Performance of the majority voting rule in solving the density classification problem in high dimensions

2011 J. Phys. A: Math. Theor. 44 445101

(http://iopscience.iop.org/1751-8121/44/44/445101)

J. Phys. A: Math. Theor. **44** (2011) 445101 (11pp) doi:10.1088/1751-8113/44/44/445101

Performance of the majority voting rule in solving the density classification problem in high dimensions

José Manuel Gómez Soto[1] **and Henryk Fukś**[2]

[1] Unidad Académica de Matemáticas, Universidad Autónoma de Zacatecas, Calzada Solidaridad entronque Paseo a la Bufa, Zacatecas, Zac., Mexico
[2] Department of Mathematics, Brock University, St. Catharines, ON, Canada

E-mail: jmgomezgoo@gmail.com and hfuks@brocku.ca

Received 2 June 2011, in final form 11 August 2011
Published 12 October 2011
Online at stacks.iop.org/JPhysA/44/445101

Abstract
The density classification problem (DCP) is one of the most widely studied problems in the theory of cellular automata. After it was shown that the DCP cannot be solved perfectly, the research in this area has been focused on finding better rules that could solve the DCP approximately. In this paper, we argue that the majority voting rule in high dimensions can achieve high performance in solving the DCP, and that its performance increases with dimension. We support this conjecture with arguments based on the mean-field approximation and direct computer simulations.

PACS number: 05.45.−a
Mathematics Subject Classification: 37B15, 68Q80

(Some figures may appear in colour only in the online journal)

1. Introduction

Classification of initial configurations is a type of computational problem which has been extensively studied in the context of cellular automata (CA) theory. For problems of this type, the initial configuration is the data and the CA evolution rule is the algorithm that processes the data and yields the solution in the form of the final configuration. One of the simplest classification tasks is the so-called density classification problem (DCP). The CA performing this task should converge to a fixed point of all ones if the initial configuration contains more ones than zeros, and to a fixed point of all zeros if the initial configuration contains more zeros than ones.

From the time when Gacs *et al* proposed this problem and its first approximate solution [11], a lot of research effort went into studying this topic. After it was proved by Land and Belew [13] that the perfect two-state rule performing this task does not exist, approximate solutions have been constructed using a variety of methods, including 'complexity engineering'

[7], genetic algorithms [15, 6, 5, 12, 16], genetic programming [2, 9], using basins of attraction [3], analytical formulation [14] and other methods. Additionally, in 1997 Fukś [10] proposed a generalized version of the problem involving two rules and found its exact solution, and later on Capcarrère and Sipper [4] established the necessary conditions to obtain a solution to the DCP involving two rules. Modified versions of the problem allowing different output specifications and different boundary conditions had been considered as well [17]. Most recently, it has been demonstrated that the stochastic version of the DCP can be solved with arbitrary precision [8].

In this paper, we consider a totalistic majority rule in high dimensions. We conjecture that this rule can solve the DCP with increasing accuracy as the dimension increases. We illustrate this claim with numerical experiments in dimensions ranging from 1 to 4, and with some arguments based on the mean-field approximation.

2. Definitions

A binary cellular automaton (BCA) is a dynamical system which evolves in discrete time steps. Let $\mathbb{Z}_L = \{0, 1, \ldots, L-1\}$. We will consider d-dimensional configuration space (space of global states) defined as $\mathcal{A}^{\mathbb{Z}_L^d}$, where $\mathcal{A} = \{0, 1\}$ is a set of allowed cell states, and where we impose periodic boundary conditions, such that all indices of elements of $\mathcal{A}^{\mathbb{Z}_L^d}$ are taken modulo L. Elements of the configuration space are thus hypercubes of binary symbols, and in the special case of $d = 1$, the configuration space is simply a set of binary strings of length L.

Let $\mathcal{N}$ be a finite subset of $\mathbb{Z}_L^d$, to be called a *neighborhood*. For a given $\mathbf{x} \in \mathbb{Z}_L^d$, the set $\mathbf{x} + \mathcal{N}$ will be called a neighborhood of $\mathbf{x}$. Let $\psi : \mathcal{A}^{\mathcal{N}} \to \mathcal{A}$ be called a *local function*. *CA (rule)* is a transformation $\Psi : \mathcal{A}^{\mathbb{Z}_L^d} \to \mathcal{A}^{\mathbb{Z}_L^d}$ defined as

$$\Psi(\mathbf{x})_{\mathbf{n}} = \psi(\mathbf{x}_{\mathbf{n}+\mathcal{N}}), \tag{1}$$

for every $\mathbf{n} \in \mathbb{Z}_L^d$ and $\mathbf{x} \in \mathcal{A}^{\mathbb{Z}_L^d}$. Ψ is sometimes called a *global function*. For a given initial configuration $\mathbf{x} \in \mathcal{A}^{\mathbb{Z}_L^d}$, the set $\{\Psi^t(\mathbf{x})\}_{t=0}^{\infty}$ of consecutive iterates of Ψ will be called an *orbit* of $\mathbf{x}$.

