1, \cdots, i_K \rangle} J_{\langle i_1, \cdots, i_K \rangle} S_{i_1} \cdots S_{i_K} \, , \tag{8}$$

where we have introduced the dynamical variable S_i to find the most suitable Ising codeword sequence $\boldsymbol{\xi}$ to provide the reproduction sequence $\hat{\boldsymbol{J}}$ in the decoding stage. Elements of the sparse connectivity tensor $\mathcal{A}_{\langle i_1, \cdots, i_K \rangle}$ take the value one if the corresponding indices of codeword bits are chosen (i.e., if all corresponding indices of the matrix A are one) and zero otherwise; C ones per i index represent the system's degree of connectivity.

For calculating the partition function $\mathcal{Z}(\mathcal{A}, \boldsymbol{J}) = \text{Tr}_{\{\boldsymbol{S}\}} \exp[-\beta \mathcal{H}(\boldsymbol{S})]$, we apply the replica method following the calculation of Kabashima and Saad [5]. To calculate replica-free energy, we have to calculate the annealed average of the n-th power of the partition function by preparing n replicas. Here we introduce the inverse temperature β, which can be interpreted as a measure of the system's sensitivity to distortions. As we see in the following calculation, the optimal value of β is naturally determined when the consistency of the replica symmetry breaking scheme is considered [13, 3]. We use integral representations of the Dirac δ function to enforce the restriction, C bonds per index, on $\mathcal{A}$ [14]:

$$\delta \left(\sum_{\langle i_2, i_3, \cdots, i_K \rangle} \mathcal{A}_{\langle i, i_2, \cdots, i_K \rangle} - C \right) = \oint_0^{2\pi} \frac{dZ}{2\pi} Z^{-(C+1)} Z^{\sum_{\langle i_2, i_3, \cdots, i_K \rangle} \mathcal{A}_{\langle i, i_2, \cdots, i_K \rangle}} , \quad (9)$$

giving rise to a set of order parameters

$$q_{\alpha, \beta, \cdots, \gamma} = \frac{1}{N} \sum_{i=1}^N Z_i S_i^\alpha S_i^\beta \cdots S_i^\gamma , \quad (10)$$

where $\alpha, \beta, \cdots, \gamma$ represent replica indices, and the average over $\boldsymbol{J}$ is taken with respect to the probability distribution:

$$\text{P}[J_{\langle i_1, i_2, \cdots, i_K \rangle}] = \frac{1}{2} \delta(J_{\langle i_1, i_2, \cdots, i_K \rangle} - 1) + \frac{1}{2} \delta(J_{\langle i_1, i_2, \cdots, i_K \rangle} + 1) \quad (11)$$

as we consider the non-biased source sequences for simplicity. Assuming the replica symmetry, we use a different representation for the order parameters and the related conjugate variables [14]:

$$q_{\alpha, \beta, \cdots, \gamma} = q \int dx \, \pi(x) \tanh^l(\beta x) , \quad (12)$$

$$\hat{q}_{\alpha, \beta, \cdots, \gamma} = \hat{q} \int dx \, \hat{\pi}(\hat{x}) \tanh^l(\beta \hat{x}) , \quad (13)$$

where $q = [(K-1)! NC]^{1/K}$ and $\hat{q} = [(K-1)!]^{-1/K} [NC]^{(K-1)/K}$ are normalization constants, and $\pi(x)$ and $\hat{\pi}(\hat{x})$ represent probability distributions related to the integration variables. Here l denotes the number of related replica indices. Throughout this paper, integrals with unspecified limits denote integrals over the range of $(-\infty, +\infty)$. We then obtain an expression for the free energy per source bit expressed in terms of the probability distributions $\pi(x)$ and $\hat{\pi}(\hat{x})$:

