Multiple testing, uncertainty and realistic pictures

Mikhail Langovoy

Mikhail Langovoy, Technische Universiteit Eindhoven,
EURANDOM, P.O. Box 513,
5600 MB, Eindhoven, The Netherlands
e-mail: `langovoy@eurandom.tue.nl`
Phone: (+31) (40) 247 - 8113
Fax: (+31) (40) 247 - 8190
and

Olaf Wittich *

Olaf Wittich, Technische Universiteit Eindhoven and
EURANDOM, P.O. Box 513,
5600 MB, Eindhoven, The Netherlands
e-mail: `o.wittich@tue.nl`
Phone: (+31) (40) 247 - 2499

Abstract: We study statistical detection of grayscale objects in noisy images. The object of interest is of unknown shape and has an unknown intensity, that can be varying over the object and can be negative. No boundary shape constraints are imposed on the object, only a weak bulk condition for the object's interior is required. We propose an algorithm that can be used to detect grayscale objects of unknown shapes in the presence of nonparametric noise of unknown level. Our algorithm is based on a nonparametric multiple testing procedure.

We establish the limit of applicability of our method via an explicit, closed-form, non-asymptotic and nonparametric consistency bound. This bound is valid for a wide class of nonparametric noise distributions. We achieve this by proving an uncertainty principle for percolation on finite lattices.

Keywords and phrases: Image analysis, signal detection, image reconstruction, percolation, noisy image, shape constraints, unsupervised machine learning, spatial statistics, multiple testing.

1. Introduction

Object detection and image reconstruction for noisy images are two of the cornerstone problems in image analysis. In this paper, we continue our work on an efficient method for quick detection of objects in noisy images. Our approach uses mathematical percolation theory.

Detection of objects in noisy images is the most basic problem of image analysis. Indeed, when one looks at a noisy image, the first question to ask is whether there is any object at all. This is also a primary question of interest in such diverse fields as, for example, cancer detection (Ricci-Vitiani et al. (2007)), automated urban analysis (Negri et al. (2006)), detection of cracks in buried pipes

---

*Corresponding author.

(Sinha and Fieguth (2006)), and other possible applications in astronomy, electron microscopy and neurology. Moreover, if there is just a random noise in the picture, it doesn't make sense to run computationally intensive procedures for image reconstruction for this particular picture. Surprisingly, the vast majority of image analysis methods, both in statistics and in engineering, skip this stage and start immediately with image reconstruction.

The crucial difference of our method is that we do not impose any shape or smoothness assumptions on the *boundary* of the object. This permits the detection of nonsmooth, irregular or disconnected objects in noisy images, under very mild assumptions on the object's interior. This is especially suitable, for example, if one has to detect a highly irregular non-convex object in a noisy image. This is usually the case, for example, in the aforementioned fields of automated urban analysis, cancer detection and detection of cracks in materials. Although our detection procedure works for regular images as well, it is precisely the class of irregular images with unknown shape where our method can be very advantageous.

We approached the object detection problem as a hypothesis testing problem within the class of statistical inverse problems in spatial statistics. We were able to extend our approach to the nonparametric case of unknown noise density in Davies et al. (2009) and Langovoy and Wittich (2010a). In Langovoy and Wittich (2009a) and Davies et al. (2009), this density was not assumed smooth or even continuous. It is even possible that the noise distribution is heavy-tailed, see Langovoy and Wittich (2009a), Davies et al. (2009) and Langovoy and Wittich (2010a).

In Langovoy and Wittich (2010b), we gave an algorithmic implementation of our nonparametric hypothesis testing procedure. We also provided a program that can be used for statistical experiments in image processing. This program was written in the statistical programming language R.

We have shown that there is a deep connection between the spatial structure chosen for the discretisation of the image, the type of the noise distribution on the image, and statistical properties of object detection. These results seem to be of independent interest for the field of spatial statistics.

In our previous papers, we considered the case of square lattices in Langovoy and Wittich (2009a) and Langovoy and Wittich (2009b), triangular lattices in Davies et al. (2009) and Langovoy and Wittich (2010a) and even the case of general periodic lattices in Langovoy and Wittich (2010a). In all those cases, we proved that our detection algorithms have linear complexity in terms of the number of pixels on the screen. These procedures are not only asymptotically consistent, but on top of that they have accuracy that grows exponentially with the "number of pixels" in the object of detection. All of our detection algorithms have a built-in data-driven stopping rule, so there is no need in human assistance to stop the algorithm at an appropriate step.

In view of the above, our method can be considered as an unsupervised learning method, in the language of machine learning. This makes our results valuable for the field of machine learning as well. Indeed, we do not only propose an unsupervised method, but also prove the method's consistency and even go as far as to prove the rates of convergence.

In our previous papers we assumed that the original image was black-and-white and that the noisy image was grayscale. In the present paper, we consider the general case where the signal intensity is completely unknown. This intensity

is only assumed to be bounded, but otherwise can vary from pixel to pixel and can be negative.

We propose a multiple testing procedure for detection of grayscale objects of unknown varying intensity in grayscale pictures corrupted by a nonparametric noise that has an unknown distribution. Instead of using a single fixed threshold, we choose a set of thresholds and perform the maximum cluster test from Langovoy and Wittich (2010a) for each of those thresholds. We show in this paper that, under mild model assumptions, if there is an object in the picture, then it is possible to choose a set of thresholds such that we will consistently detect this object, whenever the object can be even in principle detected on the basis of sizes of percolation clusters. This is one of the two parts that are necessary to prove consistency of the new test.

To establish this result, we need to find out when a signal is too weak so that it *cannot* be detected by our approach. We achieve this goal by proving the so-called uncertainty relation for percolation on finite lattices. This is the main probabilistic result of the present paper. An important distinction of our uncertainty result is that it can be formulated as an explicit condition on the noise distribution and the lattice size. Results of this type are very rare both in statistical literature and in image analysis research. To the best of or knowledge, explicit uncertainty bounds were proved only for Gaussian errors (for example, in research on wavelets by Donoho and coauthors). Our uncertainty relation is much stronger, because it holds uniformly over a wide class of nonparametric error distributions.

Since the problem of detection of greyscale objects cannot be solved in complete generality, we might also provide a set of necessary conditions on the image that makes the object detection possible. We plan to give a possible set of those conditions, as well as the full proof of the consistency theorem for our multiple testing method, in our forthcoming paper on the subject.

The paper is organized as follows. Section 2 gives a necessary minimal introduction into the mathematical percolation theory. In Section 3, we review our previous results on detection of black-and-white objects in noisy images. In Section 4, we develop an appropriate model for detection of greyscale objects of unknown varying intensity in greyscale pictures corrupted by nonparametric noise. We prove consistency of the basic building blocks of our multiple testing procedure in Theorem 3. The new uncertainty relation for percolation on finite lattices is proved in Section 5. Theorem 4 of this section is the main mathematical result of the present paper. A new multiple testing procedure for statistical image analysis is proposed in Section 6. Some important results from percolation theory are reviewed in Section 7 of Appendix; this section also contains the proof of the uncertainty relation. Section 8 in Appendix contains the discussion of bounded detector devices.