In one dimension ($d = 1$), one often considers the neighborhood of radius r, defined as $\mathcal{N} = \{-r, -r+1, \ldots, r\}$. The global rule Ψ is then defined by a local mapping $\psi : \{0, 1\}^{2r+1} \to \{0, 1\}$, where r is referred to as the radius of the rule. The global function is then defined as

$$\Psi(x)_i = \psi(x_{i-r}, x_{i-r+1}, \ldots, x_{i+r}), \tag{2}$$

for all $x \in \mathbb{Z}_L$ and $i \in \{0, 1, \ldots, L-1\}$. The above neighborhood definition can be easily generalized to higher dimensions. We define *Moore neighborhood* $\mathcal{M}_{r,d}$ of radius r in dimension d as

$$\mathcal{M}_{r,d} = \{-r, -r+1, \ldots, r\}^d. \tag{3}$$

The number of cells in this neighborhood is $(2r+1)^d$.

The DCP can now be stated as follows: given the initial configuration $\mathbf{x} \in \mathcal{A}^{\mathbb{Z}_L^d}$ containing majority of zeros or ones, find the CA rule Ψ such that after sufficiently many iterations of this rule the orbit of $\mathbf{x}$ reaches a homogeneous state where every cell is, respectively, in the state 0 or in the state 1.

The *performance* of a given rule Ψ in performing the density classification is typically defined as follows. Let I denote the number of random initial configurations consisting of N cells, drawn from a symmetric Bernoulli distribution. This means that each initial configuration is generated by setting each of its cells independently to 0 or 1, with the same probability 1/2. Suppose that we iterate the rule Ψ on each of those initial configurations for a maximum of

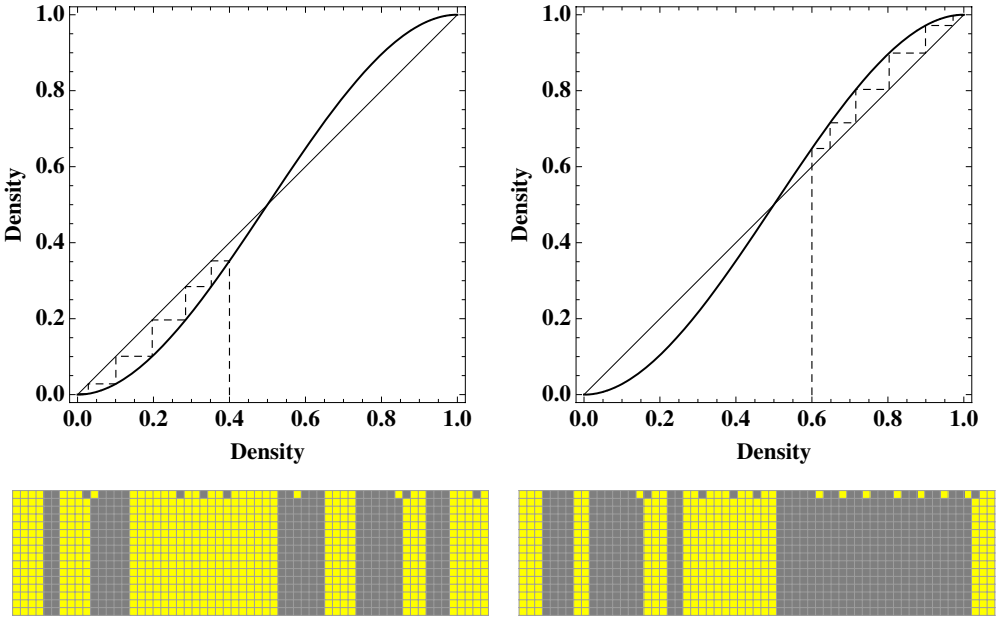

Figure 1. Spatiotemporal diagrams representing orbits of the majority rule 232 in 1D with initial densities 0.4 and 0.6 (bottom) and cobweb plots illustrating iterations of the mean-field map for this rule (top).

