Leipzig University

Faculty of Physics and Earth Sciences
Institute of Theoretical Physics
Computational Quantum Field Theory

Bachelor Thesis

Generating Long-range Power-law Correlated Disorder

submitted by
Franz Paul Spitzner

in partial fulfillment of the requirements for the degree of
Bachelor of Science (B. Sc.)
in Physics (IPSP)

Supervisors:
Prof. Dr. Wolfhard Janke
M. Sc. Johannes Zierenberg

Referees:
Prof. Dr. Wolfhard Janke
Dr. Stefan Schnabel

Leipzig, October 7, 2014

Contents

1 Introduction

In times of increasingly complex social networks, environmental problems and when terms like globalisation or gentrification are explained in school, it is not surprising that mankind is more than ever fascinated by causality and correlation. For many people those two expressions mean pretty much the same, nonetheless there is remarkable difference between the two. While causality describes a reasoned order or relationship between two phenomena, which often cannot be reduced to a few words or a formula, correlation on the other hand is in fact measurable and mathematically defined. In a way it is an expression for the similarity of two observables that depend on some common parameters. The intent of the following pages is to create and discuss a correlation that shows a power-law behaviour, to identify difficulties and suggest possible solutions how to overcome them. Power-law exhibiting phenomena are common in nature and society, e.g. earthquakes, citygrowth or infrastructure. Hence we find a long-lasting interest in versatile scientific areas such as phase transitions [1], percolation problems [2, 3], communication activity [4] or surface studies [5], that fit well in this context. The major effort will go into the reproduction and understanding of the paper by Makse et al. from 1996 [6], as it covers the general method of creating correlated disorder in great detail. Once we have successfully created correlated sequences, which show the proposed behaviour, the process will be generalized from one to two dimensions, so that a mapping onto a square lattice can be done. During the discussion we are going to work with systems of a maximum size of $N = 2^{21}$ for one dimension and a maximum size per dimension of $N = 2^{11}$ for two dimensions. Since this will lead to numerical values of the order 10^6 and larger during some of the calculations we would need to take care of machine dependant limits, especially for random number generators as has been emphasized in [7]. Unfortunately deep inquiries on this subject also require a fair amount of background, testing and care, which is why further consideration of this matter will not be included. For the generation of random numbers we use the Mersenne Twister 19937 generator (mt19937) which seems to offer decent results, as shown in the aforementioned publication. A follow-up aspiration would be to exactly enumerate a random walk in the created disorder-confinement, which is of particular interest as it offers a similar approach to the problem treated in [8].

2 Fourier Transform

Simply put, the Fourier transfom is a mathematical tool which transforms data from its time domain into the frequency domain, or in other words from real space (RS) to Fourier space (FS). The data itself does not change, rather one can think of two different ways to represent it. In general, the time domain is the common representation that is used in everyday life. Nevertheless there is a useful property to the frequency representation: Every single discrete value in FS is equivalent to a distribution over the whole RS - and vice versa. Hence any change or operation that is performed on a single point in FS has an effect upon every point in RS, which is what we are going to exploit to create the desired power-law correlation.

The Fourier transform for continuous functions is given by

$$H(f) = \int_{-\infty}^{\infty} h(t)\, e^{2\pi i f t} dt, \tag{2.1}$$

where t is the time, f is the frequency and i the imaginary number, $i = \sqrt{-1}$. The inverse transform to get back into RS is:

$$h(t) = \int_{-\infty}^{\infty} H(f)\, e^{-2\pi i f t} df \tag{2.2}$$

In most cases, especially for experimental purposes, it is required to work with discrete data rather than with continuous functions, which raises interest in the discrete Fourier transform. The aim is to construct the Fourier transform of the function $h(t)$ from a given number N of sample points, where a constant sampling interval Δ is assumed.

$$h_k = h(t_k), \qquad t_k = k\Delta, \qquad k = 0, 1, 2, ..., N-1 \tag{2.3}$$

From these N points of input data it is intuitively clear to search for no more than N points of output. There is another fact which needs to be considered, that is the so called sampling theorem. It gives birth to terminology such as aliasing, bandwith limitation and the Nyquist critical frequency. The highest frequency possible to sample with these N points is $f_c = \frac{1}{2\Delta}$, which is plausible if one thinks of a sine: it cannot be reduced to less than two discrete points per period without giving up information. This puts certain limitations on the Fourier analysis, for more information see also Ref. [9].

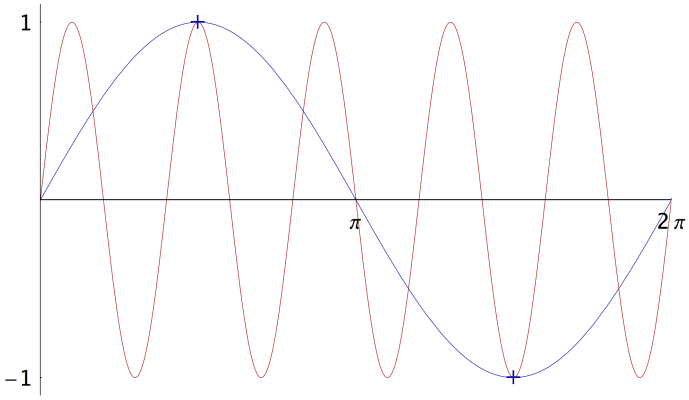


Figure 2.1 Both functions $\sin(x)$ and $\sin(5x)$ have the same functional value of ± 1 at $x = \frac{\pi}{2}$ and $x = \frac{3\pi}{2}$, but only one period of $\sin(x)$ can be sampled uniquely with the two points.

Taking the beforehand mentioned obstructions into account, one seeks

$$f_n = \frac{n}{N\Delta}, \qquad n = -\frac{N}{2}, ..., \frac{N}{2} \tag{2.4}$$

and defines

$$H_n = \sum_{k=0}^{N-1} h_k \, e^{2\pi i k \frac{n}{N}} \tag{2.5}$$

where the relation between the discrete and the continuous Fourier transform can be written as:

$$H(f_n) \approx \Delta H_n$$

We see that Δ can be treated like a unit in the considered domain. Usually one sets $\Delta = 1s$, which gives seconds as the unit in RS and hertz in FS. For the discrete inverse transform a normalization factor appears, which follows from Parseval's theorem. Due to the periodic behaviour of H_n in n, the sum can be shifted so that both indices, n and k vary over the same range.

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n \, e^{-2\pi i k \frac{n}{N}} \tag{2.6}$$

2.1 Fast Fourier Transform

Having looked at the discrete Fourier transfrom briefly, it becomes apparent that N multiplications are required to obtain every H_n, which needs to be done exactly N times, leading to a total complexity of the order N^2. The goal of the fast Fourier transform (FFT) is to deliver the same output - but faster. In fact FFT algorithms operate as efficient as $O\big(N \log(N)\big)$. This noteworthy performance gain can be achieved by applying a lemma by Danielson and Lanczos. The ingenious trick is to split the discrete sum of length N into two sums of length $N/2$, where the first one contains all coefficients with even indices and the second sum only contains the coefficients with odd indices.

$$H_n = \sum_{k=0}^{N-1} h_k \, e^{2\pi i k \frac{n}{N}} \tag{2.7}$$

$$= \sum_{k=0}^{N/2-1} h_{2k} \, e^{2\pi i (2k) \frac{n}{N}} + \sum_{k=0}^{N/2-1} h_{2k+1} \, e^{2\pi i (2k+1) \frac{n}{N}} \tag{2.8}$$

$$= \sum_{k=0}^{N/2-1} h_{2k} \, e^{2\pi i k \frac{n}{N/2}} + e^{2\pi i \frac{n}{N}} \sum_{k=0}^{N/2-1} h_{2k+1} \, e^{2\pi i k \frac{n}{N/2}} \tag{2.9}$$