2. Percolation theory

We start with some basic notions of percolation theory. Let $\mathcal{G}$ be an infinite graph consisting of *sites* $s \in \mathcal{G}$ and *bonds* between sites. The bonds determine the topology of the graph in the following sense: We say that two sites $s, s' \in \mathcal{G}$ are *neighbors* if there is a bond connecting them. We say that a subset $C \subset \mathcal{G}$ of sites is *connected* if for any two sites $s, s' \in C$ there are sites $s_1, ..., s_n$ such

that s and s_1, s_n and s', and s_k and s_{k+1} are neighbors for all $k = 1, ..., n - 1$. Considering *site percolation* on the graph $\mathcal{G}$ means that we consider random configurations $\omega \in \{0, 1\}^{\mathcal{G}}$ where the probabilites are *Bernoulli*

$$P(\omega(s) = 1) = p, \quad P(\omega(s) = 0) = 1 - p$$

independently for each $s \in \mathcal{G}$ where $0 \leq p \leq 1$ is a fixed probability. If $\omega(s) = 1$, we say that the site s is *occupied*.

Then, under mild assumptions on the graph, there is a *phase transition* in the qualitative behaviour of cluster sizes. To be precise, there is a *critical percolation probability* p_c such that for $p < p_c$ there is no infinite connected cluster and for $p > p_c$ there is one.

This statement and the very definition p_c being the location of this phase transition are only valid for infinite graphs. We can not even speak of an infinite connected cluster for finite graphs. However, a qualitative difference of sizes of connected clusters of occupied sites can already be seen for finite graphs, say with $|\mathcal{G}| = N$ sites. In a sense that will be made precise below, the sizes of connected clusters are typically of order $\log N$ for small p and of order N for values of p close to one. This will yield a criterion to infer whether p is close to zero or close to one from observed site configurations. Intuitively, for large enough values of N the distinction between the two regimes is quite sharp and located very near to the critical percolation probability of an associated infinite lattice.

3. Maximum cluster test, consistency and rates of convergence

The signal in our previous papers Langovoy and Wittich (2009a), Davies et al. (2009) and Langovoy and Wittich (2010a) was assumed to be zero-one which corresponds to images with only black and white pixels. In this paper, we will show that the consistency result can be modified to cover also the detection of grayscale objects of unknown intensity. However, first we have to describe our constructions for the basic case.

Let $\mathcal{G}$ denote a planar graph. We think of the sites $s \in \mathcal{G}$ as the pixels of a discretized image and of the graph topology as indicating neighboring pixels. In our aforementioned papers, we considered noisy signals of the form

$$Y(s) = \mathbf{1}_{\mathcal{G}_0}(\mathbf{s}) + \sigma \epsilon(\mathbf{s}) \tag{1}$$

where $\mathbf{1}_{\mathcal{G}_0}$ denotes the indicator function of a subset $\mathcal{G}_0 \subseteq \mathcal{G}$, the noise is given by independent, identically distributed random variables $\{\epsilon(s), s \in \mathcal{G}\}$ with $E\epsilon = 0$ and $V\epsilon = 1$, and $\sigma > 0$ is the noise variance. Thus, σ^{-1} was a measure for the *signal to noise ratio*. We refer to Langovoy and Wittich (2009a) or Langovoy and Wittich (2010a) for a more detailed introduction.

Definition 1. *(The detection problem) For signals of the form (1), we consider the* detection problem *meaning that we construct a test for the following hypothesis and alternative:*

$\mathbf{H}_0$*: $\mathcal{G}_0 = \emptyset$, i.e. there is no signal.*

$\mathbf{H}_1$: $\mathcal{G}_0 \neq \emptyset$, *i.e. there is a signal.*

In our previous work, we constructed tests for the detection problem given in Definition 1 above and computed explicit upper bounds for the *type I and type II errors* under some mild condition on the shape of $\mathcal{G}_0$, called the *bulk condition*. We refer to Langovoy and Wittich (2009a) and Langovoy and Wittich (2010a) for proofs.

The setup is as follows: $\mathcal{T}^{(N)} \subset \mathcal{T}$ denotes the finite triangular lattice consisting of the N^2 sites $s \in \mathcal{T}$ and bonds which are contained in the subset

$$\{z \in \mathbb{C} \,:\, \Re(z) \leq N + \frac{1}{2}, \Im(z) \leq \frac{\sqrt{3}}{2}N\}.$$

By consistency we mean that the test will deliver the correct decision, if the signal can be detected with an arbitrarily high resolution. To be precise, we think of the signal as a subset $G_0 \subset [0,1]^2$ and write

$$G_0^{(N)} := \{(N + 1/2)x + iN\sqrt{3}y/2 \,:\, (x,y) \in G_0\} \subset \mathbb{C}.$$

The model from equation (1) is now depending on N, and given by

$$Y^{(N)}(s) = \mathbf{1}_{\mathcal{G}_0^{(N)}}(s) + \sigma\,\varepsilon(s) \tag{2}$$

where the sites of the subgraph are given by $\mathcal{G}_0^{(N)} = \{s \in \mathcal{T} \,:\, s \in G_0^{(N)}\}$ and the bonds of the subgraph are all bonds in $\mathcal{T}$ that connect two points in $\mathcal{G}_0^{(N)}$.

We apply now the threshold in the following way. First, we let $\tau = 1/2$, and then define

$$Y_\tau^{(N)}(s) = \begin{cases} 1 & ,Y^{(N)}(s) > 1/2 \\ 0 & ,Y^{(N)}(s) \leq 1/2 \end{cases}.$$

We consider the following collection of black pixels

$$\hat{\mathcal{G}}_0^{(N)} := \{s \in \mathcal{T}^{(N)} \,:\, Y_\tau^{(N)}(s) = 1\}. \tag{3}$$

As a side remark, note that one can view $\hat{\mathcal{G}}_0^{(N)}$ as an (inconsistent) pre-estimator of $\mathcal{G}_0^{(N)}$. Now recall that we want to construct a test on the basis of $\hat{\mathcal{G}}_0^{(N)}$, for the hypotheses $\mathbf{H}_0^{(N)} : \mathcal{G}_0^{(N)} = \emptyset$ against the alternative $\mathbf{H}_1^{(N)} : \mathcal{G}_0^{(N)} \neq \emptyset$.

Definition 2. (The Maximum-Cluster Test) *Let $\phi(N)$ be a suitably chosen threshold depending on N. Let the test statistic T be the size of the largest connected black cluster $C \subset \hat{\mathcal{G}}_0^{(N)}$. We reject $\mathbf{H}_0^{(N)}$ if and only if $T \geq \phi(N)$.*

For this test, we have the following consistency result under the assumption that the support of the indicator function satisfies the following very weak type of a shape constraint.

Definition 3. (The Bulk Condition) *We say that the support $\mathcal{G}_0^{(N)}$ of the signal contains a square of side length $\rho(N) \leq N$ if there is a site $s \in \mathcal{G}_0^{(N)}$ such that $s + \mathcal{T}^{(\rho(N))} \subset \mathcal{G}_0^{(N)}$.*

The following consistency result was proved in Langovoy and Wittich (2010a).