$T_{\max}$ time steps. If a configuration with initial density less than 0.5 converges to the fixed point of all zeros, we consider it a successful classification, similarly as when a configuration with initial density greater than 0.5 converges to the fixed point of all ones. In all other cases we consider the classification unsuccessful. The percentage of successful classifications among all I initial conditions will be denoted by $p^I_N(\Psi)$, and called *performance* of the rule.

3. Rule 232

One of the most obvious candidates for a potential solver of DCP is the so-called majority rule. This rule returns 1 if and only if the majority of the cells in the neighborhood have the value 1. In one dimension and $r = 1$, rule 232 is the majority rule using the Wolfram numbering convention. Its local function is defined by $000 \to 0, 001 \to 0, 010 \to 0, 011 \to 1, 100 \to 0$, $101 \to 1, 110 \to 1$ and $111 \to 1$.

Let us consider the mean-field approximation of this rule, which assumes that at time t the expected value of all cells is the same and equal to c_t, and that there is no correlation between sites. Under these assumptions, one can show that $c_{t+1} = f(c_t)$, where $f(x) = 3(1-x)x^2 + x^3$. The function f has three fixed points: 0, 0.5 and 1, as illustrated in figure 1. Two of them (0 and 1) are attracting fixed points while 0.5 is a repelling fixed point. For that reason, if the initial density c_0 is smaller than 1/2, then $\lim_{t\to\infty} c_t = 0$, and if $c_0 > 0.5$, $\lim_{t\to\infty} c_t = 1$.

If the mean-field approximation was correct, the majority rule would solve the DCP problem in 1D. Thus, in a sense, the mean-field behavior is the desired behavior for the potential DCP solver. In one dimension, the mean-field approximation is not very accurate, mainly because of the strong interdependence of individual cells. It is generally known, however, that in higher dimensions the accuracy of the mean-field approximation improves;

thus, one would expect that the performance of the majority rule as a solver of DCP should be better in higher dimensions than in 1D. Before we consider this, however, we need to discuss some properties of one-dimensional generalization of rule 232.

4. Majority rule in one dimension

In the one-dimensional space, the majority rule with radius r is defined as

$$\Psi(x)_i = \text{majority}(x_{i-r}, x_{i-r+1}, \ldots, x_{i+r}). \tag{4}$$

Suppose that we apply this rule iteratively to a binary string of length N with periodic boundary conditions. It has been demonstrated [1] that one of two things will eventually happen: a fixed point or a cycle of period 2 will be reached. By a fixed point we obviously understand a string which is invariant under the action of the majority rule. As it turns out, the dynamics of this rule is dominated by its fixed points, because one can show that the fraction of possible strings of length N that lead to cycles is less than $N^{-1/2}$, thus becoming negligible for large N [1]. This means that orbits of almost all initial strings eventually evolve toward fixed points.

The number of these fixed points has been calculated by Z Agur [1]. For N odd and the radius $r \leqslant (N-1)/2$, the number of fixed points $p(N, r)$ is given by

$$p(N, r) = 2 + 2 \sum_{l=1}^{\lfloor \frac{N}{2(r+1)} \rfloor} \frac{N}{N-2lr} \binom{N-2lr}{2l}. \tag{5}$$

The structure of these points is not difficult to describe. Following [1], consider a binary string $a = a_0 a_1 \ldots a_{N-1}$. Let us define a run as a maximal substring of consecutive bits of equal value. Let k be the number of runs in a, and their sizes be $t_1, t_2, \ldots, t_k$ (since a is periodic, we can choose the origin of the coordinate system to coincide with the beginning of some run). In [1] it has been demonstrated that a string a is a fixed point of the majority rule if and only if, for even k, $t_i \geqslant r+1$ for all i, or, for odd k, $t_i \geqslant r+1$ for $i = 2, \ldots, k-1$ and $t_1 + t_k \geqslant r+1$. For example, if $r = 1$, any string which has no isolated zeros and no isolated ones is a fixed point. Figure 1 (bottom) shows spatiotemporal diagrams of orbits which reach fixed points of this type.

We will make an important observation regarding $p(N, r)$. Namely, $p(N, r)$ decreases with increasing r, eventually reaching the value 2. This happens when $r = (N-1)/2$, that is, when the neighborhood of any site includes all other sites. Looking at the above description of the structure of fixed points we immediately realize that for $r = (N-1)/2$, the condition that t_i must satisfy becomes

$$t_i \geqslant r + 1 = \frac{N+1}{2}, \tag{6}$$

and this is possible only if there is exactly one run—otherwise the total length of all runs would exceed N. This means that the fixed points in this case are all zeros and all ones. Moreover, any configuration reaches one of these fixed points in one iteration, and strings with density less than $1/2$ are mapped to all zeros, while those with density greater than $1/2$ are mapped to all ones. For odd N and $r = (N-1)/2$, therefore, the majority rule is a perfect classifier of densities.

