Published as: *Neural Comput*. 2014 July ; 26(7): 1329–1339.

On Criticality in High-Dimensional Data

Saeed Saremi and **Terrence J. Sejnowski**
Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A

Saeed Saremi: saeed@salk.edu; Terrence J. Sejnowski: terry@salk.edu

Abstract

Data sets with high dimensionality such as natural images, speech, and text have been analyzed with methods from condensed matter physics. Here we compare recent approaches taken to relate the scale invariance of natural images to critical phenomena. We also examine the method of studying high-dimensional data through specific heat curves by applying the analysis to noncritical systems: 1D samples taken from natural images and 2D binary pink noise. Through these examples, we concluded that due to small sample sizes, specific heat is not a reliable measure for gauging whether high-dimensional data are critical. We argue that identifying order parameters and universality classes is a more reliable way to identify criticality in high-dimensional data.

1 Introduction

Scale-free structures are ubiquitous in nature. The best-understood examples of these structures are in critical phenomena, which occur in the neighborhood of a second-order phase transition. At critical points, the scale of large fluctuations becomes infinite while at the same time, small fluctuations persist (Wilson, 1979). The system is characterized by an order parameter, and long-range fluctuations of the order parameter at the critical point are studied using the renormalization group (Wilson & Kogut, 1974). The hallmark of the theory of critical phenomena was to show that criticality falls into universality classes dictated by the symmetry and dimensionality of the problem (Ma, 1976). Systems in the same universality class behave the same way regarding correlation functions at long distances even though they could have very different microscopic degrees of freedom. Scale-free structures also exist in high-dimensional data; natural images are the most prominent example (Field, 1987; Ruderman & Bialek, 1994). The key question is how the scale invariance in these systems relates to the theory of critical phenomena.

The issue of scale invariance in high-dimensional data is important from both information theory and machine learning and also from the neuroscience perspective. From the machine learning perspective, understanding scale-free structures is important in efficient coding and in the quest for finding generative models (Ackley, Hinton, & Sejnowski, 1979; Hinton & Salakhutdinov, 2006). Furthermore, the neural representations of inputs are related to their statistical structure (Barlow, 1961; Simoncelli & Olshausen, 2001), and understanding their scale invariance may shed light on the way the cortex processes them. With the goal of understanding further the scale invariance in these systems, we examine the reliability of studying criticality in high-dimensional data through the qualitative behavior of specific heat curves.

2 Review of Critical Phenomena Concepts

In this section, we review some of the key concepts of the theory of critical phenomena in condensed matter physics. The field of condensed matter physics deals with emergent properties of large number of particles interacting with each other. The most prominent example of emergence is symmetry breaking, which is ubiquitous in nature, underlying different phases of matter. Symmetry breaking happens when the equilibrium phase that a system with large number of particles settles into has a reduced symmetry compared to the symmetries that underlie its dynamics. An example of symmetry breaking is a magnet, which has the same energy under a global rotation of all magnetic moments; however, as the magnetic system is cooled down, there is a temperature (the critical temperature) at which all the magnetic elements begin to align themselves in a particular direction. The "order" in the broken symmetry state is quantified by order parameter, which is the average magnetization in the case of a magnet. The phase transitions are classified by whether the order parameter changes discontinuously (first-order phase transition) or continuously (second-order phase transition), where in the first case, the first derivative of free energy (with respect to temperature) is discontinuous at the transition point, and in the second case, the second derivative has discontinuity. The order parameter is zero at the second-order phase transition, but the system is highly fluctuating, deciding which direction to pick. These fluctuations are visualized by large clusters ordered in one direction while being surrounded by clusters ordered in a different direction. In the thermodynamic limit (for infinite systems), the sizes of the largest clusters become infinite, while inside them, smaller fluctuations of all sizes persist. The size of the largest cluster is a measure of the correlation length in the system. At the critical point, the correlation length is infinite, and the thermodynamic state is known as the critical state; it is scale free because of the infinite correlation length. The scale invariance appears in the correlation functions, taking algebraic form, and the exponents characterizing the algebraic correlation functions at critical points are known as critical exponents. Critical points fall into universality classes even for systems with different microscopic degrees of freedom, such that critical states that are in the same universality class have the same critical exponents. In the quest to relate the scale invariance of high-dimensional data to critical phenomena, an important step is to identify an order parameter that has critical fluctuations and also the universality class of the system.