Theorem 1. *For the maximum cluster test, we have*

1. *There is some constant $K_0 > 0$ such that for $\phi(N) = K_0 \log N$, we have for the* type I error

$$\lim_{N \to \infty} \alpha(N) = 0.$$

2. *Let $\phi(N)$ be as above. Let the support $\mathcal{G}_0^{(N)}$ of the signal contain squares of side length $\rho(N)$. If $\rho(N) \geq K_0 \log N$, we have for the* type II error

$$\lim_{N \to \infty} \beta(N) = 0.$$

In particular, in the limit of arbitrary large precision of sampling, the test will always produce the right detection result.

The next Theorem strengthens Theorem 1 and delivers the actual rates of convergence for both types of testing errors. It is a remarkable fact that both types of errors in our method tend to zero exponentially fast in terms of the size of the object of interest. See Davies et al. (2009) or Langovoy and Wittich (2010a) for the proof.

Theorem 2. *Suppose assumptions of Theorem 1 are satisfied. Then there are constants $C_1 > 0$, $C_2 > 0$ such that*

1. *The type I error of the maximum cluster test does not exceed*

$$\alpha(N) \leq \exp(-C_2 \phi(N))$$

for all $N > \phi(N)$.

2. *The type II error of the maximum cluster test does not exceed*

$$\beta(N) \leq \exp(-C_1 \rho(N))).$$

for all $N > \rho(N)$.

4. Realistic pictures

Instead of the above idealized model, in the present paper we consider the non-distorted signal of interest to be a bounded function $f \in L^\infty(\mathcal{G})$, i.e. $f(s), s \in \mathcal{G}$ is a collection of pixel intensities and there exists a $c > 0$ such that $|f(s)| < c$ for all $s \in \mathcal{G}$. In the sequel, we will call these functions *realistic pictures*.

The underlying model for the noisy signal is now as in the indicator signal case given by

$$Y(s) = f(s) + \sigma \epsilon(s) \tag{4}$$

and we assume the same properties of the noise as before in Langovoy and Wittich (2010a). More precisely, we assume the following

Noise Properties. For a given graph $\mathcal{G}$, the noise is given by random variables $\{\varepsilon(s) : s \in \mathcal{G}\}$ such that

1. the variables $\varepsilon(s)$ are independent, identically distributed with $E\varepsilon = 0$ and $V\varepsilon = 1$,

2. the noise distribution is *symmetric*,
3. the distribution of the noise is *non-degenerate* with respect to a critical probability p_c meaning that if F denotes the cumulative distribution function of the noise and we define

$$m_c^+ = \inf\{x \in \mathbb{R} \,:\, F(x) \geq 1 - p_c\}, \quad m_c^- = \sup\{x \in \mathbb{R} \,:\, F(x) \leq 1 - p_c\}$$

then we have $m_c^+ = m_c^-$ where we denote the common value by m, and *either*

$$F(m) > \lim_{h \to 0, h > 0} F(m - h), \tag{5}$$

or

$$F'(m) > 0. \tag{6}$$

Furthermore, we assume a *bounded detector device* meaning that only signal intensities $|Y| \leq r$ can be properly displayed, and we assume that this is actually sufficient, i.e. that $|Y| < r$ for the incoming signal. This is explained more closely in the appendix.

The test that has to be performed now reads as $\mathbf{H}_0 : f = 0$ versus the alternative $\mathbf{H}_1 : f \neq 0$ where we assume in an analogous way as before, that $f : [0, 1]^2 \to \mathbb{R}$ is a bounded continuous function. Thus, in a a similar fashion as before, we construct tests for different resolutions, namely for the hypotheses $\mathrm{H}_0^{(N)} :$ $f^{(N)} = 0$ against the alternatives $\mathrm{H}_1^{(N)} : f^{(N)} \neq 0$ where the discretized function $f^{(N)} : \mathcal{T}^{(N)} \to \mathbb{R}$ is given by

$$f^{(N)}(s) = f(x, y), \quad s = \left(N + \tfrac{1}{2}\right) x + i \tfrac{\sqrt{3}}{2} N y. \tag{7}$$

and the corresponding signal is given by

$$Y^{(N)}(s) = f^{(N)}(s) + \sigma \, \varepsilon(s).$$

We now have to slightly modify the test, in particular since we do not have any information about the signal strength. This is the main difference to the situation with the indicator function and also the main reason to introduce a *bounded detector device*. By that property (and assuming as explained before that the intensity scale provided by the detector is actually sufficient to properly display the signal, or – likewise – if we condition on that event) we have a compact scale of thresholds that has to be explored.

Let now $\tau > 0$ and $\mathcal{G}_{\tau,+}^{(N)} \subset \mathcal{T}^{(N)}$ denote the *super level set*

$$\mathcal{G}_{\tau,+}^{(N)} := \{s \in \mathcal{T}^{(N)} \,:\, Y^{(N)}(s) \geq \tau\}$$

and $\mathcal{G}_{\tau,-}^{(N)} \subset \mathcal{T}^{(N)}$ denote the *sub level set*

$$\mathcal{G}_{\tau,-}^{(N)} := \{s \in \mathcal{T}^{(N)} \,:\, Y^{(N)}(s) \leq -\tau\}.$$

Assume furthermore, that the bounded detector device under consideration has range $r > 0$. As a threshold, we use the same $\phi(N) = K_0 \log N$ as in Theorem 1.

We attempt to do signal detection using the following test statistics

$$
\begin{aligned}
T_+^{(N)}(a) &:= \max\{|C| : C \subset \mathcal{G}_{a,+}^{(N)} \text{ black cluster}\}, \\
T_-^{(N)}(a) &:= \max\{|C| : C \subset \mathcal{G}_{a,-}^{(N)} \text{ black cluster}\}
\end{aligned}
\tag{8}
$$

where $a \in [0, r/2]$. It is immediate that we have the following properties as in the case of indicator functions.

Lemma 1. *Under the null hypothesis, the probabilities that a pixel is erroneously marked black are*

1. $p_E = P(s \in \mathcal{G}_{a,+}^{(N)}) = P(\varepsilon \geq a/\sigma) < 1/2 = p_c,$
2. $p_E = P(s \in \mathcal{G}_{a,-}^{(N)}) = P(\varepsilon \leq -a/\sigma) < 1/2 = p_c$

and hence subcritical.

Lemma 2. (i) *Let $Q_+^{(N)} \subset \{f^{(N)} \geq a\}$ be a square. Then we have for all $s \in Q_+^{(N)}$ that*

$$
p_B = P(s \in \mathcal{G}_{a/2,+}^{(N)}) = P(\varepsilon \geq -a/2\sigma) > 1/2 = p_c.
$$

(ii) *Let $Q_-^{(N)} \subset \{f^{(N)} \leq -a\}$ be a square. Then we have for all $s \in Q_-^{(N)}$ that*

$$
p_B = P(s \in \mathcal{G}_{a/2,-}^{(N)}) = P(\varepsilon \leq a/2\sigma) > 1/2 = p_c.
$$

By these two lemmas, we see that for the test statistics considered above, we are essentially in the same situation as we were in Davies et al. (2009) and Langovoy and Wittich (2010a). Both previous lemmas were valid without change if we would consider the respective models

$$
\begin{aligned}
Y_+^{(N)} &= \mathbf{1}_{\{f^{(N)} \geq a\}} + \frac{\sigma}{a}\varepsilon, \\
Y_-^{(N)} &= \mathbf{1}_{\{f^{(N)} \leq -a\}} + \frac{\sigma}{a}\varepsilon
\end{aligned}
$$

for suitably chosen indicator functions. That implies, we may draw the following conclusion by applying exactly the same proof as in Theorem 1.