This suggests that perhaps the DCP performance of the majority rule increases with the radius (for fixed N). While we are not able to construct a rigorous proof of this statement, we will offer some experimental evidence. Figure 2 shows the plot of the performance of the majority rule as a function of its radius. We can clearly see that the performance increases with r and reaches 100% when the ratio $(2r+1)/N$ becomes 1. In this case, since we took $N = 10\,001$, this happens when $r = 5000$.

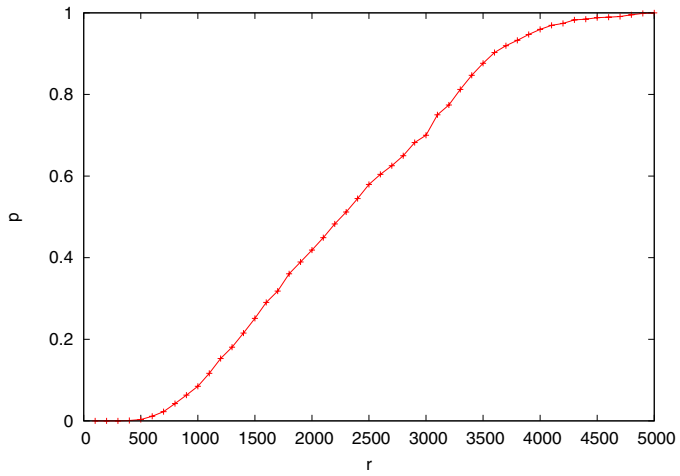

Figure 2. Plot of performance $p_{10\,001}^{10\,000}$ of the 1D majority rule as a function of radius r.

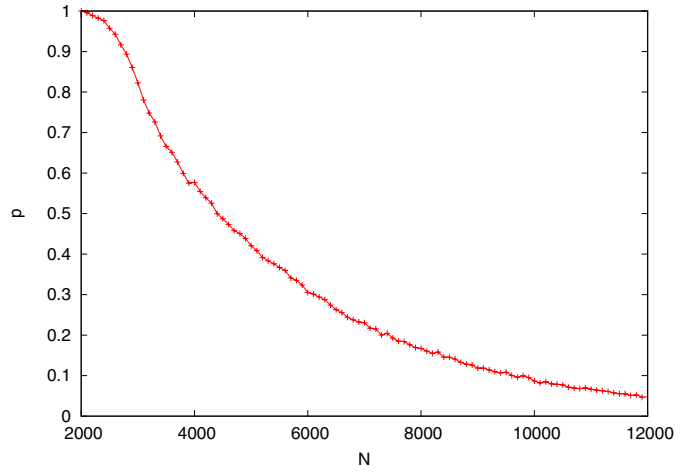

Figure 3. Plot of performance $p_N^{10\,000}$ of the 1D majority rule with radius $r = 1000$ as a function of lattice size N.

One can also ask what happens in the converse case, that is, when the radius r is fixed and the size N of the lattice changes. Figure 3 demonstrates results of such an experiment. It shows how the performance varies with N when the radius is held constant at $r = 1000$. A decrease of performance can clearly be observed.

5. Mean-field approximation of the majority rule

Let us again consider the local function of the majority voting rule with n inputs:

$$f(x_1, x_2, \ldots, x_n) = \text{majority}\{x_1, x_2, \ldots, x_n\}, \tag{7}$$

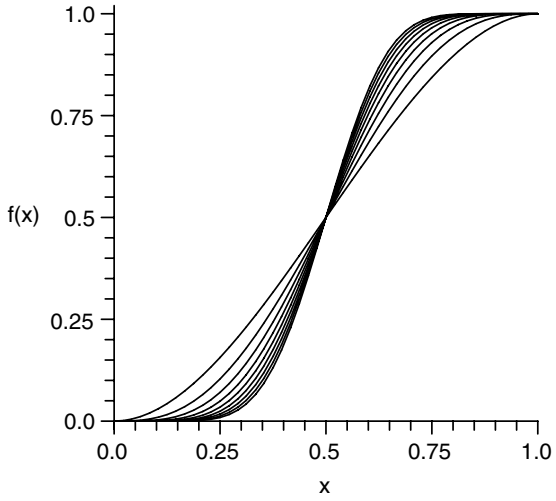

Figure 4. Plots of $f_{\mathrm{MF}}(x)$ for $n = 3, 5, \ldots, 19$.

where n is assumed to be a positive odd integer. This is the form of the local function used in the majority CA regardless of dimension. Of course, the inputs $x_1, x_2, \ldots, x_n$ are arranged linearly only in 1D, but in 2D they form a square array, etc. Details of the arrangement of inputs are irrelevant for considerations in this section; thus, we will ignore them for now.