3 Criticality in Natural Images

We focus on natural images to frame the issues. We start by comparing two recent studies in analyzing the scale invariance of natural images (Stephens, Mora, Tkačik, & Bialek, 2013; Saremi & Sejnowski, 2013). In the first approach, grayscale images were studied using thermodynamic quantities after first transforming them into an Ising system by binarizing images. We summarize the formalism in Stephens et al. (2013) for clarity, adopting a notation similar to that of Saremi and Sejnowski (2013). When we denote the grayscale image as the matrix I, the image is transformed to the binary image Θ (see Figure 1),

$$\mathscr{I} \to \Theta = \theta(\mathscr{I} - \mathrm{median}(\mathscr{I})), \quad (3.1)$$

where θ is the step function. In this way, the grayscale image database is transformed to an Ising system, where at each site *i*, the pixel value is either 0 or 1, depending on whether the pixel intensity is below or above the median intensity. The ensemble Θ has the following properties: (1) by construction, there are equal numbers of zeros and ones on average, and the average magnetization is zero (magnetic variables are obtained by changing variables 0 $\rightarrow -1$, 1 $\rightarrow$ 1), and (2) the spin correlation function in the ensemble decays with the distance as a power law, and with an exponent similar to the corresponding exponent for images *I*. These two facts motivated the search in Stephens et al. (2013) for signs of criticality by taking samples of size $N = n \times n$ from ensemble Θ and systematically increasing *n*. For a system of size *N*, the space of all possible configurations Ω_N has 2^N elements. The distribution $P(\omega)$ ($\omega \in \Omega_N$) was estimated by taking samples from the ensemble Θ. Assuming the system is in equilibrium, $P(\omega)$ is related to the energy $E(\omega)$ of the configuration ω by the Boltzmann distribution:

$$P(\omega)=\exp(-E(\omega))/Z, \quad (3.2)$$

where *Z* is the normalizing constant, known as the partition function, and the temperature was taken to be $T = 1$ for the ensemble Θ. Knowing $P(\omega)$ at $T = 1$, the distribution at a different temperature *T* is given by

$$P_T(\omega)=\exp(-E(\omega)/T)/Z(T), \quad (3.3)$$

$$Z(T)=\sum_{\omega\in\Omega_N}\exp(-E(\omega)/T). \quad (3.4)$$

The entropy $S(T)$ and the specific heat $C(T)$ can then be constructed from the distribution $P_T(\omega)$:

$$S(T)=-\sum_{\omega\in\Omega_N}P_T(\omega)\log P_T(\omega), \quad (3.5)$$

$$C(T)=T\frac{dS(T)}{dT}. \quad (3.6)$$

Stephens et al. (2013) observed that the peak in $C(T)/N$ versus *T* increased with *N*, while the width became narrower and the location of the peak moved closer to $T = 1$. Their analysis is reproduced here in Figure 2 for the van Hateren database (van Hateren & van der Schaaf, 1998). Since Ω_N grows exponentially with *N*, this analysis is feasible only for systems of size (for square systems) $N = \{2 \times 2, 3 \times 3, 4 \times 4\}$. The conclusion was that in the thermodynamic limit ($N \rightarrow \infty$), $C(T)/N$ diverges at or near $T = 1$, which would imply that the original system Θ was at or near a critical point.