Theorem 3. *For the test statistics considered above, we have:*

1. *There is some constant $K_0 > 0$ such that for $\phi(N) = K_0 \log N$, we have under the null hypothesis*

   $$
   \lim_{N \to \infty} P(T_+^{(N)}(a) \geq \phi(N)) = \lim_{N \to \infty} P(T_-^{(N)}(a) \geq \phi(N)) = 0.
   $$

   For K_0, we may use the same choice as in Theorem 1.
2. *Let $\phi(N)$ be as above. Let $Q_+^{(N)} \subset \{f^{(N)} \geq a\}$ contain squares of side length $\rho(N)$. If If $\rho(N) \geq K_0 \log N$, we have*

   $$
   \lim_{N \to \infty} P(T_+^{(N)}(a/2) \leq \phi(N)) = 0.
   $$

3. Let $\phi(N)$, $\rho(N)$ be as above. Let $Q_-^{(N)} \subset \{f^{(N)} \leq -a\}$ contain squares of side length $\rho(N)$. Then we also have

$$\lim_{N \to \infty} P(T_-^{(N)}(a/2) \leq \phi(N)) = 0.$$

In particular, the test statistic associated to the correct scale parameter $a/2$ will asymptotically always produce the right detection result.

At first sight, the situation seems rather similar as for indicator functions in Theorem 1. However, it is completely different in the sense that the consistency result only holds if we pick the right signal strength in advance. We might be able to overcome this problem by considering a scale of tests for some positive $a > 0$.

5. Uncertainty

It is intuitively clear that, for principal reasons, it is not possible to detect a signal with arbitrarily small signal to noise ratio on a lattice of finite size, no matter which method is used for detection. However, for every particular method, it is very difficult to provide a "horizon of consistency" in explicit form. Results of this type a very rare in hypothesis testing, image analysis or machine learning. Typically, one proves those results in special cases like Gaussian errors.

In this section, we provide an explicit, closed-form, non-asymptotic and non-parametric consistency bound for our method. This bound is valid for a wide class of nonparametric noise distributions and is given in Theorem 4.

Recall from the proof of Theorem 1 that the constant K_0 in the threshold was given by the inequality

$$K_0 = 2C > \lambda(p_E)^{-1}$$

where p_E is the (subcritical) probability under the null hypothesis that a pixel is erroneously marked black and $\lambda(p_E)$ is the constant from the Aizenman - Newman theorem, see Langovoy and Wittich (2010a). Thus, we have to begin by finding a proper estimate of $\lambda(p)$ for a subcritical p.

The classical Aizenman-Newman theorem reads as follows.

Proposition 1. (Aizenman-Newman Theorem) *Consider percolation with subcritical probability $p < p_c = 1/2$ on the infinite triangular lattice $\mathcal{T}$. Then there is a constant $\lambda(p) > 0$ depending on p such that*

$$P(|C| \geq n) \leq e^{-n\,\lambda(p)} \tag{9}$$

for all $n \geq 1$ where C denotes the black cluster containing the origin.

Remark. Please note that we use asymptotics-oriented estimates to prove statements about finite lattices. For instance in the case of (10) below, these estimates are not the best possible. So we can by no means expect that the bound in Theorem 4 is sharp. But it is good enough to serve as an illustration of the basic principle.

In the sequel, $\chi(p)$ denotes the expected size of the cluster containing $0 \in \mathcal{T}$ in the infinite lattice depending on the subcritical occupation probability $p < p_c = 1/2$.

Lemma 3. *For the infinite triangular lattice, we have*

$$\chi(p) \leq \frac{1}{18} |p - p_c|^{-1}. \tag{10}$$

Proof: See Appendix 7.3. ∎

By the Aizenman - Newman Theorem (Proposition 1), we obtain an upper bound for this expectation value by

$$\chi(p) = \sum_{n \geq 1} P(|C| \geq n) \leq \sum_{n \geq 1} e^{-n\lambda(p)} = \frac{e^{-\lambda(p)}}{1 - e^{-\lambda(p)}}.$$

Thus, we have

$$\lambda(p) \leq -\log\left(\frac{\chi(p)}{1 + \chi(p)}\right) = \log\left(1 + \frac{1}{\chi(p)}\right). \tag{11}$$

Combining these two results yields

Lemma 4. *We have*

$$\lambda(p)^{-1} \geq \frac{1}{\log\left(1 + 18 |p - p_c|\right)}.$$

Proof: (10) together with (11) implies

$$\lambda(p) \leq \log\left(\frac{1 + \chi(p)}{\chi(p)}\right) = \log\left(1 + \frac{1}{\chi(p)}\right) \leq \log\left(1 + 18 |p - p_c|\right).$$

∎

This implies that an intrinsic feature of the procedure is the following form of *uncertainty: By our procedure, we cannot – even in principle – detect signals with arbitrary low signal to noise ratio on a finite lattice of fixed size.*

Of course, it is clear that something like the above statement is valid for any statistical testing procedure. Therefore, something similar is also valid for signal detection. An important distinction of our uncertainty result is that we can give an explicit condition on the noise level and the lattice size, such that this condition implies that our test *does not* work. Results of this type are very rare both in statistical literature and in image analysis research. To the best of or knowledge, explicit uncertainty bounds were proved only for Gaussian errors (for example, in research on wavelets by Donoho and coauthors). Our uncertainty relation is much stronger, because it holds irrespectively of the actual noise distribution, uniformly over a wide class of nonparametric error distributions.

To be precise, we consider again the threshold $\phi(N) = K_0 \log N$ and the fact that in the proof of Theorem 1, we had to choose $K_0 = 2C > 2\lambda(p_E)^{-1}$. That implies together with Lemma 4 that

$$\phi(N) = K_0 \log(N) > \lambda(p)^{-1} \log N^2 \geq \frac{\log N^2}{\log\left(1 + 18 |p_E - p_c|\right)}. \tag{12}$$

But that means, that for values of p which are very close to the critical probability, the threshold $\phi(N)$ may exceed the lattice site N^2 and our method breaks down. To be precise, we have the following statement.