The mean-field polynomial associated with this function is defined as

$$f_{\mathrm{MF}}(x) = \sum_{(a_1,a_2,\ldots,a_n)\in\{0,1\}^n} f(a_1, a_2, \ldots, a_n) \prod_{i=1}^{n} x^{a_i}(1-x)^{a_i}, \tag{8}$$

where we take $x^{a_i} = 1$ if $a_i = 0$ and $(1-x)^{a_i} = 1$ if $a_i = 0$. Here, x represents the density of ones. Several examples of f_{MF} are shown below:

$n = 1 : f_{\mathrm{MF}}(x) = x$

$n = 3 : f_{\mathrm{MF}}(x) = -2x^3 + 3x^2$

$n = 5 : f_{\mathrm{MF}}(x) = 6x^5 - 15x^4 + 10x^3$

$n = 7 : f_{\mathrm{MF}}(x) = -20x^7 + 70x^6 - 84x^5 + 35x^4$

$n = 9 : f_{\mathrm{MF}}(x) = 70x^9 - 315x^8 + 540x^7 - 420x^6 + 126x^5$

$n = 11 : f_{\mathrm{MF}}(x) = -252x^{11} + 1386x^{10} - 3080x^9 + 3465x^8 - 1980x^7 + 462x^6.$

One can easily show that the following formula for $f_{\mathrm{MF}}(x)$ holds for arbitrary odd n:

$$f_{\mathrm{MF}}(x) = \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} x^{n-i}(1-x)^i. \tag{9}$$

Graphs of f_{MF} for $n = 3, 5, \ldots, 19$ are shown in figure 4. One can see from that figure that for all n, $f_{\mathrm{MF}}(x)$ has stable fixed points at $x = 0$ and $x = 1$, and an unstable (repelling) fixed point at $x = 1/2$. The fixed point at $x = 1/2$ corresponds to the inflection point of $f_{\mathrm{MF}}(x)$, and as n increases, the slope at $x = 1/2$ increases too. In fact, for $n = 3, 5, \ldots, 19$, the values of $f'_{\mathrm{MF}}(1/2)$ form the sequence

$$\frac{3}{2}, \frac{15}{8}, \frac{35}{16}, \frac{315}{128}, \frac{693}{256}, \frac{3003}{1024}, \frac{6435}{2048}, \frac{109\,395}{32\,768}, \frac{230\,945}{65\,536}, \ldots \tag{10}$$

or, in decimal form,

$$1.5,\ 1.875,\ 2.1875,\ 2.460\,9375,\ 2.707\,031\,25,\ 2.932\,617\,188,\ 3.142\,089\,844,$$
$$3.338\,470\,459,\ 3.523\,941\,040,\ \ldots.$$

This sequence is clearly increasing with n, which means that the unstable fixed point at $1/2$ is becoming more and more strongly repelling. Using equation (9), one can demonstrate that

$$f'_{\mathrm{MF}}(1/2) = \binom{n+1}{\frac{n+1}{2}} 2^{-n-1}(n+1), \tag{11}$$

and therefore

$$\lim_{n\to\infty} f'_{\mathrm{MF}}(1/2) = \infty. \tag{12}$$

This means that the mean-field approximation is becoming more and more like a step function, and the slope at $x = 1/2$ tends to infinity.

We can also see from figure 4 that both $x = 0$ and $x = 1$ are strongly attracting fixed points, with basins of attractions, respectively, $[0, 1/2)$ and $(1/2, 1]$. In fact, both these points are superattracting, meaning that $f'_{\mathrm{MF}}(0) = f'_{\mathrm{MF}}(1) = 0$. For superattracting fixed points one can measure the 'strength of attraction' by defining the *degree of superattraction* as the smallest order of a derivative which does not vanish at the fixed point. Since for a given odd n the smallest power of x occurring in $f_{\mathrm{MF}}(x)$ is $(n+1)/2$, the first non-vanishing derivative at $x = 0$ will be of order $(n+1)/2$. Similar reasoning holds for $x = 1$, so that both $x = 0$ and $x = 1$ are superattracting fixed points of degree $(n+1)/2$. Again, this means that the strength of attraction increases with n, just like the strength of repulsion for the fixed point at $x = 1/2$.