Next we outline the approach in Saremi and Sejnowski (2013). Images were mapped to a stack of binary layers $\mathscr{B}_\lambda$ (bit planes) obtained uniquely through the relation

$$\mathscr{I} = \sum_{\lambda=1}^{L} 2^{L-\lambda} \mathscr{B}_\lambda, \quad (3.7)$$

where L is the bit length of the representation ($L = 15$ for the van Hateren database). A grayscale $[0, 1, \ldots, 2^L - 1]$ image was thus replaced with L images, each being binary $\{0, 1\}$ (see Figure 1). This representation imposes a hierarchy in intensities, which was emphasized schematically by taking $\lambda = 1$ to be the "top" layer and $\lambda = 15$ the "bottom" one. Going from the top to the bottom layer, there was a qualitative change from order to disorder. Changing to magnetic variables, the average magnetization of the top layers was close to −1, and the bottom ones were close to 0, with a sharp transition near $\lambda = 6$ reminiscent of a second-order phase transition (Saremi & Sejnowski, 2013). The scaling exponent of the power spectrum for layers near $\lambda = 6$ was close to that of the 2D ferromagnetic Ising model at its critical point. It was therefore argued that the scaling of natural images has its roots in the layers close to the phase transition ($\lambda = 6$).

The link between these two approaches is as follows. A binary image obtained by thresholding based on its median intensity (see equation 3.1) is approximately equal to the disjunction of layers above the median layer $\mu = L - \log_2 \text{median}(I)$), by applying the logical OR operator. This is so because in binary representation, if any of the units above μ are active, it makes the value bigger than the median value. However, this is approximate because μ may not necessarily be an integer. In the van Hateren database, the median intensity lies, on average, between layers 5 and 6 ($\mu = 5.7 \pm 0.46$). Therefore:

$$\theta(\mathscr{I} - \text{median}(\mathscr{I})) \simeq \mathscr{B}_1 \vee \mathscr{B}_2 \vee \cdots \vee \mathscr{B}_6. \quad (3.8)$$

In short, the binary ensemble obtained through $I \to \theta(I - \text{median}(I))$ can be obtained by nonlinear mixing of bit planes $\mathscr{B}_\lambda$ above the phase transition with the OR operator (see Figure 1). This mixing, however, changes the scaling exponent of the power spectrum. For a scale-invariant system, in Fourier space, the power spectrum has the scaling form $1/|k|^\alpha$, as $|k| \to 0$.. The exponent α for layer $\lambda = 6$ is close to 1.75 (which is the exponent of the 2D Ising critical system); however, it is closer to 2 for binary images Θ. Is it possible that both $\mathscr{B}_6$ and Θ are close to a critical point? If that is the case, we know from the theory of critical phenomena that they must belong to different universality classes, since they have different scaling behaviors. Addressing the universality class of Θ remains an interesting problem from the perspective of critical phenomena. However, comparing the two approaches inspired us to examine the formalism developed in Stephens et al. (2013) by studying noncritical Ising systems: a one-dimensional system extracted from natural images and two-dimensional binary pink noise. These exhibit (see Figures 3 and 4) qualitative behavior similar to that in Figure 2, which questions the reliability of specific heat curves for gauging criticality.

4 Specific Heat Curves for Noncritical Systems

4.1 1D Samples from Natural Images

Here we construct a one-dimensional ensemble extracted from natural images. The database was constructed by sampling 1D stripes from median-thresholded images Θ. They are sampled at random locations and random orientations (either horizontal or vertical). We examined this system for 1D samples of $N = \{2, 3, \ldots, 16\}$. The results for the specific heat curves are given in Figure 3a. What is surprising is that a qualitative behavior similar to samples taken from Θ (see Figure 2) is observed here: the size of the peak $C(T)/N$ increases by increasing N (see Figure 3b), the width becomes narrower (see Figure 3c), and the location of the peak shifts toward $T = 1$ (see Figure 3d). However, interactions in natural images are local, and we know from the theory of critical phenomena that a one-dimensional system with local interactions cannot go through a phase transition. The advantage of studying 1D ensemble is that we could study the finite size scaling of the peak and the width (see Figure 2) to see if the increase in peak and the decrease in the width saturate for larger systems. However in this case, the saturation does not occur up to system of size $N = 16$.