$$\begin{aligned}
-\beta f =& \frac{1}{M} \langle \langle \ln \mathcal{Z}(\mathcal{A}, \boldsymbol{J}) \rangle \rangle \\
=& \ln \cosh \beta \\
& + \int \prod_{l=1}^K dx_l \, \pi(x_l) \left\langle \ln \left(1 + \tanh \beta J \prod_{l=1}^K \tanh \beta x_l \right) \right\rangle_J \\
& - K \int dx \, \pi(x) \int d\hat{x} \, \hat{\pi}(\hat{x}) \ln(1 + \tanh \beta x \, \tanh \beta \hat{x}) \\
& + \frac{C}{K} \int \prod_{l=1}^C d\hat{x}_l \, \hat{\pi}(\hat{x}_l) \ln \left[\text{Tr}_S \prod_{l=1}^C (1 + S \tanh \beta \hat{x}_l) \right] ,
\end{aligned} \quad (14)$$

where $\langle \langle \cdots \rangle \rangle$ denotes the average over quenched randomness of $\mathcal{A}$ and $\boldsymbol{J}$. The saddle point equations with respect to probability distributions provide a set of relations between

$\pi(x)$ and $\hat{\pi}(\hat{x})$:

$$
\pi(x) = \int \left[\prod_{l=1}^{C-1} d\hat{x}_l \, \hat{\pi}(\hat{x}_l) \right] \delta \left(x - \sum_{l=1}^{C-1} \hat{x}_l \right) ,
$$

$$
\hat{\pi}(\hat{x}) = \int \left[\prod_{l=1}^{C-1} dx_l \, \pi(x_l) \right] \left\langle \delta \left[\hat{x} - \frac{1}{\beta} \tanh^{-1} \left(\tanh \beta J \prod_{l=1}^{K-1} \tanh \beta x_l \right) \right] \right\rangle_J . \tag{15}
$$

By using the result obtained for the free energy, we can easily perform further straight-forward calculations to find all the other observable thermodynamical quantities, including internal energy:

$$
e = \frac{1}{M} \left\langle \left\langle \mathrm{Tr}_{\boldsymbol{S}} \mathcal{H}(\boldsymbol{S}) e^{-\beta \mathcal{H}(\boldsymbol{S})} \right\rangle \right\rangle = -\frac{1}{M} \frac{\partial}{\partial \beta} \langle\langle \ln \mathcal{Z}(\mathcal{A}, \boldsymbol{J}) \rangle\rangle , \tag{16}
$$

which records reproduction errors. Therefore, in terms of the considered replica symmetric ansatz, a complete solution of the problem seems to be easily obtainable; unfortunately, it is not.

This set of equations (15) may be solved numerically for general β, K, and C. However, there exists an analytical solution of this equations. We first consider this case. Two dominant solutions emerge that correspond to the paramagnetic and the spin glass phases. The paramagnetic solution, which is also valid for general β, K, and C, is in the form of $\pi(x) = \delta(x)$ and $\hat{\pi} = \delta(\hat{x})$; it has the lowest possible free energy per bit $f_{\mathrm{PARA}} = -1$, although its entropy $s_{\mathrm{PARA}} = (R-1)\ln 2$ is positive only for $R \geq 1$. It means that the true solution must be somewhere beyond the replica symmetric ansatz. As a first step, which is called the one-step replica symmetry breaking (RSB), n replicas are usually divided into n/m groups, each containing m replicas. Pathological aspects due to the replica symmetry may be avoided making use of the newly-defined freedom m. Actually, this one-step RSB scheme is considered to provide the exact solutions when the random energy model limit is considered [15], while our analysis is not restricted to this case.