Considering $W = e^{2\pi i/N} = const$

$$H_n = \sum_{k=0}^{N/2-1} h_{2k} \, e^{2\pi i k \frac{n}{N/2}} + W^n \sum_{k=0}^{N/2-1} h_{2k+1} \, e^{2\pi i k \frac{n}{N/2}} \tag{2.10}$$

$$= H_n^e + W^n \, H_n^o \tag{2.11}$$

Where H^e denotes the sum of all even elements and H^o the odd ones. What is left to do is a multiplication with a constant factor: some power of W^n. The beauty of this method is, it can be done recursively until the sums only have length one, evaluating merely

$$H_n^{eeo\ldots eoe} = h_k \tag{2.12}$$

for each k. The challenge for a working algorithm is to get the book keeping right, on which index to apply the correct factor. Furthermore, this idea of repeatedly splitting sums into two requires the input data to have length 2^n, $n \in \mathbb{N}$. Of course there are ways to work around this limitation, such as padding with zeros, but since the input data in this particular case is created at random, it might as well just satisfy the condition. The actual aftermath of the precedent contemplation of the FFT is the

way data is stored in FS. Since this is where we are going to manipulate the data, an exact understanding of the arrangement of frequencies is required. Because different notations are common, depending on the reference, it becomes very hard to be self consistent, yet congruent with those references. One important convention about the storage layout is explained in very much detail in Numerical Recipes [9, p. 612]. We will mostly stick to it with subtle modifications. It turned out to be useful to work with complex numbers over treating real and imaginary parts as separate elements in the storage array. In compliance with Makse et al. [6] the angular frequency q will be used as the argument of the soon to be introduced spectral density $S(q)$.

$$q = 2\pi f, \qquad -\frac{1}{2} \leq f \leq +\frac{1}{2}$$

Where f varies with increments of $\frac{1}{N}$, N being the system size in one dimension. Another notation is commonly used, that offers an increment size normalized to 1, which has evident advantages during the implementation.

$$q = \frac{2\pi}{N}\, m, \qquad -\frac{N}{2} \leq m \leq +\frac{N}{2} \tag{2.13}$$

Index	RS	FS	
0	$t = 0$	$f = 0$	$m = 0$
1	$t = 1$	$f = +\frac{1}{N}$	$m = 1$
2	$t = 2$	$f = +\frac{2}{N}$	$m = 2$
...			
$\frac{N}{2} - 1$	$t = \frac{N}{2} - 1$	$f = +\frac{N/2-1}{N}$	$m = \frac{N}{2} - 1$
$\frac{N}{2}$	$t = \frac{N}{2}$	$f = \pm\frac{1}{2}$	$m = \pm\frac{N}{2}$
$\frac{N}{2} + 1$	$t = \frac{N}{2} + 1$	$f = -\frac{N/2-1}{N}$	$m = -\left(\frac{N}{2} - 1\right)$
...			
$N - 2$	$t = N - 2$	$f = -\frac{2}{N}$	$m = -2$
$N - 1$	$t = N - 1$	$f = -\frac{1}{N}$	$m = -1$

Table 2.1 Storing complex numbers at each index leads to arrays of size N instead of $2N$. The convention in the very right column where m only takes integer values will be used in the implementation. Values in RS are in units of $\Delta = 1s$ and in FS $\frac{1}{\Delta} = \frac{1}{s}$.

3 Long-range Power-law Correlated Disorder

3.1 Correlation in One Dimension

"The possibility that the elements in a percolation problem experience a long-range spatial correlation has been of long-standing interest." [2] Promising algorithms have been developed to create such correlations. The usual purpose is to end up with a correlated sequence of the form below, where γ is the correlation exponent.

$$\langle \eta_i \, \eta_{i+l} \rangle \propto l^{-\gamma} \tag{3.1}$$

The general line of action is the following:

1. Generate a random, uncorrelated spatial normal distribution - $\{u_i\}$.
2. Bring it into the frequency domain using FFT - $\{u_q\}$.
3. Multiply every element of data in FS with a factor that depends on the position within FS - $\{\eta_q\}$.
4. Transform back into real space - $\{\eta_i\}$.

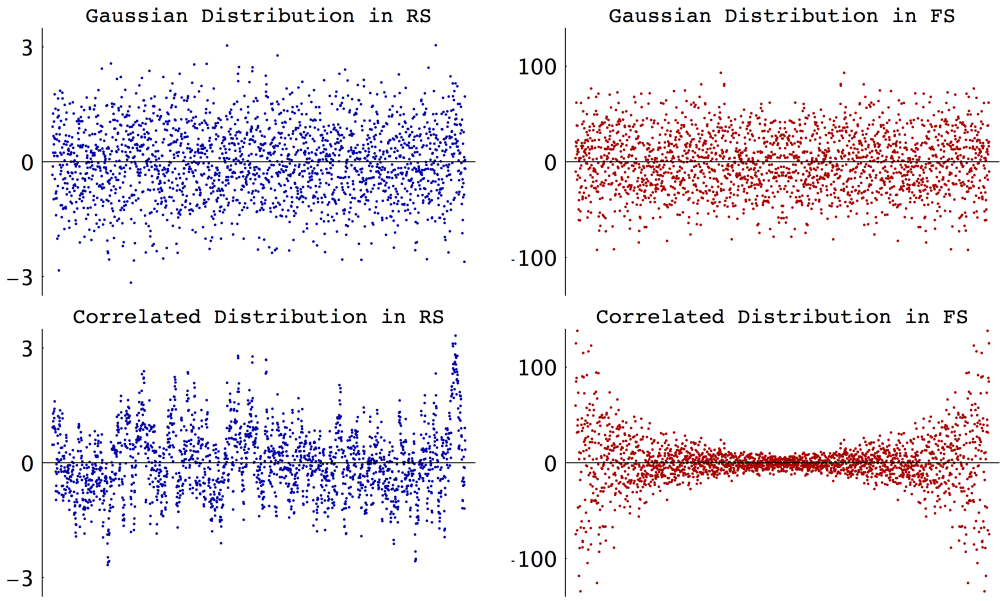


Figure 3.1 A sample of random data of size 2^{11} during the four steps

In our case, the factor is the square root of the spectral density $S(q)$, which is nothing else than the correlation function in Fourier space. It can be obtained by doing the apparent: applying a Fourier transfrom to the correlation function $C(l)$ in real space. Starting with an intuitive approach for the correlation function in the form of Eq. 3.1 - as has been done in [2], the so called Fourier filtering method - it becomes evident, that the integration from $-\infty$ to $+\infty$ during the Fourier transform is not well defined due to the singularity at $l = 0$. This causes the asymptotic behaviour to also occur in FS, leading to a correlation that does not cover the whole system [6]. This can be circumvented by setting

$$C(l) = \left(1 + l^2\right)^{-\gamma/2} \tag{3.2}$$

which exhibits the same behaviour for large l but avoids the singularity.

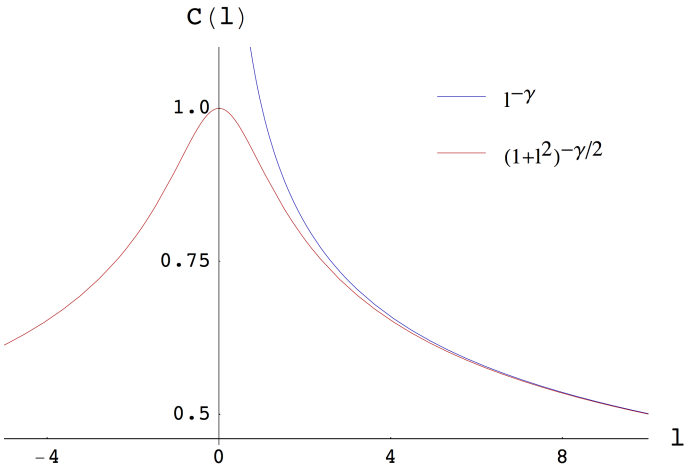


Figure 3.2 The two different approaches for the correlation function in RS, $\gamma = 0.3$

Now, what does the spectral density in Fourier space $S(q)$ look like? By using Mathematica to perform an analytical Fourier transform of $C(l)$ one obtains

$$S(q) = \sqrt{\pi}\, 2^{\frac{3}{2}-\frac{\gamma}{2}}\, |q|^{\frac{\gamma-1}{2}}\, \frac{K_{\frac{\gamma-1}{2}}(|q|)}{\Gamma\left(\frac{\gamma}{2}\right)} \tag{3.3}$$

where K is the modified Bessel function of second kind and Γ denotes the gamma function. Using $\beta = \frac{\gamma-1}{2}$ we can simplify to

$$S(q) = \frac{2\sqrt{\pi}}{\Gamma\left(\beta + 0.5\right)} \left(\frac{|q|}{2}\right)^{\beta} K_{\beta}(|q|) \tag{3.4}$$

and notice that $S(q) = S(-q)$. This is a quite beneficial property, since the data on which $S(q)$ will be applied in FS shows a similar symmetry, see Table 2.1. By definition

we also know $K_\beta = K_{-\beta}$, which determines the domain in which it is appropriate to define γ, that is $0 < \gamma < d$ with $d = 1$ for one dimension. Furthermore, this equation is similar to the one by Makse et al. [6]:

$$S(q) = \frac{2\sqrt{\pi}}{\Gamma(\beta + 1)} \left(\frac{q}{2}\right)^{\beta} K_\beta(q) \tag{3.5}$$