To demonstrate clearly the lack of long-range correlations in the 1D system, the power spectrum as a function of spatial frequency k (for the original gray scale samples and after binarization) is given in Figure 3e. It shows that over long distances (low spatial frequencies), the 1D system behaves similar to the white noise. It also shows the large effect that binarization has (special to 1D) in the intermediate scales in flattening the slope of power versus spatial frequency.

4.2 2D Binary Pink Noise

We generated pink noise from the gaussian white noise, followed by low-pass-filtering the Fourier components by $1/k^{\alpha/2}$; the resulting samples in real space were then thresholded by their median intensity to get binary samples. The power spectrum of the resulting ensemble (denoted by $\wp\alpha$) is $1/k^{\alpha}$. The specific heat curves for different values of α, together with representative samples, are given in Figure 4 for systems of sizes $N = \{2 \times 2, 3 \times 3\text{m } 4 \times 4\}$. For a range of α, the specific heat curves have similar qualitative behavior to the ensemble Θ obtained from images. We suspect that as in the 1D case, the specific heat behavior is an artifact of $P(\omega)$ for small system sizes. This remains an open problem from an analytical point of view, especially since the “pseudo-critical” behavior of specific heat curves were not seen for all values of α. To check the robustness of these results, we also studied the specific heat for samples after high-pass-filtering them (see Figure 4.) Some examples of the pink noise samples and their high-pass-filtered version are given in Figure 4.

5 Discussion

The focus of this letter is on natural images, but the issues are general and relevant to other systems including data from ganglion cells in the retina (Schneidman, Berry, Segev, & Bialek, 2006; Tkačik, Schneidman, Berry, & Bialek, 2009; Ganmor, Segev, & Schneidman, 2011). We examined the methods developed in Stephens et al. (2013) by applying them to 1D binary samples and 2D binary pink noise, and found that the analysis of the specific heat curves may lead to false positives in gauging whether a system is critical. Identifying an

order parameter and the universality class is essential in addressing criticality in these systems. However, it could be that some of these systems have a universality class of their own, which would also be very interesting. This is a difficult problem, since in physics, order parameters are usually dictated by symmetries in the system. But in the case of neural activities, for example, it is not clear what the underlying symmetries in the process of generating spiking activities are. Regarding natural images, the issue of universality class and order parameter remains open despite the attempts made in Saremi and Sejnowski (2013) in addressing the issue. We expand on this in an upcoming article.

Finally, we mention the work of Macke, Opper, and Bethge (2011), where the divergence of specific heat was reported for "dichotomized gaussian" samples. In that work, the connectivity is all to all, but a short-range connectivity constructed with their method could also lead to the pseudo-divergence of the specific heat, which would strengthen the results we have reported here.

Acknowledgments

We thank William Bialek, Mehran Kardar, Matthias Bethge, and Yashar Ahmadian for their comments and suggestions.

References

Ackley D, Hinton G, Sejnowski T. A learning algorithm for Boltzmann machines. Cognitive Sci. 1979; 9:147–169.

Barlow, H. Possible principles underlying the transformation of sensory messages. In: Rosenblith, W., editor. Sensory communication. Cambridge, MA: MIT Press; 1961. p. 217-234.

Field D. Relations between the statistics of natural images and the response properties of cortical cells. J Optical Soc Am A. 1987; 12:2379–2394.

Ganmor E, Segev R, Schneidman E. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. Proceedings of the National Academy of Sciences. 2011; 108(23):9679–9684.

Hinton G, Salakhutdinov R. Reducing the dimensionality of data with neural networks. Science. 2006; 313(5786):504–507. [PubMed: 16873662]

Ma, SK. Modern theory of critical phenomena. New York: Benjamin; 1976.