The spin glass solution can be calculated for both the replica symmetric and the one-step RSB ansatz. The former reduces to the paramagnetic solution ($f_{\mathrm{RS}} = f_{\mathrm{PARA}}$), which is unphysical for $R < 1$, while the latter yields $\pi_{\mathrm{1RSB}}(x) = \delta(x)$, $\hat{\pi}_{\mathrm{1RSB}}(\hat{x}) = \delta(\hat{x})$ with $m = \beta_g(R)/\beta$ and β_g obtained from the root of the equation enforcing the non-negative replica symmetric entropy

$$
s_{\mathrm{RS}} = \ln \cosh \beta_g - \beta_g \tanh \beta_g + R \ln 2 = 0 , \tag{17}
$$

with a free energy

$$
f_{\mathrm{1RSB}} = -\frac{1}{\beta_g} \ln \cosh \beta_g - \frac{R}{\beta_g} \ln 2 . \tag{18}
$$

Since the target bit of the estimation in this model is $J_{\langle i_1, \cdots, i_K \rangle}$ and its estimator the product $S_{i_1} \cdots S_{i_K}$, a performance measure for the information corruption could be the per-bond energy e. According to the one-step RSB framework, the lowest free energy can be calculated from the probability distributions $\pi_{\mathrm{1RSB}}(x)$ and $\hat{\pi}_{\mathrm{1RSB}}(\hat{x})$ satisfying the saddle point equation (15) at the characteristic inverse temperature β_g, when the replica symmetric entropy s_{RS} disappears. Therefore, f_{1RSB} equals e_{1RSB}. Let the Hamming distortion be our fidelity criterion. The distortion D associated with this code is given by the fraction of the free energies that arise in the spin glass phase:

$$
D = \frac{f_{1RSB} - f_{RS}}{2|f_{RS}|} = \frac{1 - \tanh \beta_g}{2} . \tag{19}
$$

Here, we substitute the spin glass solutions into the expression, making use of the fact that the replica symmetric entropy s_{RS} disappears at a consistent β_g, which is determined by (17). Using (17) and (19), simple algebra gives the relation between the rate $R = N/M$ and the distortion D in the form

$$R = 1 - H(D) \, ,$$

which coincides with the rate distortion function retrieving Theorem 2.1. Surprisingly, we do *not* observe any first-order jumps between analytical solutions. Recently, we have seen that many approaches to the family of codes, characterized by the linear *encoding* operations, result in a quite different picture: the optimal boundary is constructed in the random energy model limit and is well captured by the concept of a first-order jump. Our analysis of this model, viewed as a kind of inverse problem, provides an exception. Many optimal conditions in textbook information theory may be well described without the concept of a first-order phase transitions from a view point of statistical physics.

We will now investigate the possiblity of the other solutions satisfying (15) in the case of finite K and C. Since the saddle point equations (15) appear difficult for analytical arguments, we resort to numerical evaluations representing the probability distributions $\pi_{\mathrm{1RSB}}(x)$ and $\hat{\pi}_{\mathrm{1RSB}}(\hat{x})$ by up to 10^5 bin models and carrying out the integrations by using Monte Carlo methods. Note that the characteristic inverse temperature β_g is also evaluated numerically by using (17). We set $K = 2$ and selected various values of C to demonstrate the performance of stable solutions. The numerical results obtained by the one-step RSB senario show suboptimal properties [Figure 2]. This strongly implies that the analytical solution is *not* the only stable solution. This conjecture might be verified elsewhere, carrying out large scale simulations.

5   Conclusions

Two points should be noted. Firstly, we found that the consistency between the rate distortion theory and the Parisi one-step RSB scheme. Secondly, we conjectured that the analytical solution, which is consistent with the Shannon's result, is not the only stable solution for some situations. We are currently working on the verification.

Acknowledgments

We thank Yoshiyuki Kabashima and Shun-ichi Amari for their comments on the manuscript. We also thank Hiroshi Nagaoka and Te Sun Han for giving us valuable references. This research is supported by the Special Postdoctoral Researchers Program at RIKEN.

References

[1] H. Nishimori. *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press, 2001.

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

[3] V. Dotsenko. *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge University Press, 2001.

[4] N. Sourlas. Spin-glass models as error-correcting codes. *Nature*, 339:693–695, 1989.