If we consider the majority CA rule with fixed radius, and increase the dimensionality of space, the number of cells in the neighborhood will increase with dimension, and the strength of attraction/repulsion of fixed points of the mean-field map will increase too. Since the accuracy of the mean-field approximation improves with increasing dimension, we can expect that the performance of the majority rule in solving the DCP will increase too. The above argument explains the fact that the increasing number of cells in the neighborhood is the main factor which could lead to increased performance of the majority rule in solving the DCP in higher dimensions. But this is not the only factor, as we will see in the next section.

6. Simulation results

The arguments presented in the previous two sections can be verified by direct simulations, iterating the generalized majority voting rule with a Moore neighborhood $\mathcal{M}_{r,d}$ of radius r and dimension d:

$$\Psi(\mathbf{x})_{\mathbf{n}} = \mathrm{majority}(\mathbf{x}_{\mathbf{n}+\mathcal{M}_{r,d}}), \tag{13}$$

where $\mathbf{n} \in \mathbb{Z}_L^d$ and $\mathbf{x} \in \mathcal{A}^{\mathbb{Z}_L^d}$. In what follows N will denote the total number of cells in the lattice, $N = L^d$.

Performing such simulations in various dimensions, we found that for a given fixed radius of the neighborhood, the DCP performance of the majority rule increases with dimension. We performed simulations of the DCP for $N = 10^4$ cells in dimensions 1–4, varying r from 1 to a value for which the performance approaches 100%. Obviously, the radius cannot be increased *ad infinitum*, because at some point the neighborhood becomes large enough to include the entire lattice, and then the performance reaches 100%.

We took the size N of the configuration space hypercube to be approximately 10^4 for all dimensions. Specifically, we used $L = 10001, 100, 22$ and 10, respectively, for $d = 1$,

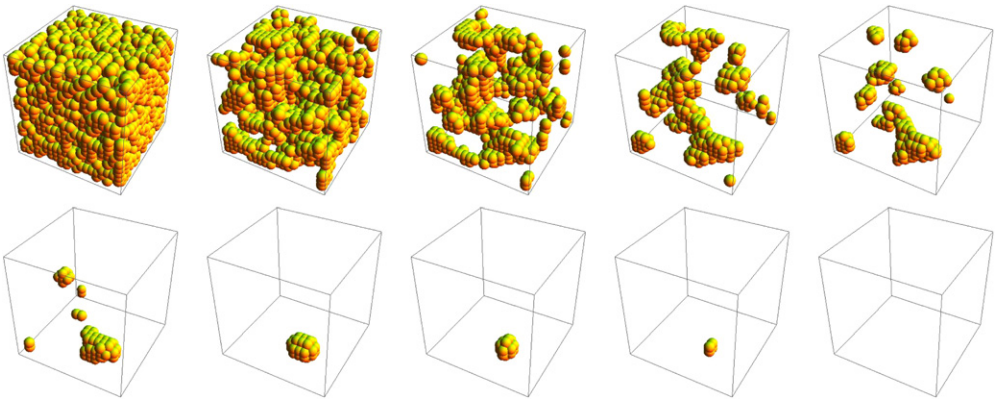

Figure 5. Evolution of the majority rule in 3D starting from the initial configuration with density 0.43.

2, 3, and 4, yielding $N = 10\,001$, 100^2, 22^3 and 10^4. Note that for $d = 3$ this resulted in $N = 22^3 = 10\,648$, that is, slightly more than 10 000, but it was the only choice as the cubic root of 10 000 is not an integer.

Also note that in the case of the 1D majority rule, we used an odd lattice size $N = 10\,001$, to make sure that every configuration can be classified. In higher dimensions this was not possible, and we had to use an even lattice size. Once could ask at this point, what happens when the majority rule is applied to a perfectly symmetric configuration, that is, configuration with equal number of ones and zeros? It turns out that the orbit of such a configuration reaches either the homogeneous fixed point or otherwise some other, non-homogeneous fixed point. The probability of reaching the fixed point of all zeros is the same as the probability of reaching the fixed point of all ones, so there is no breaking of symmetry. Each such case would then be counted as incorrect classification.

In order to avoid the problem we slightly modified the Bernoulli distribution for dimensions $d > 1$. Each time when a configuration was generated with exactly the same number of zeros and ones, we replaced 1 bit of this configuration by its complement, to break the symmetry. The effect of this change on the performance numbers is almost negligible, slightly increasing these numbers, as configurations which would otherwise be always counted as incorrectly classified now have a chance to be classified correctly. This modification, however, has no effect on the overall conclusion of this paper.