Proposition 2. *If the lattice size N^2 is fixed, the threshold $\phi(N)$ is larger than the lattice size, and therefore, the null hypothesis will never be rejected, if we have*

$$|p_E - p_c| < \frac{1}{18} \left\{ \left(N^2\right)^{\frac{1}{N^2}} - 1 \right\}.$$

Proof: By (12), we have $\phi(N) > N^2$ if

$$\frac{\log N^2}{\log\left(1 + 18\,|p_E - p_c|\right)} > N^2.$$

$\blacksquare$

Finally, we want to relate this statement directly to the signal strength. Thus, if $|f^{(N)}| \le a$ we say that the *signal to noise ratio* is given by $\rho = a/\sigma$. Let us now assume that the distribution function of the noise F is continuous at zero. Then

$$|p_E - p_c| = \frac{1}{2} - p_E = P(0 < \varepsilon < a/\sigma) = F(\rho) - F(0)$$

and we finally obtain

Theorem 4. (Uncertainty) *Assume that the distribution function of the noise is continuous at zero. A signal $f^{(N)}$ with $|f^{(N)}| \le a$ and signal to noise ratio $\rho = a/\sigma$ can only be detected by our method, if*

$$\frac{P(0 < \varepsilon < \rho)}{(N^2)^{\frac{1}{N^2}} - 1} > \frac{1}{18}$$

that means if either the lattice size is sufficiently large or the signal to noise ratio is sufficiently small.

Remark. (i) Note that this statement only means that – as a matter of principle – we can not detect signals of arbitrarily small strength on a finite lattice of a given size. That does not at all mean that detection of signals that respect the bound above is automatically possible in an effective way. Topics like type I and type II error are not at all touched by the uncertainty bound. In other words, from the uncertainty relation we can derive only a necessary condition for the signal to be detectable via our method. Usually this condition will not be sufficient.

(ii) From studying the behavior of the function

$$s(x) = \frac{1}{18}\left(e^{-x\ln x} - 1\right)$$

on the unit interval, we see that the bound is always fulfilled if $P(0 < \varepsilon < \rho) > 0.25 \approx \max_{x \in [0,1]} s(x)$.

The proof of Theorem 4 consists of a simple reformulation of the preceding proposition and is therefore omitted. However, we still have to justify why we

use the word *uncertainty* in connection with this statement. This discussion can only be purely informal. The analogy simply is that a function of the signal to noise ratio times another function of the lattice size have to exceed a certain value for a signal to be detectable. Otherwise, the signal is virtually not existing. A weaker version of the statement might provide another argument: There is some number $M > 0$ such that

$$s(x) \leq M\sqrt{x}$$

for all $x \in [0, 1]$. If we assume now that F has a continuous and sufficiently smooth density f with $f''(0) < 0$, we have the weaker statement that the signal can be detected only if

$$f(0)\rho \geq F(\rho) - F(0) > M\sqrt{1/N^2}$$

or, if

$$N\rho > \frac{M}{f(0)}. \tag{13}$$

Thus, for a signal to be detectable, the product of two conjugate parameters may not exceed a bound given by the circumstances. Otherwise, the signal is not detectable, even in principle.

6. Multiple testing for realistic pictures

By the uncertainty principle, we obtain a minimal threshhold value below which it does not make any sense to try to detect a signal. So there is a natural lower bound τ_0 for a threshold. The upper bound is provided by the size r of the bounded detector device. That means, the range of intensities of detectable signals is $[-r, -\tau_0] \cup [\tau_0, r]$. Thus, if f is the signal, and we assume bulk conditions for the super-level sets as in Corollary 1, taking into account the the simple *monotonicity property* that $a > a'$ implies $\mathbf{1}_{\{f^{(N)} \geq a\}} \leq \mathbf{1}_{\{f^{(N)} \geq a'\}}$, we will certainly be able to consistently detect an object (if the object can be potentially detected on the basis of percolation clusters), via the following scheme:

1. Consider the threshold scheme

$$a_k = 2^{-k}r, k = 1, ..., N.$$

2. Beginning with $a = a_1$, calculate the test statistics $T_-^{(N)}(a)$, $T_+^{(N)}(a)$. Terminate, if either the null hypothesis is rejected (at a properly adjusted level, if necessary) or if you reach a_k with $k \geq \log(r/\tau_0)$.

It can be shown that, under certain conditions on f and σ, we would have to repeat the maximum cluster test at most $O(\log N)$ times. Since each repetition of the maximum cluster test takes $O(N^2)$ operations, the new multiple testing procedure is going to take at most $O(N^2 \log N)$ operations overall. Since the input size is N^2, this implies that under some conditions our initial procedure (of linear complexity) slows down by a logarithmic factor. Asymptotically, this is only slightly slower than the original test, but the new test is adaptive with respect to the unknown image color intensity.

A point that needs to be addressed carefully here is the probability of false rejection of the null hypothesis. Indeed, we perform here not a single test, but a collection of up to $O(\log N)$ tests, and results of those tests are *not* independent. This is a basic question that always occurs in the field of multiple testing. Luckily for us, for each of the thresholds a_k the direct analog of Theorem 2 holds: the type I error of any single test tends to zero exponentially fast, while the power tends exponentially fast to one. Moreover, our tests are "monotonous" with respect to the threshold value (since the maximum cluster size is an increasing event). This also implies that we have to pay attention only to those thresholds a_k where at least one of the level sets $\mathcal{G}_{\tau,-}^{(N)}$ and $\mathcal{G}_{\tau,+}^{(N)}$ doesn't contain black clusters crossing the whole screen. Using those properties, we will be able to combine the results of not more than $O(\log N)$ tests $T_{-}^{(N)}(a_k)$ and $T_{+}^{(N)}(a_k)$ and get a unique decision out of them, while keeping the type I error of the multiple test controlled. We plan to present those results in succeeding papers.

Acknowledgments. The authors would like to thank the EURANDOM Report Series reviewers for carefully reading this manuscript.

References

P. L. Davies, M. Langovoy, and O. Wittich. Randomized algorithms for statistical image analysis based on percolation theory. *Submitted*, 2009.

C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971. ISSN 0010-3616.

Geoffrey Grimmett. *Percolation*, volume 321 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1999. ISBN 3-540-64902-6.

Harry Kesten. *Percolation theory for mathematicians*, volume 2 of *Progress in Probability and Statistics*. Birkhäuser Boston, Mass., 1982. ISBN 3-7643-3107-0.

M. Langovoy and O. Wittich. *Detection of objects in noisy images and site percolation on square lattices*. EURANDOM Report No. 2009-035. EURANDOM, Eindhoven, 2009a.

M. Langovoy and O. Wittich. *Robust nonparametric detection of objects in noisy images*. EURANDOM Report No. 2010-049. EURANDOM, Eindhoven, 2010a.

M. Langovoy and O. Wittich. *Computationally efficient algorithms for statistical image processing. Implementation in R*. EURANDOM Report No. 2010-053. EURANDOM, Eindhoven, 2010b.

M. A. Langovoy and O. Wittich. Detection of objects in noisy images and site percolation on square lattices. *Submitted*, 2009b.

M. Negri, P. Gamba, G. Lisini, and F. Tupin. Junction-aware extraction and regularization of urban road networks in high-resolution sar images. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(10):2962–2971, Oct. 2006. ISSN 0196-2892. .

Lucia Ricci-Vitiani, Dario G. Lombardi, Emanuela Pilozzi, Mauro Biffoni, Matilde Todaro, Cesare Peschle, and Ruggero De Maria. Identification and

expansion of human colon-cancer-initiating cells. *Nature*, 445(7123):111–115, Oct. 2007. ISSN 0028-0836.