Macke JH, Opper M, Bethge M. Common input explains higher-order correlations and entropy in a simple model of neural population activity. Physical Review Letters. 2011; 106(20):208102. [PubMed: 21668265]

Ruderman DL, Bialek W. Statistics of natural images: Scaling in the woods. Physical Review Letters. 1994; 73:814–817. [PubMed: 10057546]

Saremi S, Sejnowski TJ. Hierarchical model of natural images and the origin of scale invariance. Proceedings of the National Academy of Sciences. 2013; 110(8):3071–3076.

Schneidman E, Berry MJ II, Segev R, Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature. 2006; 440(7087):1007–1012. [PubMed: 16625187]

Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annual Review of Neuroscience. 2001; 24(1):1193–1216.

Stephens GJ, Mora T, Tkačik G, Bialek W. Statistical thermodynamics of natural images. Physical Review Letters. 2013; 110:018701. [PubMed: 23383852]

Tkačik G, Schneidman E, Berry MJ II, Bialek W. Spin glass models for a network of real neurons. 2009 arXiv:0912.5409.

van Hateren J, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings: Biological Sciences. 1998; 265:359–366.

Wilson K. Problems in physics with many scales of length. Scientific American. 1979; 241:158–179.

Wilson KG, Kogut J. The renormalization group and the ε expansion. Physics Reports. 1974; 12(2): 75–199.

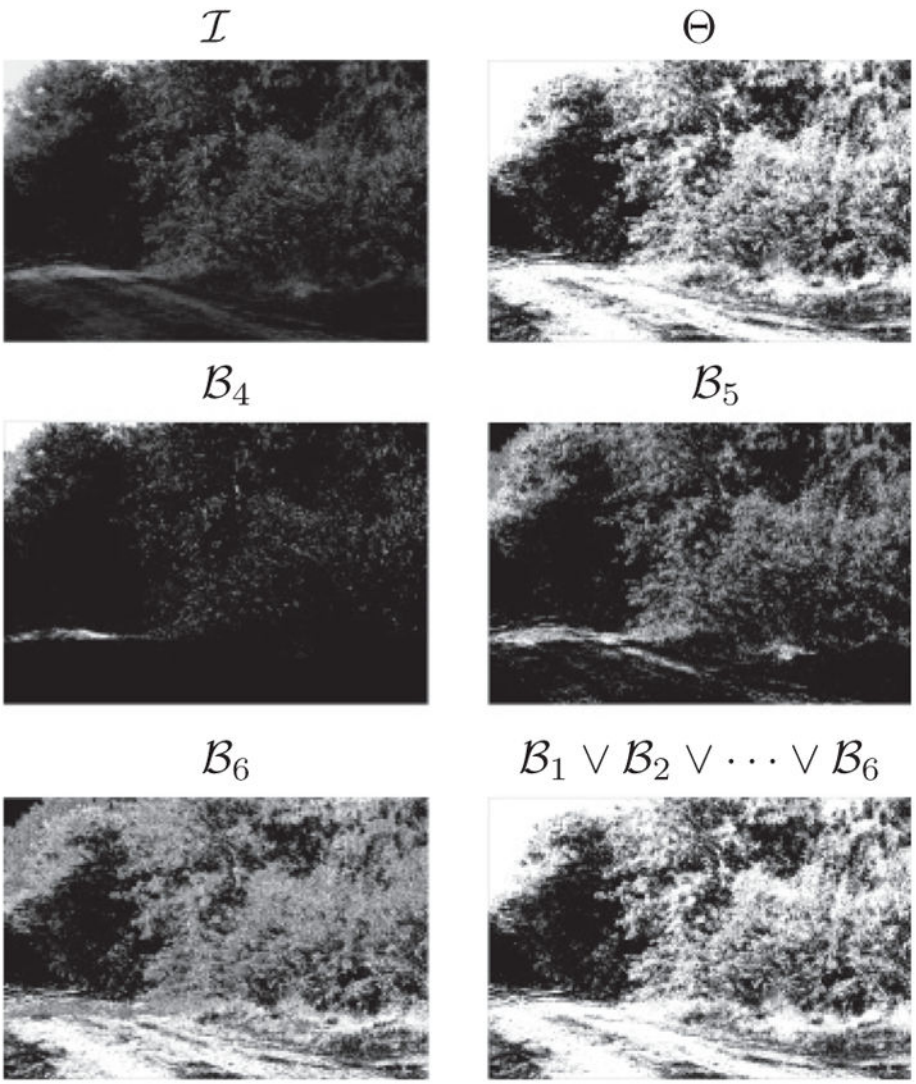

Figure 1.