It is worth mentioning, that we concluded the argument of the gamma function in Eq. 3.5 is $\beta + 1$ only due to typing error, which might have been caused by the γ to β conversion in higher dimensions. As a matter of fact, rewriting Eq. 3.4 in terms of β does not serve another reason than to give an equation that looks a bit clearer. After correlating the sequence, we need a measure to ensure the sequence does in fact show the aspired behaviour. To this extend we calculate $C'(l)$ for appropriate values of l

$$C'(l) = \frac{1}{N} \frac{\sum_{i=0}^{N} (x_i - \mu)(x_{i+l} - \mu)}{\sigma^2} \tag{3.6}$$

with expectation value μ and variance σ^2.

$$\mu = \frac{1}{N} \sum_{i=0}^{N} x_i, \qquad \sigma^2 = \frac{1}{N} \sum_{i=0}^{N} (x_i - \mu)^2$$

The actual measurement of $C'(l)$ during the analysis has a huge impact on the results. Though one would assume to achieve a better distribution for exactly evaluted expectation values μ, the contrary turns out to be true. Only setting μ to zero in Eq. 3.6 altogether - consistent with the intitially created Gaussian - allowed for the intended outcome. This is non-trivial, since applying the correlation to the initial input data is quite a strong modification and it is not directly deducible that the expectation value would remain unchanged. An empirical argument is shown in Fig. 3.3, where we see indeed that the correlated data is on average normally distributed around zero, although individual realizations for different seeds are not. Increased detail in this matter will be given for the two dimensional case.

Once measured, the correlation values are plotted together with the reference $C(l)$ in the same graph. When observing Fig. 3.4, a discrepancy between measured values and the $C(l)$ is notable for small γ. We examined that strong correlations with small exponents γ react delicately to the way the singularity $S(0)$ is treated, which originates in the Bessel function. While the fluctuation for large γ is coherent with [6, 10], the general trend that those sequences tend to be overcorrelated beyond a certain l is not. In one dimension we will comply with Makse et al. [6] by "assigning a suitable

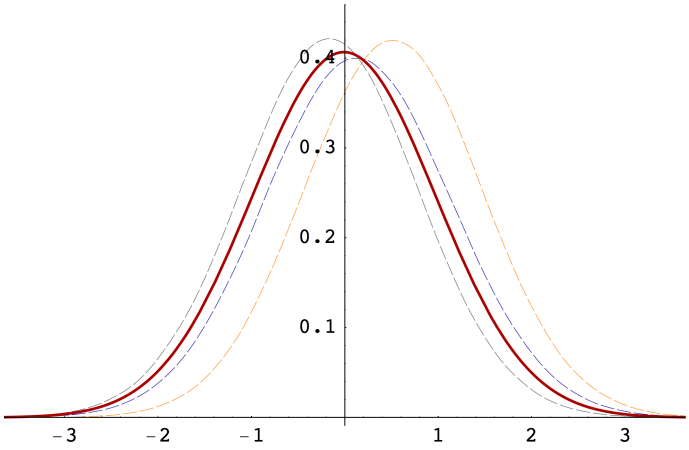


Figure 3.3 The distribution of 50 realizations of correlated sequences in 1D for $N = 2^{21}$: the average (red) and individual seeds (dashed).

numerical value $0 < m_0 < 1$".

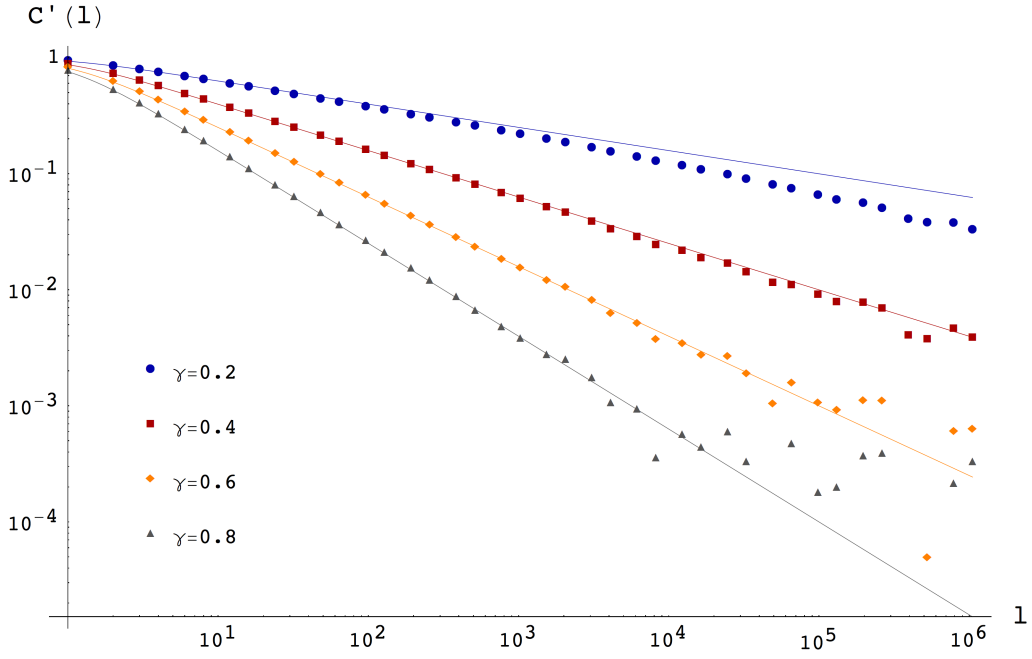


Figure 3.4 Measured correlations and according references $C(l)$ for different γ and 1D systems of size $N = 2^{21}$. The singularity $S(0)$ was treated with $m_0 = 0.1$. Averages were taken over 50 samples.

3.2 Correlation in Two Dimensions

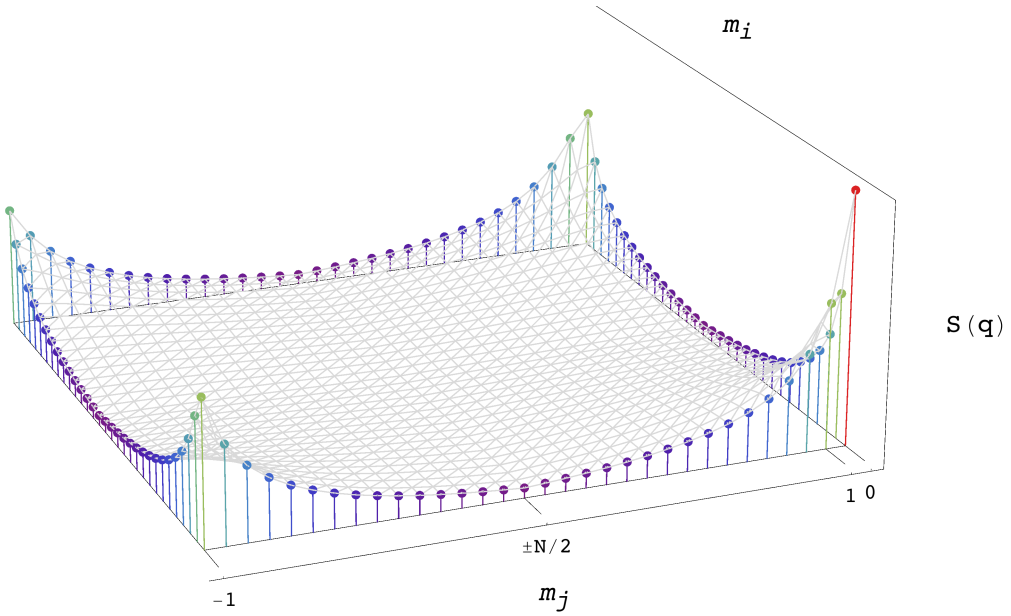


Figure 3.5 3D plot of the spectral density for a system size of $2^5 \times 2^5$ and $\gamma = 1.2$. Note the symmetry for $S(q)$ that arises from the FFT. $S(0)$ has great impact on the correlation in RS, although there exist no analytical solution.