We assumed $T_{\max} = 200$ and verified that a further increase of $T_{\max}$ had no detectable influence on the performance numbers. This is because the convergence to the homogeneous state is very fast. An example of a few iterations of the majority rule in 3D is shown in figure 5.

Simulation results are presented in table 1. This table shows performance of the majority rule in dimensions 2–4. Owing to the large number of data points, numbers for $d = 1$ are not shown in the table, they are instead presented in figure 2. The numbers shown in table 1 reveal the same pattern as what has been observed in one dimension, namely the performance increases with the increasing radius r. The convergence toward the perfect performance is faster in higher dimensions, and one could think that this is simply because the number of cells in the neighborhood grows faster in higher dimensions.

Table 1. Performance of the majority rule in dimensions 2–4 for different radii of the neighborhood. See figure 2 for the $d = 1$ case.

	Performance (%)		
Radius r	$d = 2$	$d = 3$	$d = 4$
1	0	9.52	78.3
2	0	84.58	93.68
3	9.06	87.25	97.35
4	45.18	89.01	98.92
5	51.42	90.36	
6	52.04	92.57	
7	54.12	95.57	
8	54.86	98.08	
9	56.3		
10	56.62		
15	62.51		
20	68.37		
25	77.15		
30	85.02		
35	94.36		
40	98.17		

We claim, however, that even if we had two rules (in different dimensions) with the same number of cells in their neighborhoods, the rule with higher dimensions would perform better. Obviously, having exactly the same number of cells in the neighborhood is difficult to achieve, as we would need to find numbers r_1, r_2 and d_1, d_2 such that $(2r + 1)^{d_1} = (2r + 2)^{d_2}$. There are, of course, many such examples, for instance $d_1 = 2$, $r_1 = 4$ with performance 45.15% and $d_2 = 4$, $r_2 = 1$ with performance 98.92%, but it is difficult to see a general trend by considering only such selected cases.

In order to better illustrate the influence of dimension on performance let us define the *connectivity* c of the neighborhood to be the ratio of the number of cells in the neighborhood to the number of cells in the lattice:

$$c = \frac{(2r + 1)^d}{N}. \tag{14}$$

We then plot the performance of the majority rule as a function of connectivity, as shown in figure 6. For each dimension, the graph of the performance as a function of c resembles an S-shaped curve. This is especially visible in one dimension, while in higher dimensions the number of data points is getting smaller; thus, the shape becomes less pronounced. Nevertheless, we can observe that generally a curve corresponding to a higher dimension lies above the curve corresponding to the lower dimension. We could therefore say that if the connectivity (and thus the number of cells in the neighborhood) is fixed while the dimension increases, then the performance increases as well. Since normally there are not too many data points that would have exactly the same c yet belonged to different curves, we can make a stronger statement in a more formal language, as follows.

Conjecture 1. *Let* $\Psi_1 : \mathcal{A}^{\mathbb{Z}_{L_1}^{d_1}} \to \mathcal{A}^{\mathbb{Z}_{L_1}^{d_1}}$ *and* $\Psi_2 : \mathcal{A}^{\mathbb{Z}_{L_2}^{d_2}} \to \mathcal{A}^{\mathbb{Z}_{L_2}^{d_2}}$ *be two majority CA rules with Moore neighborhoods of radii* r_1 *and* r_2*, containing, respectively,* $n_1 = (2r_1 + 1)^{d_1}$ *and* $n_2 = (2r_2 + 1)^{d_2}$ *cells, where* $L_1^{d_1} = L_2^{d_2}$*. If* $d_1 > d_2$ *and* $n_1 \geqslant n_2$*, then the performance of* Ψ_1 *in the DCP is strictly greater than the performance of* Ψ_2*.*

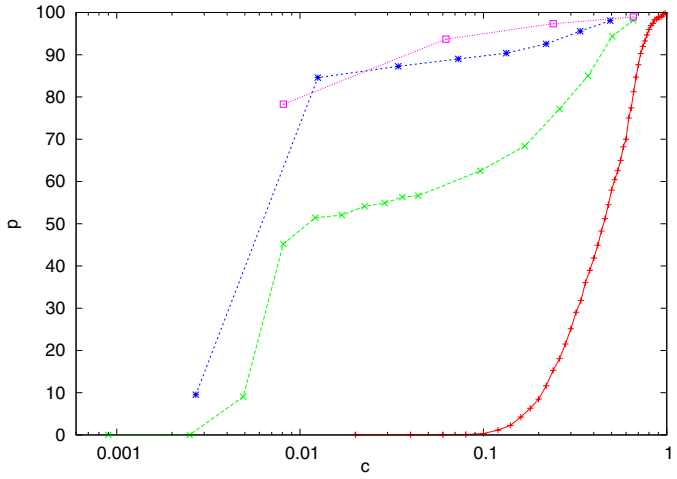

Figure 6. Plot of performance $p_{10\,000}^{10\,000}$ of the majority rule as a function of connectivity of the neighborhood $c = (2r-1)^d/N$, for dimensions $d = 1(+)$, $d = 2(\times)$, $d = 3(\star)$ and $d = 4(\square)$.