Example of an image I from the van Hateren database. Its median thresholded image Θ and the bit planes $\{\mathscr{B}_4, \mathscr{B}_5, \mathscr{B}_6\}$ are shown. $\theta(I - \text{median}(I)) \simeq \mathscr{B}_1 \vee \mathscr{B}_2 \vee \ldots \vee \mathscr{B}_6$ is also demonstrated, which is a good approximation for this example as $\mu L - \log_2 \text{median}(I)) = 5.9447$ is close to 6.

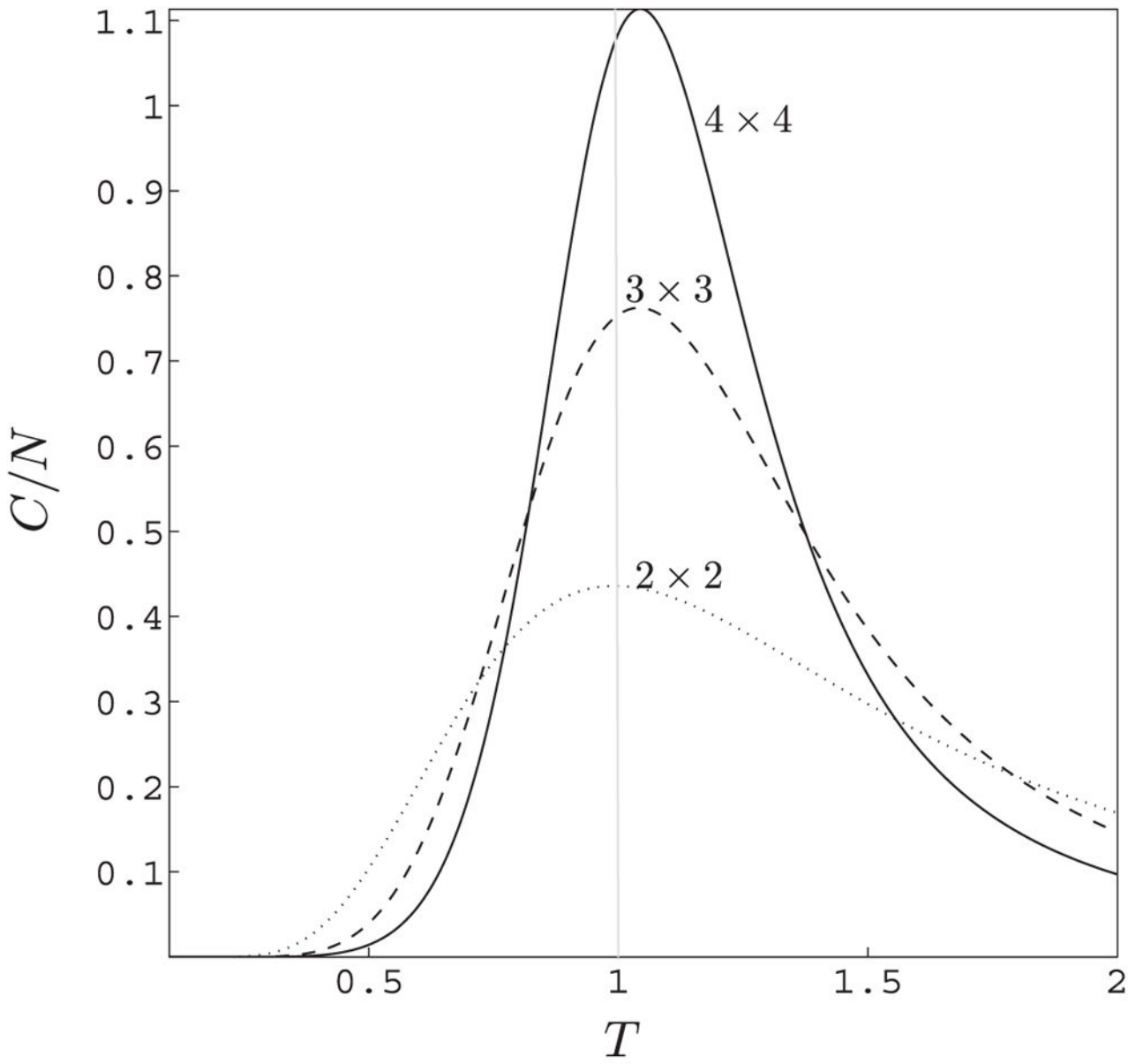

Figure 2.

The specific heat curves *C/N* calculated the ensemble $\Theta = \theta(I - \text{median}(I))$. They are plotted versus temperature *T* for systems of size $N = \{2 \times 2, 3 \times 3, 4 \times 4\}$. The ensemble Θ is at $T = 1$.

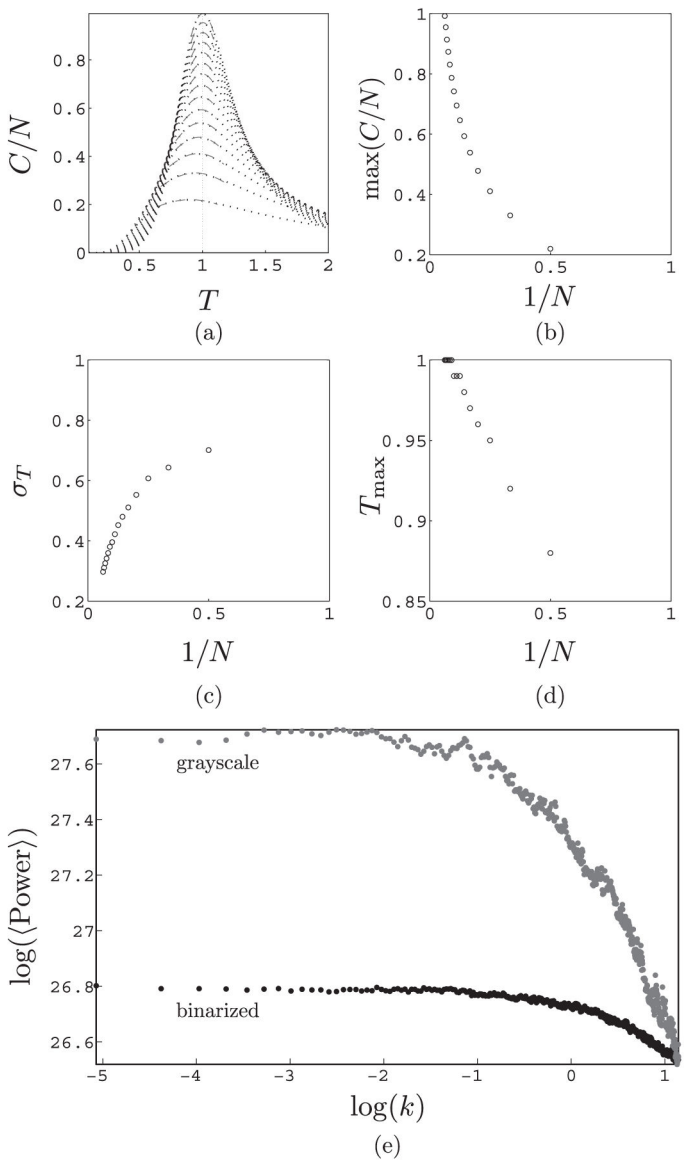

Figure 3.