The general procedure for higher dimensions builds upon the one dimensional case, even though one needs to be aware of some additional circumstances. For one, the FFT in higher dimensions is computationally more intensive. This causes us to decrease the system size to $N = 2^{11}$ in each of the two dimensions. In terms of storage we will stick to the convention from Numerical Recipies [9, p. 629] "store by rows", with complex numbers. Moreover, $C(l)$ and $S(q)$ defined in Eq. 3.2 and Eq. 3.4 need to be generalized to higher dimensions. Following Makse et al. [6] and considering periodic boundary conditions, the correlation function is given by

$$C\left(\vec{l}\right) = \left(1 + \sum_{i=1}^{d} l_i^2\right)^{-\gamma/2} \tag{3.7}$$

which leads to

$$S(\vec{q}) = \frac{2\,\pi^{d/2}}{\Gamma\left(\frac{\gamma}{2}\right)} \left(\frac{q}{2}\right)^{\frac{\gamma-d}{2}} K_{\frac{\gamma-d}{2}}(q) \tag{3.8}$$

or

$$S(\vec{q}) = \frac{2\,\pi^{d/2}}{\Gamma\left(\beta_d + \frac{d}{2}\right)} \left(\frac{q}{2}\right)^{\beta_d} K_{\beta_d}(q) \tag{3.9}$$

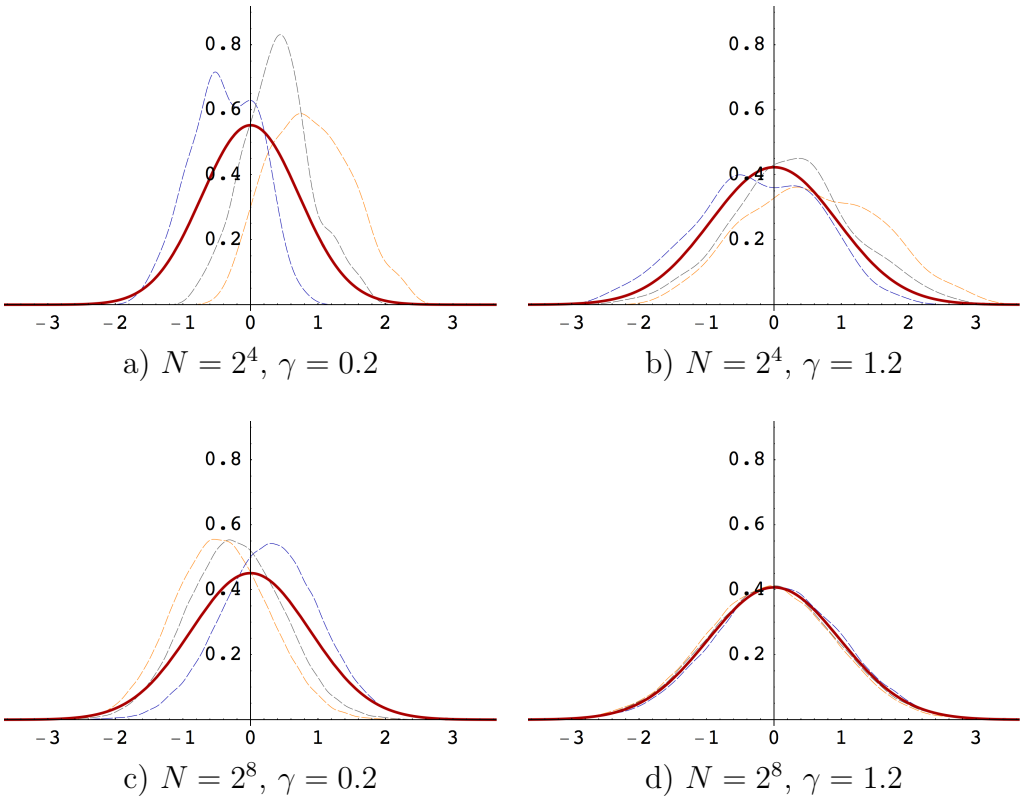
for the spectral density. Again, the argument of the gamma function appears to be

inconsistent in the paper but seems correct for $d = 2$.

$$\beta_d = \frac{\gamma - d}{2}, \qquad q = |\,\vec{q}\,|, \qquad \vec{q} = \frac{2\pi}{N}\,\vec{m}, \qquad -\frac{N}{2} \leq m_i \leq +\frac{N}{2}$$

For now we will use the Euclidean norm for $|\,\vec{q}\,|$, but some future inquiries could go into using a different definiton e.g. the Manhattan norm. It follows for two dimensions:

$$q = \frac{2\pi}{N}\,\sqrt{m_i^2 + m_j^2}$$



a) $N = 2^4$, $\gamma = 0.2$ b) $N = 2^4$, $\gamma = 1.2$

c) $N = 2^8$, $\gamma = 0.2$ d) $N = 2^8$, $\gamma = 1.2$

Figure 3.6 Shown are distributions of correlated 2D sequences for two different system sizes: an average of 50 seeds (red) and example seeds (dashed). Note that individual realizations are closer to the average for larger systems.

Before we discuss the measurement of $C'(l)$ in 2D, it seems sensible to back up the claim, that after correlating a Gaussian, the given set is still normally distributed, since this assumption will be used not only while measuring the correlation itself, but also when deriving the mapping algorithm. On that account 50 realizations of correlated sequences in 2D were created for different system sizes and different correlation exponents γ. The system size was varied between 2^4 and 2^8 per dimension and γ was in the range from 0.2 (strong correlation) to 1.8 (weak correlation). Fig. 3.6 shows some examples. In each case, the average distribution is centered around $\mu = 0$. Further-

more we see that the fluctuation of individual seeds around the average is dependant on both, γ and the system size.

Next, the way the created correlation is measured has to be adopted to 2D. To this end there are two possibilities, either one can simply measure $C'_{\parallel}(l)$ in one component - keeping the other one constant, which should reproduce the results from 1D - or one determines a radial correlation. Admittedly we were not able to come up with a satisfactory algorithm to measure the correlation for a given length in an arbitrary direction. The problem lies within the discrete lattice, one would have to check which lattice points are connectable with that given length at some angle. Thus we settled for a more pragmatic solution, in which we investigate only in diagonal directions and call this $C'_{\perp}(l)$. The clue is not trying to achieve an exact distance from x^i_j to $x^{i'}_{j'}$ but to adjust the argument l afterwards. We will start with the former case, in this notation i and j are the indicies of the matrix element, equivalent to the positions on a square lattice. x^i_j is the corresponding value, analogeous to 1D.

$$C'_{\parallel}(l) = \frac{1}{N^2} \frac{\sum_{i,j=0}^{N} \left(x^i_j - \mu\right) \left(x^{i+l}_j - \mu\right)}{\sigma^2} \tag{3.10}$$

Notice $j' = j$ and $i' = i + l$, hence the correlation is only measured in the dimension corresponding to i. For the diagonal correlation we use:

$$C'_{\perp}\left(\sqrt{2}\, l\right) = \frac{1}{2N^2} \frac{\sum_{i,j=0}^{N} \left\{\left(x^i_j - \mu\right) \left(x^{i+l}_{j+l} - \mu\right) + \left(x^i_j - \mu\right) \left(x^{i+l}_{j-l} - \mu\right)\right\}}{\sigma^2} \tag{3.11}$$

On account of shifting both i and j by an amount l in the latter equation, the correlation that is evaluated is not $C'_{\perp}(l)$ but $C'_{\perp}(\sqrt{2}\, l)$. Once again periodic boundary conditions were implemented in the algorithm. It turned out, that both $C'_{\parallel}(l)$ and $C'_{\perp}(\sqrt{2}\, l)$ yield comparably good results, which made us use the diagonal, more sophisticated version.