Sunil K. Sinha and Paul W. Fieguth. Automated detection of cracks in buried concrete pipe images. *Automation in Construction*, 15(1):58 – 72, 2006. ISSN 0926-5805. .

M. F. Sykes and J. W. Essam. Exact critical percolation probabilities for site and bond problems in two dimensions. *J. Mathematical Phys.*, 5:1117–1127, 1964. ISSN 0022-2488.

Appendix.

7. Some facts from percolation theory

In this section, we collect some basic statements and techniques from the theory of percolation. In particular, we are going to prove the inequality (10) which is basic for the introduction of *uncertainty principle.*

7.1. FKG and BK inequality

Recall the partial ordering

$$\omega_1 \preceq \omega_2 \quad :\Longleftrightarrow \quad \omega_1(s) \leq \omega_2(s) \text{ for all } s \in \mathcal{T}.$$

on the set $\Omega = \{0,1\}^{\mathcal{T}}$ of all *percolation configurations* from Definition **??** and that an *event* $A \subset \Omega$ is *increasing* if we have

$$\mathbf{1}_A(\omega_1) \leq \mathbf{1}_A(\omega_2)$$

for the corresponding indicator variable whenever $\omega_1 \preceq \omega_2$.

The FKG inequality was already stated before and is just added here another time for completeness.

Proposition 3. (FKG inequality) *If A and B are both increasing (or both decreasing) events, then we have*

$$P(A \cap B) \geq P(A)\,P(B).$$

Proof: Fortuin et al. (1971) ∎

Let $\mathcal{G} \subset \mathcal{T}$ be a *finite* subgraph and

$$\mathcal{F}_{\mathcal{G}}\sigma(\{0,1\}^{\mathcal{G}}) \subset \sigma(\{0,1\}^{\mathcal{T}}) =: \mathcal{F}_{\mathcal{T}}$$

the sub sigma - algebra associated to the percolation configurations on $\mathcal{G}$ (in the canonical version). Let now $A, B \in \mathcal{F}_{\mathcal{G}}$ be two increasing events. We define the support of $\omega \in \{0,1\}^{\mathcal{G}}$ to be

$$\mathrm{supp}\,\omega := \{s \in \mathcal{G} \,:\, \omega(s) = 1\}.$$

and for a subset $H \subset \text{supp}\,\omega$, we write

$$\omega|_H := \left\{ \begin{array}{ll} 1 & s \in H \\ 0 & \text{el}se \end{array} \right. .$$

Definition 4. *Let A, B be as above. The event $A \circ B$ that A and B occur disjointly is given by*

$$A \circ B := \{\omega \in \{0,1\}^{\mathcal{T}} : \exists_{H(\omega) \in \text{supp}\,\omega}\, \omega|_{H(\omega)} \in A, \omega|_{\text{supp}\,\omega - H(\omega)} \in B\}.$$

The BK inequality now reads as follows.

Proposition 4. (BK inequality) *Let $A, B \in \mathcal{F}_{\mathcal{G}}$ be increasing events. Then*

$$P(A \circ B) \le P(A)\, P(B).$$

Proof: Grimmett (1999), p. 38 ff. ∎

7.2. Russo's formula

Let $s \in \mathcal{T}$ be a site. We consider the involution $j_s : \{0,1\}^{\mathcal{T}} \to \{0,1\}^{\mathcal{T}}$ given by

$$j_s(\omega)(s') := \left\{ \begin{array}{ll} \omega(s') & s' \neq s \\ 1 - \omega(s') & s' = s \end{array} \right. .$$

From this definition, we see that the configuration space is a disjoint union $\{0,1\}^{\mathcal{T}} = \Omega(s)^+ \cup j_s\Omega(s)^+$, where

$$\Omega(s)^+ := \{\omega \in \{0,1\}^{\mathcal{T}} : \omega(s) = 1\}.$$

Definition 5. (Pivotal sites) *Let $\mathcal{G} \subset \mathcal{T}$ be a finite subgraph and $A \subset \mathcal{F}_{\mathcal{G}}$ be an increasing event. The event* the site s is pivotal for A *is given by*

$$\text{Piv}(A,s) := \{\omega \in \{0,1\}^{\mathcal{T}} : \mathbf{1}_A(\omega) \neq \mathbf{1}_A \circ j_s(\omega)\}.$$

Russo's formula is a statement about how the probability of a certain event changes if the individual site occupation probability p is changed. We denote by $P_p(A)$ the probability of the event A if this probability is p and by

$$N_A := \sum_{s \in \mathcal{G}} \mathbf{1}_{\text{Piv}(A,s)}$$

the number of pivotal elements for A.

Proposition 5. (Russo's formula) *Let $\mathcal{G} \subset \mathcal{T}$ be a finite subgraph and $A \subset \mathcal{F}_{\mathcal{G}}$ be an increasing event. Then*

$$\frac{d}{dp} P_p(A) = E_p N_A. \tag{14}$$

Proof: (i) First of all, since A is increasing and $\omega \preceq j_s(\omega)$ for all $\omega \in \Omega(s)^+$, we have on the set $\Omega(s)^+$

$$(\mathbf{1}_A - \mathbf{1}_A \circ j_s)(\omega) = \left\{ \begin{array}{ll} 1 & \omega \in \text{Piv}(A,s) \cap \Omega(s)^+ \\ 0 & \text{else} \end{array} \right. = \mathbf{1}_{\text{Piv}(A,s) \cap \Omega(s)^+}.$$

(ii) In the sequel, we write $\Omega(s)^- = j_s\Omega(s)^+$. Let $\mathbb{P}|_{\Omega(s)^-}$ denote the restriction of the measure to $\Omega(s)^-$. Then, the image measure under j_s is a measure on $\Omega(s)^+$ with density

$$\frac{d\mathbb{P}|_{\Omega(s)^-} \circ j_s}{d\mathbb{P}} = \frac{\mathbb{P}(\Omega(s)^+)}{\mathbb{P}(\Omega(s)^-)}.$$

That implies

$$
\begin{aligned}
E_p\left[\mathbf{1}_A|\Omega(s)^-\right] &= \frac{\int_{\Omega(s)^-}\mathbf{1}_A(\omega)\mathbb{P}(d\omega)}{\mathbb{P}(\Omega(s)^-)} = \frac{\int_{\Omega(s)^+}\mathbf{1}_A \circ j_s(\omega')\mathbb{P}\circ j_s(d\omega')}{\mathbb{P}(\Omega(s)^-)} \\
&= E_p\left[\mathbf{1}_A \circ j_s|\Omega(s)^+\right].
\end{aligned}
$$

(iii) Now let $p' > p$ and denote by $E_{p',s}$ the expectation with respect to the product measure $\mathbb{P}_s$ with marginals

$$
\mathbb{P}_s(\omega(s') = 1) = \left\{
\begin{array}{ll}
p' & s' = s \\
p & \text{else}
\end{array}
\right. .
$$