Studying 1D samples from the ensemble Θ for systems of linear size $N = \{2, 3, \ldots, 16\}$. (a) C/N is plotted versus T. The dashed curves are the gaussian fit around the peak. (b) Finite size scaling. The peak of C/N is plotted versus $1/N$. (c) Here the width σ_T (obtained from the gaussian fit near the peak) is plotted versus $1/N$. (d) The location of the peak is plotted versus $1/N$. Unlike the curves in panels b and c, it saturates in the large N limit. (e) The power spectrum of 1D stripes with 1000 pixels before (in gray) and after binarization (in black) are given. The data points for binarized samples are shifted up by 14.7.

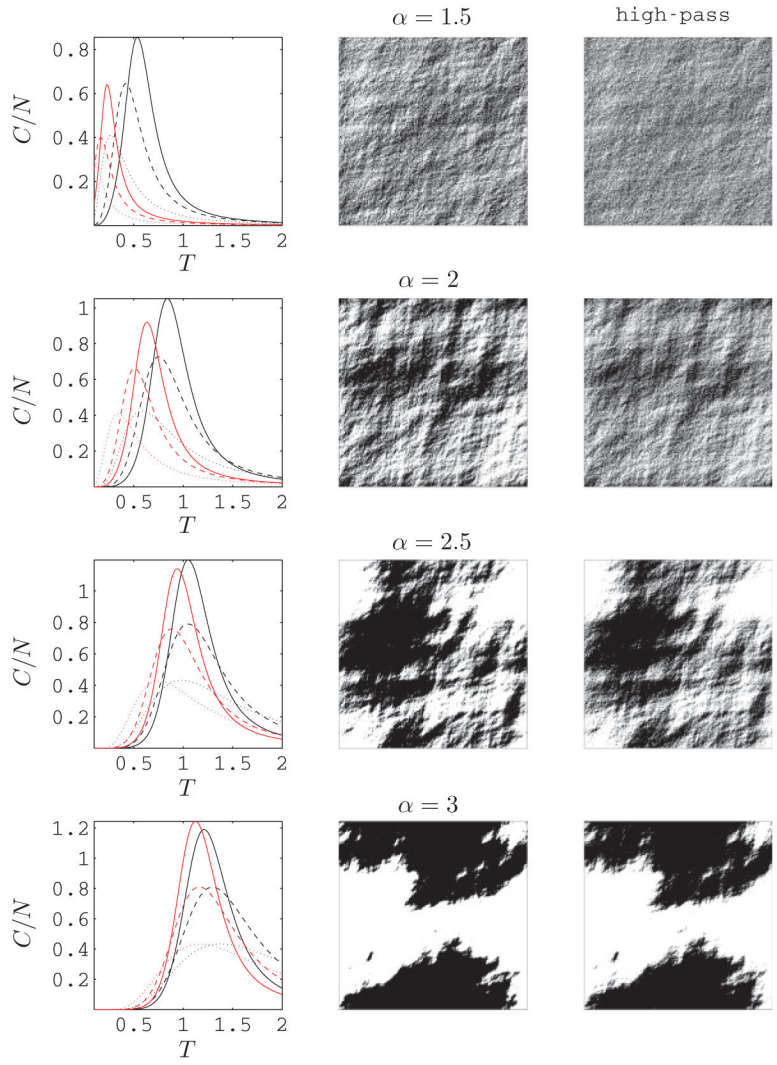

Figure 4.

(Left) Here *C/N* is calculated for the 2D binary pink noise (black curves) and taking samples with sizes $N = \{2 \times 2, 3 \times 3, 4 \times 4\}$. The pink noise ensemble is at $T = 1$. The red curves are the corresponding curves after high-pass-filtering the samples. (Middle) Binary pink noise samples (500 × 500) with the exponent α obtained by thresholding the gaussian pink noise. (Right) High-pass-filtered binary pink noise samples obtained by high-pass-filtering gaussian pink noise before thresholding.