When looking at Fig. 3.8, we can identify the same issues as in the 1D case, remember that γ is varied in the range $0 < \gamma < d$, in d dimensions. For one, sequences with strong correlations $0 < \gamma < 0.5$ tend to be undercorrelated. This is adressed in the section “Difficulties” and arises from the singularity. Secondly, sequences with weak correlations $1.5 < \gamma < 2$ start to spread heavily for large l, occasionally $C'(l)$ even attains negative values, hence is not shown in the logarithmic plot.

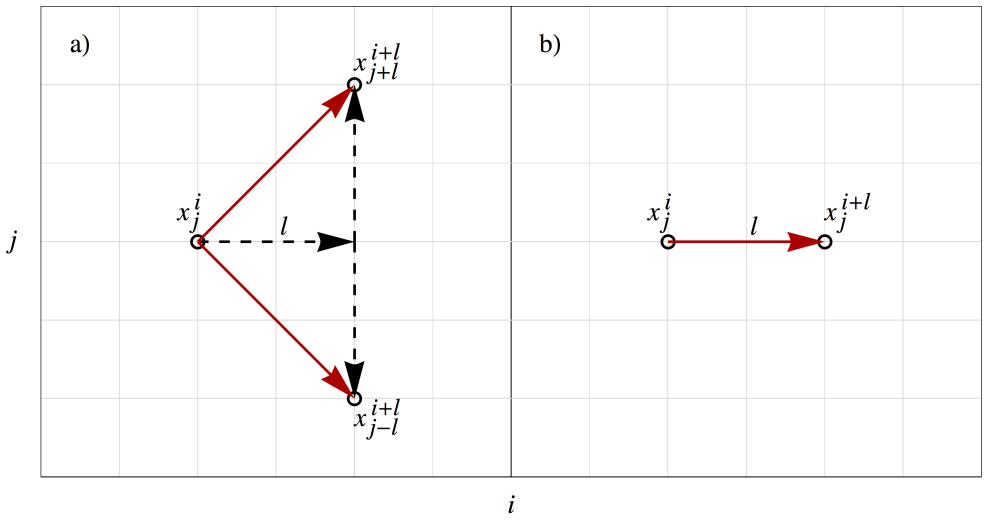


Figure 3.7 Simplified illustration of the measurement of a) $C'_{\perp}(\sqrt{2}\,l)$ and b) $C'_{\parallel}(l)$.

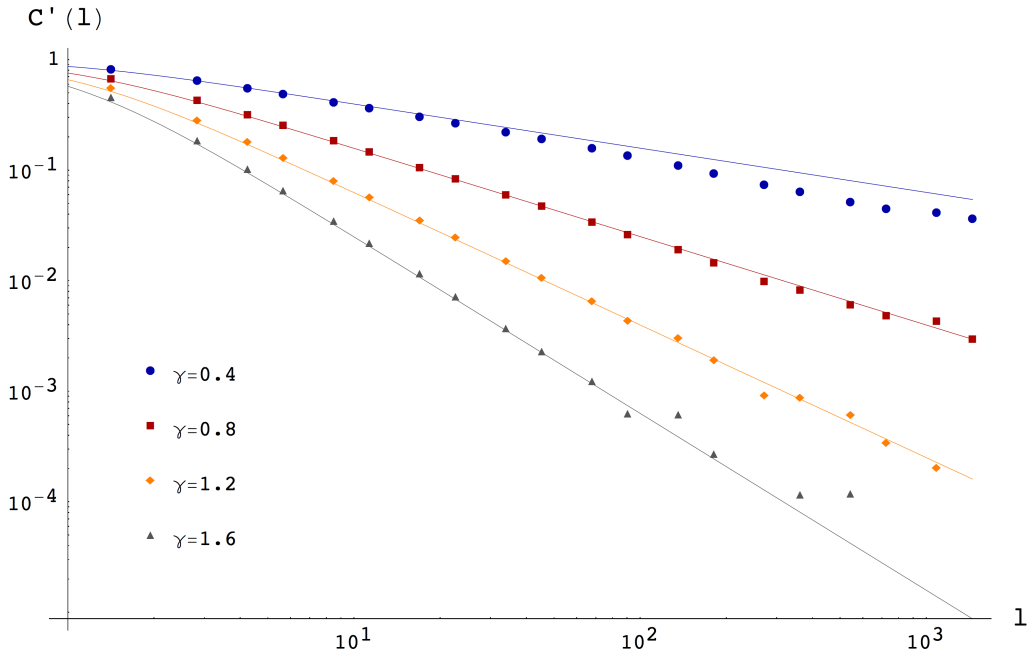


Figure 3.8 Measured correlations for 2D systems of size $2^{11} \times 2^{11}$ and different γ. To avoid the singularity, the approach that will be discussed in the section "Difficulties" was used. See Eq. 3.14.

3.3 Mapping Disorder

In preparation for a potential enumeration of a 2D random walk (RW) in the correlated disorder, it is necessary to translate the disorder into a confinement on a square lattice. There are only two possibilities on the lattice, either the RW is allowed to access a particular position or not. Since we have a real value for each lattice point in the correlated sequence, that could hypothetically be anything between $-\infty$ and ∞ the straightforward approach is to introduce a threshold θ and compare it to the value of each point. Anything below θ is mapped to zero, everything else to one.

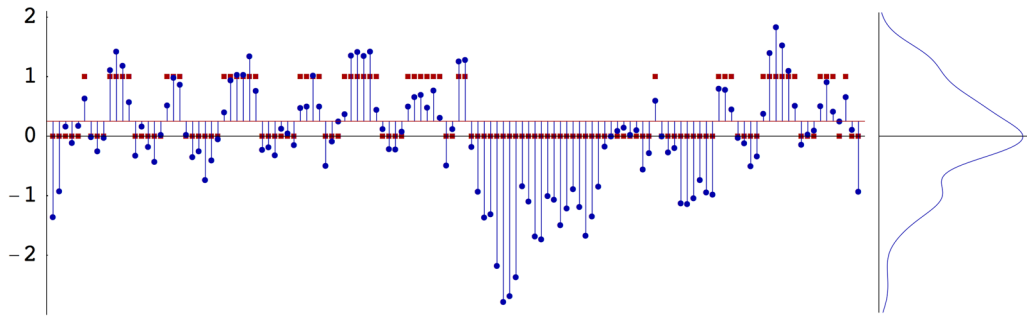


Figure 3.9 The mapping of a 1D correlated sequence (blue) is shown. Every point is mapped to either zero or one (red squares) depending on the threshold θ (red line). On the right, the unmapped sequence is given as a distribution.

It would be desirable to set the amount of accessible and non-accessible lattice sites via a probabilty to find an occupied site

$$p = \frac{n_1}{N^2} \tag{3.12}$$

where n_1 is the number of sites that are not allowed and N^2 is the size of the 2D grid. Fortunately we have already seen that the correlated sequence has a normal distribution.

$$P(x_i) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

For the mapping it is not required to differentiate between the two dimensions, hence we can omit the second index j and only work with x_i. Further we assume $\mu = 0$ and $\sigma = 1$ for $N \to \infty$. We need to be aware that this cannot be assumed for small systems, in which case σ has to be adjusted.

$$p = \int_{-\infty}^{\theta} P(x_i)\, dx_i = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\mu-\theta}{\sigma\sqrt{2}}\right) \tag{3.13}$$

As p is an input parameter and the complementary error function erfc is a well-defined function $f(\theta)$, it is possible to determine θ for any p with $0 < p < 1$ using the bisection method.

One interesting topic to discuss is the combination of mapping and correlation. To this end we can vary the two parameters, the correlation γ and the probability to occupy a lattice point p, hence create some sort of transition table. See Fig. 3.13. What we see is, that very densely occupied areas can be created by either increasing the correlation or the occupation probabibility. Likewise, many unpopulated areas can be achieved using small p, but this does not work analogeously for weakening the correlation, simply because we start with a random Gaussian noise for the sequence. The mapping is done afterwards. Therefore a very even distribution is genereated as in the left column of the figure and the only way to get very light areas, even for weak correlations, is to use a small p, too. This consideration gives reason to think about how the probabibility influences the correlation of points of the mapped lattice. Clearly the same correlation measurements for $C'_\perp(\sqrt{2}\, l)$ can be done for the discrete lattice. The results are shown in Fig. 3.10. What we see in the figure is that the mapping decreases the correlation. Further, $p = 0.5$ gives the strongest mapped correlation, while any deviation in p from one half leads to weaker correlations.