[5] Y. Kabashima and D. Saad. Statistical mechanics of error-correcting codes. *Europhys. Lett.*, 45:97–103, 1999.

[6] Y. Kabashima, T. Murayama, and D. Saad. Typical performance of Gallager-type error-correcting codes. *Phys. Rev. Lett.*, 84:1355–1358, 2000.
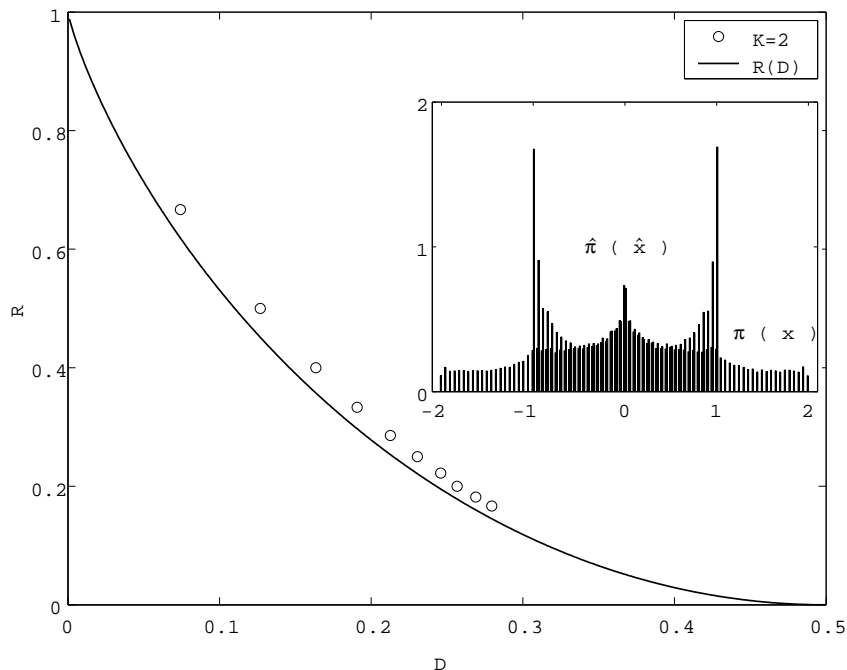
Figure 2: Numerically-constructed stable solutions: Stable solutions of (15) for the finite values of K and L are calculated by using Monte Carlo methods. We use 10^5 bin models to approximate the probability distributions $\pi_{1RSB}(x)$ and $\hat{\pi}_{1RSB}(\hat{x})$, starting from various initial conditions. The distributions converge to the continuous ones, giving suboptimal performance. ($\circ$) $K = 2$ and $L = 3, 4, \cdots, 12$; Solid line indicates the rate distortion function $R(D)$. Inset: Snapshots of the distributions, where $L = 3$ and $\beta_g = 2.35$.

[7] T. Murayama. Statistical mechanics of the data compression theorem. *J. Phys. A*, 35:L95–L100, 2002.

[8] A. Montanari and N. Sourlas. The statistical mechanics of turbo codes. *Eur. Phys. J. B*, 18:107–119, 2000.

[9] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record, Part 4*, pages 142–163, 1959.

[10] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, 1971.

[11] T. Hosaka, Y. Kabashima, and H. Nishimori. Statistical mechanics of lossy data compression using a non-monotonic perceptron. *cond-mat/0207356*.

[12] Y. Matsunaga and H. Yamamoto. A coding theorem for lossy data compression by LDPC codes. In *Proceedings 2002 IEEE International Symposium on Information Theory*, page 461, 2002.

[13] M. Mezard, G. Parisi, and M. Virasoro. *Spin-Glass Theory and Beyound.* World Scientific, 1987.

[14] K. Y. M. Wong and D. Sherrington. Graph bipartitioning and spin glasses on a random network of fixed finite valence. *J. Phys. A*, 20:L793–L799, 1987.

[15] B. Derrida. The random energy model, an exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, 1981.