Thus

$$
\begin{aligned}
& P_{p',s}(A) - P_p(A) = E_{p',s}\mathbf{1}_A - E_p\mathbf{1}_A \\
=\ & P_{p',s}(\Omega(s)^+)E_{p',s}\left[\mathbf{1}_A|\Omega(s)^+\right] + P_{p',s}(\Omega(s)^-)E_{p',s}\left[\mathbf{1}_A|\Omega(s)^-\right] \\
& -P_p(\Omega(s)^+)E_p\left[\mathbf{1}_A|\Omega(s)^+\right] - P_p(\Omega(s)^-)E_p\left[\mathbf{1}_A|\Omega(s)^-\right] \\
=\ & (p'-p)E_p\left[\mathbf{1}_A - \mathbf{1}_A \circ j_s|\omega \in \Omega(s)^+\right] + E_{p',s}\left[\mathbf{1}_A|\Omega(s)^-\right] - E_p\left[\mathbf{1}_A|\Omega(s)^-\right] \\
=\ & (p'-p)E_p\left[\mathbf{1}_{\text{Piv}(A,s)\cap\Omega(s)^+}|\omega \in \Omega(s)^+\right] \\
=\ & (p'-p)E_p\left[\mathbf{1}_{\text{Piv}(A,s)}\right] = (p'-p)P_p(\text{Piv}(A,s)).
\end{aligned}
$$

That implies finally

$$\frac{\partial P_p(A)}{\partial p(s)} = P_p(\text{Piv}(A,s)).$$

(iv) By $A \in \mathcal{F}_\mathcal{G}$, we have

$$E_p\mathbf{1}_A = E_p\left[\mathbf{1}_A|\mathcal{F}_\mathcal{G}\right] = \sum_{\omega \in \{0,1\}^\mathcal{G}} \Pi_{s\in\mathcal{G}}1_A(\omega)\Pi_{s\in\mathcal{G}}P_p(\omega(s))$$

that means, we may think of the distribution P_p as a distribution depending on *finitely many* real parameters $\{p(s) : s \in \mathcal{G}\}$. That implies together with (iii)

$$\frac{d}{dp}P_p(A) = \sum_{s\in\mathcal{G}}\frac{\partial P_p(A)}{\partial p(s)}\frac{\partial p(s)}{\partial p} = \sum_{s\in\mathcal{G}}P_p(\text{Piv}(A,s)) = \sum_{s\in\mathcal{G}}E_p\mathbf{1}_{\text{Piv}(A,s)} = E_pN_A.$$

∎

7.3. *The proof of (10)*

We follow closely the proof in Grimmett (1999), p. 263 ff. Let $P_p(x,y)$ denote the probability that there is a path connecting the sites x and y and $P_p^{(N)}(x,y)$ the probability that there is a path connecting x and y which lies entirely in $\mathcal{T}^{(N)}$. Now

$$\chi_N(p,y) := \sum_{x\in\mathcal{T}^{(N)}} P_p^{(N)}(x,y)$$

is the expected size of the connected cluster around y in $\mathcal{T}^{(N)}$ and

$$\chi(p, y) := \sum_{x \in \mathcal{T}} P_p(x, y)$$

the expected cluster size in $\mathcal{T}$. Note that $\chi(p) = \chi(p, 0)$. Furthermore, we write

$$\chi_N(p) := \max\{\chi_N(p, y) \, : \, y \in \mathcal{T}^{(N)}\}.$$

(i) First of all,

$$\chi(p) \geq \chi_N(p) \geq \chi_N(p, 0) = \sum_{x \in \mathcal{T}^{(N)}} P_p^{(N)}(x, 0) \to \sum_{x \in \mathcal{T}} P_p(x, y) = \chi(p)$$

implies that we have by bounded convergence

$$\lim_{N \to \infty} \chi_N(p) = \chi(p).$$

(ii) Denote by $A_N(x, y)$ the event that there is a path connecting x and y in $\mathcal{T}^{(N)}$. Then, by Russo's formula,

$$\frac{d}{dp} \chi_N(p, y) = \sum_{x \in \mathcal{T}^{(N)}} \sum_{s \in \mathcal{T}^{(N)}} P_p(\mathrm{Piv}(A_N(x, y), s)).$$

A site $s \in \mathcal{T}^{(N)}$ is now pivotal for $A_N(x, y)$, if and only if

1. s is adjacent to two different and non - adjacent sites x' and y'.
2. There is a path connecting x and x'.
3. There is a *disjoint* path connecting y and y', meaning that no site in this path is adjacent to any site in the path connecting x ans x'.

This means that switching s on or off will switch a connection between x and y on or off (which changes the value of the corresponding indicator function). Having disjoint paths between different pairs of sites is a typical example of a disjointly occuring event and therefore we can write the three conditions above shortly by saying that for all $x, y \neq z \in \mathcal{T}^{(N)}$ and all $x' \neq y'$ adjacent to and different from s, we have

$$A_N(x, x') \circ A_N(y, y') \subset \mathrm{Piv}(A_N(x, y), s)$$

and that on the other hand

$$\mathrm{Piv}(A_N(x, y), s) = \bigcup_{x' \neq y' \text{ adjacent to } s} A_N(x, x') \circ A_N(y, y').$$

That implies by BK inequality

$$P_p(\mathrm{Piv}(A_N(x, y), s)) \leq \sum_{x' \neq y' \text{ adjacent to } s} P_p^{(N)}(x, x') \, P_p^{(N)}(y, y').$$

Finally inserting this into Russo's formula yields

$$
\begin{aligned}
\frac{d}{dp}\chi_N(p,y) &= \sum_{x\in\mathcal{T}^{(N)}}\sum_{s\in\mathcal{T}^{(N)}} P_p(\mathrm{Piv}(A_N(x,y),s)) \\
&\leq \sum_{x\in\mathcal{T}^{(N)}}\sum_{s\in\mathcal{T}^{(N)}}\sum_{x'\neq y'\ \text{adjacent to}\ s} P_p^{(N)}(x,x')\,P_p^{(N)}(y,y') \\
&= \sum_{s\in\mathcal{T}^{(N)}}\sum_{x'\neq y'\ \text{adjacent to}\ s} \chi_N(p,x')\,P_p^{(N)}(y,y') \\
&\leq \chi_N(p)\sum_{s\in\mathcal{T}^{(N)}}\sum_{x'\neq y'\ \text{adjacent to}\ s} P_p^{(N)}(y,y') \\
&= 3\,\chi_N(p)\sum_{s\in\mathcal{T}^{(N)}}\sum_{y'\ \text{adjacent to}\ s} P_p^{(N)}(y,y') \\
&= 3\times 6\,\chi_N(p)\sum_{s\in\mathcal{T}^{(N)}} P_p^{(N)}(y,s) \\
&= 18\,\chi_N(p)\,\chi_N(p,y) \leq 18\,\chi_N(p)^2.
\end{aligned}
$$

(iii) Integrating this differential inequality over the interval $[p,p_c]$ yields

$$
\frac{1}{\chi_N(p)} - \frac{1}{\chi_N(p_c)} \leq 18\,(p-p_c)
$$

(for details, see the above mentioned proof in Grimmett (1999)) and by (i) $\chi_N \to \chi$ and the fact that $\chi(p_c)=\infty$ we finally obtain

$$
\chi(p) \geq \frac{1}{18\,(p-p_c)}.
$$

7.4. Matching graphs and $p_c = 1/2$

In this subsection, we will shortly review the material from Sykes and Essam (1964) about site percolation and matching graphs. We start with a finite graph $\mathcal{G}$ with N sites. The probability that a site is marked *active* (or *black*) is given by p, the probability that it is marked *inactive* (or *white*) is $q = 1 - p$. Denote a connected cluster of black points by C and its boundary by

$$
\partial C := \{s \in \mathcal{G} - C \,:\, s \text{ is adjacent to some site } s' \in C\}.
$$