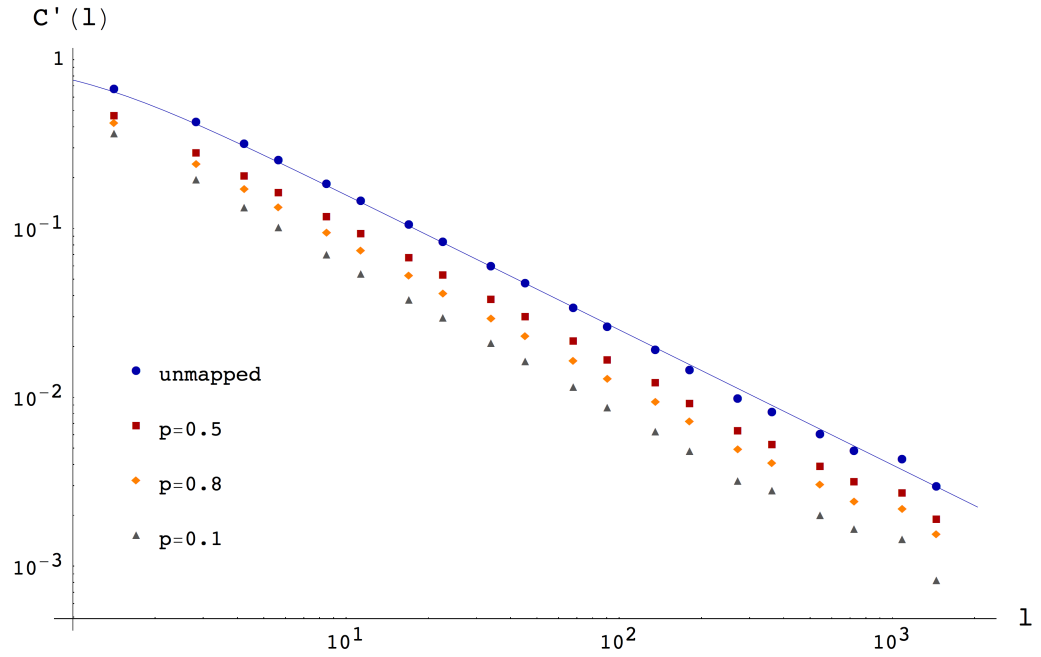
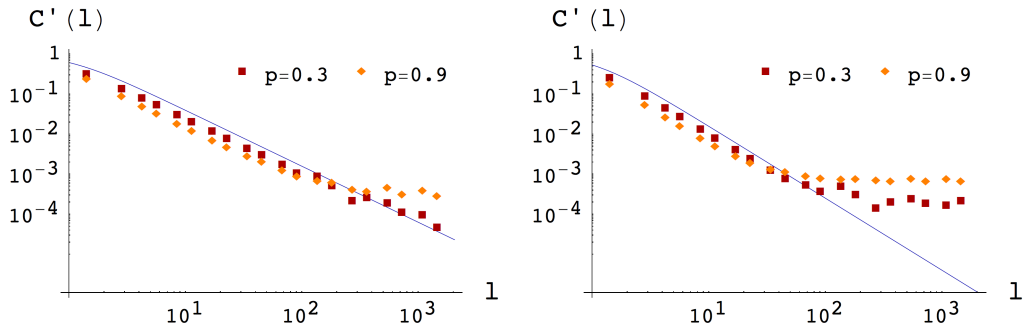


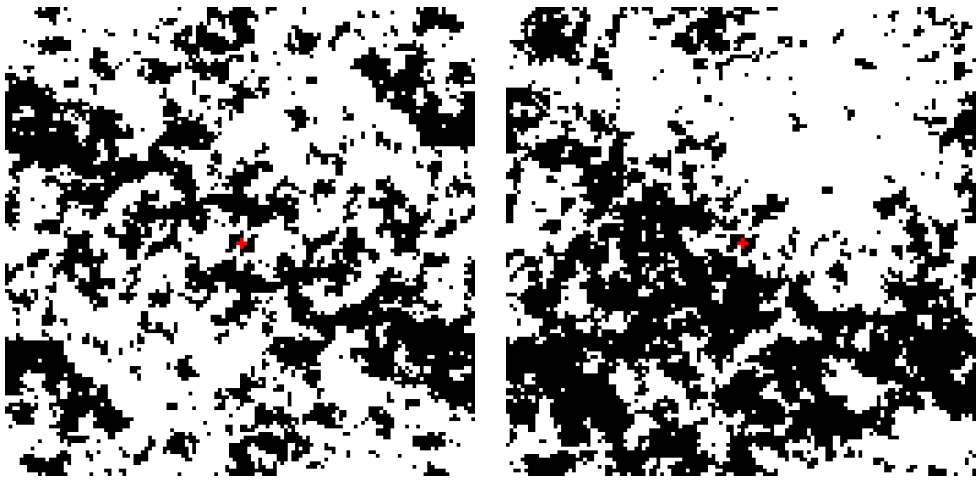
Figure 3.10 Measured correlations of mapped 2D systems for different p and $\gamma = 0.8$, as a reference the unmapped value is given. The mapping with $p = 0.5$ is closest to the unmapped case, the further the probabibility is changed away from 0.5 in either direction, the weaker the correlation.



a) $\gamma = 1.4$ b) $\gamma = 1.8$

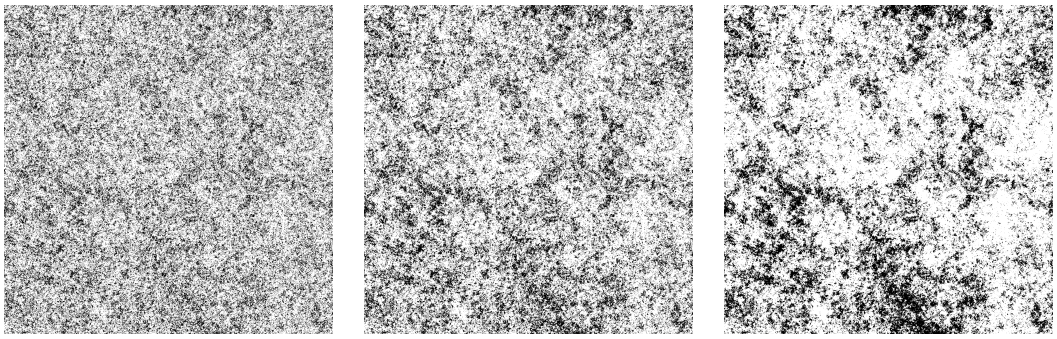
Figure 3.11 For weak correlations where γ gets closer to $d = 2$, the mapped values start to be overcorrelated for certain length scales.

Another effect is shown in Fig. 3.11. This was only observable for weaker correlations, i.e. $d/2 < \gamma < d$: For length scales beyond some l, the correlation was almost constant in l, but this regime was influenced by γ and p. Possibly the mapping smoothes out the fluctuations in $C'(l)$, that were already observable in the unmapped data for according γ. See again Fig. 3.4 for 1D and Fig. 3.8 for 2D. During the process of eliminating causes for possibly wrong correlations, we tried setting the imaginary part of the sequence in FS to zero, while applying $S(q)$ only to the real part. Retrospectively, this had no effect on the measured correlation values but caused a point symmetry in RS.