7. Conclusion

Both the numerical results and theoretical considerations support the conjecture that the simple majority rule improves its performance in the DCP as the dimensionality of space increases. We hope that this observation may stimulate further research on the DCP in higher dimensions, as well as investigations of high-dimensional versions of other problems of this type, such as, for example, the parity problem or the non-symmetric density classification problem.

Acknowledgments

HF acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of Discovery Grant, and JMGS acknowledges financial support from Research Council of México (CONACYT). This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada.

References

[1] Agur Z 1991 Fixed points of majority rule cellular automata with application to plasticity and precision of the immune system *Complex Syst.* **5** 351–7

[2] Andre D, Bennett F H III and Koza J R 1996 Discovery by genetic programming of a cellular automata rule that is better than any known rule for the majority classification problem *Genetic Programming 1996: Proceedings of the 1st Annual Conference* (Cambridge, MA: MIT Press) pp 28–31

[3] Bossomaier T, Sibley-Punnett L and Cranny T 2000 Basins of attraction and the density classification problem for cellular automata *Virtual Worlds (Lecture Notes in Computer Science* vol 1834*)* ed J-C Heudin (Berlin: Springer) pp 245–55

[4] Capcarrère M S and Sipper M 2001 Necessary conditions for density classification by cellular automata *Phys. Rev.* E **64** 036113

[5] Crutchfield J P and Mitchell M 1995 The evolution of emergent computation *Proc. Natl Acad. Sci.* **92** 10742–6

[6] Das R, Mitchell M and Crutchfield J P 1994 A genetic algorithm discovers particle-based computation in cellular automata *Proc. Int. Conf. on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature* ed Y Davidor, H Schwefel and R Männer (Berlin: Springer) pp 344–53

[7] de Sá P G and Maes C 1992 The Gacs–Kurdyumov–Levin automaton revisited *J. Stat. Phys.* **67** 507–22

[8] Fatès N 2011 Stochastic cellular automata solve the density classification problem with an arbitrary precision *STACS 2011: 28th Int. Symp. on Theoretical Aspects of Computer Science* vol 9, ed Thomas Schwentick and Christoph Dürr (Dagstuhl: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik) pp 284–95

[9] Ferreira C 2001 Gene expression programming: a new adaptive algorithm for solving problems *Complex Syst.* **13** 87–129

[10] Fukś H 1997 Solution of the density classification problem with two cellular automata rules *Phys. Rev.* E **55** R2081–4

[11] Gacs P, Kurdymov G L and Levin L A 1987 One-dimensional uniform array that wash out finite islands *Probl. Peredachi Inf.* **14** 92–8

[12] Juillé H and Pollack J B 1998 Coevolving the ideal trainer: application to the discovery of cellular automata rules *Genetic Programming 1998 Proceeding of the 3rd Annual Conference* ed M H Garzon, D E Golberg, H Iba and R Riolo (Madison, WI: Morgan Kaufmann) pp 519–27

[13] Land M and Belew R K 1995 No perfect two-state cellular automata for density classification exists *Phys. Rev. Lett.* **74** 5148–50

[14] Maiti N S, Munshi S and Chaudhuri P P 2006 An analytical formulation for cellular automata (CA) based solution of density classification task (DCT) *ACRI 2006: Proc. 7th Int. Conf. on Cellular Automata, for Research and Industry (Lecture Notes in Computer Science* vol 4173*)* (Berlin: Springer) pp 147–56

[15] Mitchell M, Crutchfield J P and Hraber P T 1994 Evolving cellular automata to perform computations—mechanisms and impediments *Physica* D **75** 361–91

[16] Morales F J, Crutchfield J P and Mitchell M 2001 Evolving two-dimensional cellular automata to perform density classification: a report on work in progress *Parallel Comput.* **27** 571–85

[17] Sipper M, Capcarrère M S and Ronald E 1998 A simple cellular automaton that solves the density and ordering problems *Int. J. Mod. Phys.* C **9** 899–902