That means, the expected cluster size is a polynomial in p and q given by

$$
K(p,q,\mathcal{G}) = E\,|C| = \sum_{C\subset\mathcal{G}} |C|\,p^{|C|}q^{|\partial C|}.
$$

By reversing the roles of p and q, we obtain the expected numbers of white clusters. To extend this concept to infinite graphs, we consider the *mean cluster size per site*

$$
k(p,q,\mathcal{G}) = E(|C|/|\mathcal{G}|) = \frac{1}{|\mathcal{G}|}\sum_{C\subset\mathcal{G}} |C|\,p^{|C|}q^{|\partial C|}, \tag{15}
$$

use a proper exhaustion $\mathcal{G}_1 \subset \mathcal{G}_2 \subset ... \subset \mathcal{G}$ of an infinite graph $\mathcal{G}$ and consider the formal power series

$$k(p, q, \mathcal{G}) = \lim_{k \to \infty} \frac{1}{|\mathcal{G}_k|} \sum_{C \subset \mathcal{G}_k} |C| \, p^{|C|} q^{|\partial C|}$$

which shows that we obtain in this case the expected *finite* cluster size per size, taking into account only finite subclusters from $\mathcal{G}$. By

$$k_L(p, \mathcal{G}) = k(p, 1 - p, \mathcal{G}), \quad k_H(q, \mathcal{G}) = k(1 - q, q, \mathcal{G}), \tag{16}$$

we clearly have $k_L(p, \mathcal{G}) = k_H(1 - p, \mathcal{G})$ and $k_H(q, \mathcal{G}) = k_L(1 - q, \mathcal{G})$.

Definition 6. *We call a (possibly infinite) graph $\mathcal{G}$ self - matching, if there is a polynomial $\varphi_{\mathcal{G}}(p)$ such that*

$$k_L(p, \mathcal{G}) = \varphi_{\mathcal{G}}(p) + k_H(p, \mathcal{G}). \tag{17}$$

$\varphi_{\mathcal{G}}$ is called the matching polynomial.

Theorem 5. *The triangular lattice is self-matching with*

$$\varphi_{\mathcal{T}}(p) = p - 3p^2 + 2p^3.$$

Proof: See Sykes and Essam (1964). ∎

When we assume that $k_L(p, \mathcal{T})$ has precisely one pole at the *critical percolation probability p_c* for the triangular lattice (see for instance Kesten (1982)), we can actually use the preceding theorem to determine p_c. Here, the special form of the matching polynomial does not play any role, only the fact that it is a polynomial and hence bounded on $p \in [0, 1]$ is important. Therefore $k_L(p_c, \mathcal{T}) = \infty$ implies $k_H(p_c, \mathcal{T}) = k_L(1 - p_c, \mathcal{T}) = \infty$. The assumption that there is only one pole immediately implies $p_c = 1 - p_c$ and thus $p_c = 1/2$.

Remark. If the graph $\mathcal{G}$ is not *self - matching*, we can construct a so called *matching graph $\mathcal{G}^*$* (for the construction, see again Sykes and Essam (1964), or Kesten (1982)) with the same number of vertices such that instead of (17), we have

$$k_L(p, \mathcal{G}) = \varphi(p) + k_H(p, \mathcal{G}^*), \quad k_L(p, \mathcal{G}^*) = \varphi^*(p) + k_H(p, \mathcal{G}),$$

together with some relations between φ and φ^* and these equations can be used in a similar way as above to obtain some information about the critical probability. $(\mathcal{G}, \mathcal{G}^*)$ is called a *matching pair*. For self - matching graphs, we have $\mathcal{G}^* = \mathcal{G}$.

8. Bounded detector devices

In the discussion of realistic signals, we introduced the notion of a *bounded detector device*. In statistical terminology, those devices from an instance of the method of truncation. A *bounded detector device of range $r > 0$* is only able to display signal strengths $Y(s)$ with intensities between $-r$ and r. Thus, the

effect of the detector device on a signal Y is that instead of the full information about $Y(s)$, $s \in S$, only the information contained in the cutoff signal

$$D(Y) = \max\{\min\{Y, r\}, -r\} \tag{18}$$

is used for further analysis. Intensities of absolute value larger than r can simply not be registered and all information about the behavior of the signal above and below the cutoff is lost before the signal processing even starts.

The detection results in the present paper were proved for bounded signals. What happens if this assumptions doesn't hold? First of all, from a purely mathematical point of view, the notion of bounded detector devices can be equivalently reformulated by saying that all considerations are only valid as statements that are obtained while *conditioning* on the event $\{|Y| < r\}$. In other words, all results are still valid without any change if we understand them as being obtained by *conditioning on the event*

$$D_0 := \{D(Y) = Y\}. \tag{19}$$

Of course, the probability $\pi_D := P(D_0)$ then yields an important characteristic of the detector device, and it could be often desirable to have π_D close to one. However, a deeper analysis of biological, engineering and cybernetical aspects of the problem leads us to the following extremely useful observation.

We think of signal processing as consisting of at least three different parts,

1. a *filter* which has the purpose to transform the incoming signal to fit in an optimal way into the bounds of the detector device,
2. the *bounded detector device* as described above, and
3. the *processor*, which analyses the detected signal $D(Y)$ and determines what is finally perceived.

We thus arrive at the following scheme

$$\text{Signal} \rightarrow \boxed{\text{Filter}} \rightarrow \boxed{\text{Detector}} \rightarrow \boxed{\text{Processor}} \rightarrow \text{Perception}$$

where the detector is the fixed component, the filter is chosen on the basis of the incoming signal and the bounds of the detector and the processor algorithm is chosen on the basis of knowledge about the detector and the chosen filter. Choosing an appropriate filter for a given environment is thus another problem of perception, a problem that we will not address in these notes.

Example. As an example, as the *detector device* of the human eye, we only consider the photo receptors situated at the retina, the processor obviously is the brain, and the filter is given by lens and iris which adapt to different light intensities for instance in night vision, but can also be those parts together with another device like, for instance, sun glasses.

For a visual perception of any system in biology or cybernetics, the meaning of a *good* Filter is exactly to filter out (or to transform) the incoming information in such a way that the Detector might still perceive a reasonable part of reality, despite the fact that the Detector works with signals in the diapason $[-r, r]$ only. Say, in the above Example, a human eye doesn't have to properly perceive

ultraviolet and infrared light frequencies in order to be able to see trees. A human brain doesn't need to process any information that could come with ultraviolet and infrared lights either.

This implies that our consideration of bounded detector devices fits many important biological situations. Moreover, working with bounded detector devices can be profitable for construction of artificial vision systems in robotics. A robot needs to perceive and to process only signals and information within the diapason that fits his tasks.