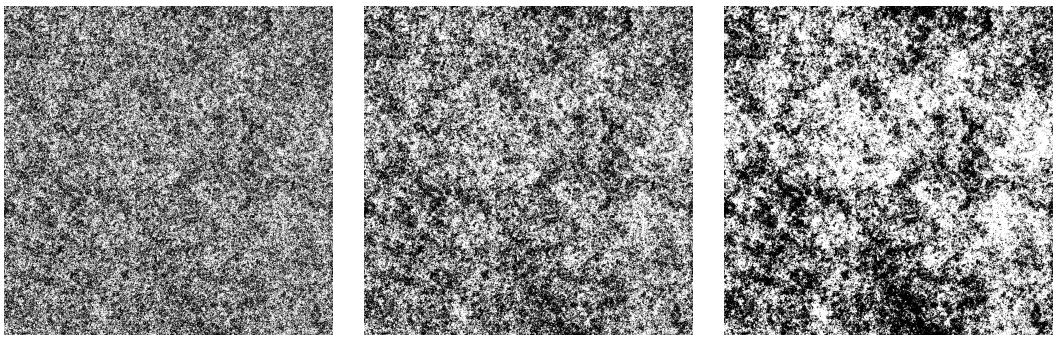


a) b)

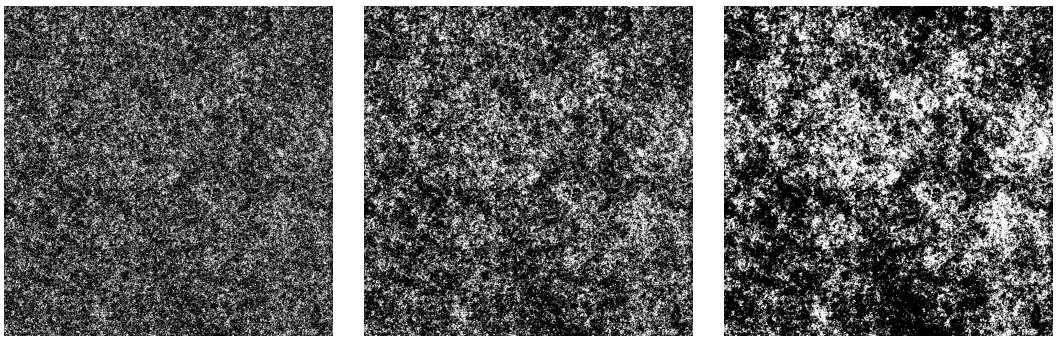
Figure 3.12 Two mapped sequences of size $2^7 \times 2^7$, both were created using the same seed for the random number generator. During the correlation, in case a) the imaginary part in FS was set to zero, while in b) $S(q)$ was applied to both parts of the complex numbers.



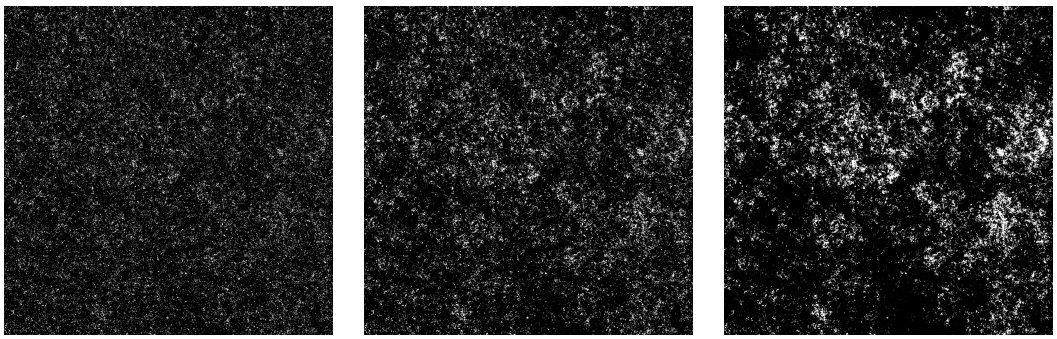
a) $\gamma = 1.2$, $p = 0.3$ b) $\gamma = 0.8$, $p = 0.3$ c) $\gamma = 0.4$, $p = 0.3$



d) $\gamma = 1.2$, $p = 0.5$ e) $\gamma = 0.8$, $p = 0.5$ f) $\gamma = 0.4$, $p = 0.5$



g) $\gamma = 1.2$, $p = 0.7$ h) $\gamma = 0.8$, $p = 0.7$ i) $\gamma = 0.4$, $p = 0.7$



j) $\gamma = 1.2$, $p = 0.9$ k) $\gamma = 0.8$, $p = 0.9$ l) $\gamma = 0.4$, $p = 0.9$

Figure 3.13 A system of size $2^{11} \times 2^{11}$ for increasing correlation - left to right - and increasing occupation probability p - top to bottom. Occupied sites are black, unoccupied ones white. The same seed has been used for the random number generator of the initial uncorrelated data.

3.4 Difficulties

After ensuring that any obvious mistakes like the implementation of the FFT, proper normalization and correct indexing when applying $S(q)$ were avoided, a few major issues arose that required further thought. For one, even though $C(l)$ is well defined in RS, $S(q)$ in FS still has the analytical solution $+\infty$ for $q = 0$, which is caused by the Bessel function. This numerically impractical obstacle could be tackled in various ways, the most innate approach was to use a very small substitue value for q close to zero, as suggested in the reference [6]. The problem with this ansatz is, that the resulting correlation is heavily dependant on the particular value chosen, especially for strong correlations, i.e. small γ. See Fig. 3.14.

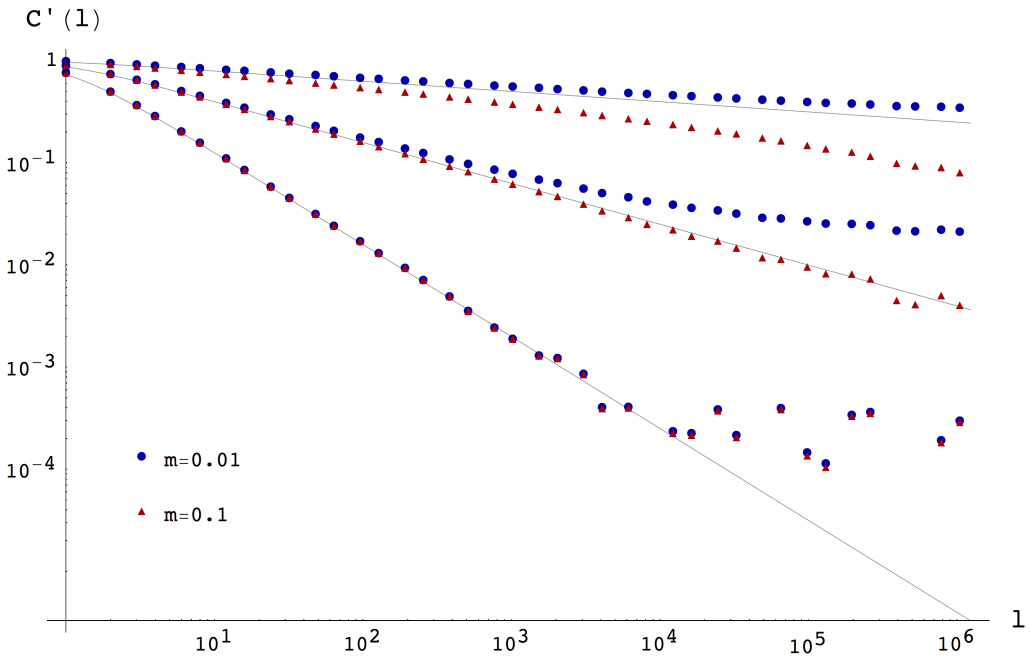


Figure 3.14 Comparison of the resulting correlations in a 1D system of size 2^{21} for different substitute arguments at the singularity. The influence of $S(0)$ is strongest for small γ and decreases as $\gamma \approx d$, from top to bottom: $\gamma = 0.1, 0.4, 0.9$.

We tried several other methods to settle with an analytical expression for $S(0)$ that would deliver correct results, independant of system size and correlation but none would work sufficently. Here, we denote the spectral density by:

$$S_m = S\left(\frac{2\pi}{N}\, m\right) = S(q)$$

First, a simple cut off, which might be promising for $N \to \infty$, where $S(0) = S_1$, the next neighbouring element in the array, gave undercorrelated results for our realizable system sizes. Similarly using an approach where $S(0)$ was determined by a linear interpolation of S_1 and S_2, was undercorrelated, too. A third ansatz, that works decently well for 2D systems $d = 2$ of size $N = 2^{11}$ was to define:

$$S(0) = S_1 \, c^{\gamma - d}, \qquad c = 0.5 \tag{3.14}$$

While c was determined purely by trial and error, the general form of this equation was deduced regarding two arguments. On the one hand, we know that

$$S(q) \propto q^{\gamma - d}, \qquad q \to 0$$

which is explicitly given in [6] for 1D. On the other hand, S_1 seems to be a modest prefactor of the right magnitude, as it is undoubtedly dependant upon the system size and γ.

Conjectures aside, when reapplying this idea to smaller system sizes and 1D we had to realize that it was not satisfactory and had to conclude for now, that the only reliable way to create a correlation that gives exact power-law behaviour is to estimate $S(0)$ empirically. In Table 3.1 you see this applied to a small 1D systems of size $N = 64$ where the value was altered by guessing and checking, until agreeable results were produced.

γ_{in}	γ_{fit}	error	m_{in}
0.1	0.098	0.002	0.0055
0.2	0.204	0.005	0.017
0.3	0.307	0.008	0.03
0.4	0.402	0.010	0.04
0.5	0.501	0.012	0.05
0.6	0.594	0.015	0.055
0.7	0.692	0.016	0.06
0.8	0.801	0.017	0.07
0.9	0.908	0.017	0.08

Table 3.1 Overview of the γ-dependance of substitute m values needed for $S(0)$ in a small 1D system of size 2^6. To estimate m, fits of the form $C'(l) = (1 + l^2)^{\frac{-\gamma_{fit}}{2}}$ were done with γ_{fit} as the fit parameter.

While looking at the distribution of correlated sequences earlier, we did not discuss the standard deviation of the normal distribution but explicitly assumed $\sigma = 1$ for $N \to \infty$ in Eq. 3.13. To fortify this claim, Fig. 3.15 shows the dependance of the standard deviation of correlated 2D sequences upon the system size. It is noteworthy that $\gamma \approx d/2$ yields values for σ closest to 1. Larger γ produce values in the range of the ones for $0.3 < \gamma < 0.9$, which have not been included to keep the figure clear. As we see, the data shows the presumed bias, but for systems of small sizes, the standard deviation is not of the described scale. This has the following effect on the mapped sequences: The occupation probabibility of a site, that can be measured after mapping - by counting in the spirit of Eq. 3.12 - is different from the probabibility desired and given as an input parameter. A way to fix this for such "small" systems would be, to generate the needed amount of differently seeded sequences first, then calculate the average σ during runtime and use it to adjust the mapping. Having said that, using such a method in a real world application quickly leads to very calculation and memory intensive programs, as either the whole set of sequences has to stay in memory before the mapping can be done or each sequence has to be generated twice - which is the most expensive part since two FFTs are performed during this step.

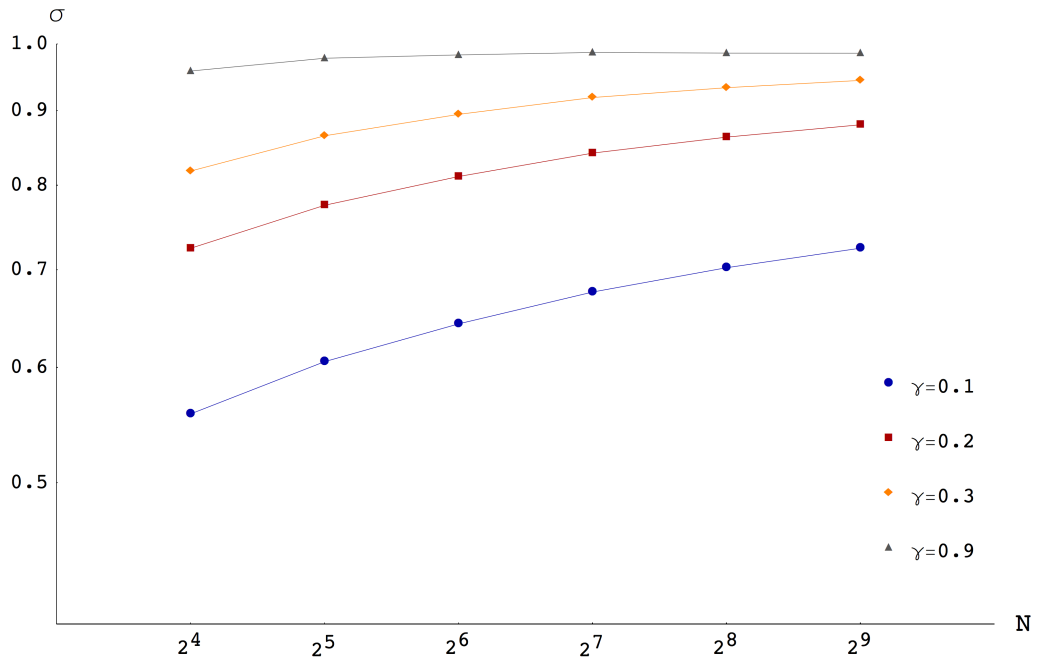


Figure 3.15 σ-dependance on the system size of 2D correlated sequences. Average values were taken from 1000 samples with different seeds for the random number generator.

4 Outlook

In the course of this bachelor thesis a large part of work was put into understanding and reproducing the publications by Makse et al. [6, 10]. Due to the extensiveness and amount of detail involved with generating long-range correlations, this became a very conceptual study and had to be done with comparably small sets of underlying data. Therefore it would be nice to check whether observations made so far remain unchanged for systems of greater sizes and averages taken from a larger basis, which requires the source code to be optimized. Undoubtedly the challenge of finding a way to treat the singularity $S(0)$ remains. It seems highly inconvenient to determine any parameter in a computer simulation by trial and error or best guess. If no basic ansatz can be found, an approach similar to the one proposed for σ in the previous section could be considered. That is to monitor during runtime how accurate the currently used substitute value is and to adjust accordingly. Even though this would not be scientific in the sense of solving problems by understanding them, it could still give the desired correlation, which is the initial objective after all. Future effort could go into the treatment of 3D or even higher dimensional systems, but there is also room to improve the techniques used in 2D. The correlation $C'(l)$ could be measured in arbitrary directions or one could create anisotropic correlations. As a final outlook we want to mention the interest in exactly enumerating a random walk in the power-law correlated confinement. Study of this matter would link the rather theoretical topic of correlation to a more everyday application, since random walks are commonly used to simulate polymers, for instance. Whereas enumerations of such walkers on a plain square lattice have been done for years, there is still a great interest to date and vast studies of related and subsequent problems, such as adding confinement to the lattice.

Acknowledgments

I want to thank M. Sc. Johannes Zierenberg, who supervised me during the past weeks and was always happy to help with advise, patient discussions and cunstructive criticism. Special thanks go to Prof. Dr. Wolfhard Janke for offering the opportunity to work on this very interesting subject and enabling me to write this thesis in the CQT group. I thank Dr. Stefan Schnabel for acting as referee. I am very grateful for the support from friends and family, especially Julia Tesch, Stefan Horn, Dorian Nothaaß and Christoph Pohl for various discussions about the thesis, programming and physics in general. There were many more helpful influences, who even though not explicitly named are very much appreciated.

Bibliography

[1] A. Weinrib and B. I. Halperin, "Critical phenomena in systems with long-range-correlated quenched disorder," *Physical Review B*, vol. 27, no. 1, 1983.

[2] S. Prakash, S. Havlin, M. Schwartz, and H. E. Stanley, "Structural and dynamical properties of long-range correlated percolation," *Physical Review A*, vol. 46, no. 4, pp. R1724–R1727, 1992.

[3] A. Weinrib, "Long-range correlated percolation," *Physical Review B*, vol. 29, no. 1, pp. 387–395, 1984.

[4] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, "Communication activity in a social network: relation between long-term correlations and inter-event clustering.," *Scientific Reports*, 2012.

[5] K. J. Schrenk, N. A. M. Araújo, R. M. Ziff, and H. J. Herrmann, "Retention capacity of correlated surfaces," *Physical Review E*, vol. 89, no. 6, p. 062141, 2014.

[6] H. A. Makse, S. Havlin, M. Schwartz, and H. E. Stanley, "Method for generating long-range correlations for large systems," *Physical Review E*, vol. 53, no. 5, pp. 5445–5449, 1996.

[7] F. Radicchi, "Underestimating extreme events in power-law behavior due to machine-dependent cutoffs," *arXiv:1405.0058*, 2014.

[8] N. Fricke, J. Bock, and W. Janke, "Diffusion and Polymers in Fractal, Disordered Environments," *diffusion-fundamentals.org*, vol. 20, 2013.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C+," 2002.

[10] H. A. Makse, S. Havlin, H. E. Stanley, and M. Schwartz, "Novel method for generating long-range correlations," *Chaos, Solitons & Fractals*, vol. 6, pp. 295–303, 1995.