From phenomenological modelling of anomalous diffusion through continuous-time random walks and fractional calculus to correlation analysis of complex systems

Dissertation
zur Erlangung des Doktorgrades der Naturwisschaften
(Dr. rer. nat.)

dem Fachbereich Chemie
der Philipps-Universität Marburg
vorgelegt von
Daniel Fulger
aus Vaslui

Marburg/Lahn 2009

Vom Fachbereich Chemie der Philipps-Universität Marburg als Dissertation am

....

angenommen.

Erstgutachter: Prof. Dr. Guido Germano
Zweitgutachter: Prof. Dr. Enrico Scalas

Tag der mündlichen Prüfung: 20. März 2009

> *'But I don't want to go among mad people,' said Alice. 'Oh, you can't help that,' said the cat. 'We're all mad here.'* ...
>
> Charles Lutwidge Dodgson alias Lewis Carroll

Dodgson condensation is a method of computing the determinants of square matrices, named after its inventor.

Contents

List of Tables

List of Figures

Zusammenfassung

Diese Arbeit enthält einen Themen*komplex.* Bei genauerer Betrachtung ergeben sich starke Zusammenhänge entweder durch physikalische Analogie, numerische Analogie oder Verwendung der gleichen Methodik oder Analyse bei unterschiedlicher Fragestellung. Desweiteren sind auch einige Teile der vorgestellten Ergebnisse oder neuen Methoden Bausteine innerhalb von anderen. Die übergeordneten Themen sind: Anomale Diffusion, Modellierung stilisierter Fakten[1] mittels eines random walk basierten Diffusionsmodells, Null-Hypothesentests in der Datenanalyse unter Verwendung des selbigen Diffusionsmodells und Korrelationsmatrizen ungekoppelter random walks. Es wird gezeigt, dass das Diffusionsmodell mathematisch rigoros die Zeit- und Raum-fraktionale Diffusionsgleichung löst. Letzteres ist ein neues Resultat aus der mathematischen Literatur. Mittels der vorgestellten Methode wird die Monte-Carlo-Lösung um ein Vielhundertfaches im Vergleich zu bisherigen Methoden beschleunigt. Die darauf folgende Behandlung rotationsinvarianter Zufallsfelder ist ein Baustein auf dem Weg zu Korrelationsmatrizen, die aus eben jenen random walks gewonnen werden. In diesem Zusammenhang wird ein analytisches Ergebnis aus der Literatur numerisch validiert. Innerhalb dieser Themen befindet sich ein Kapitel über die Erzeugung von Zufallszahlen. Darin werden zwei neue Methoden vorgestellt, eine davon zur schnellen Erzeugung univariater reeller Zufallszahlen mit fast beliebiger Dichte. Die Behandlung von Zufallszahlen ergibt sich aus dem Bedarf, millionenfach synthetische Diffusionsprozesse zu simulieren. Eine weitere dritte Methode aus dem Zufallszahlenbereich und ebenso neu, erzeugt rotationsinvariante Matrizen, und spart Matrixrotationen und Summationen ein. Rotationsinvariante Matrizen werden hundertfach für die obig erwähnte Validierung gebraucht.

Es wird dargelegt, dass das Diffusionsmodell geeignet ist zur Modellierung empirischer Zeitreihen mit ggf. asynchronen Datenpaaren. Die random walk Prozesse sind in einer Dimension interpretierbar als synthetische Zeitreihen. Unter gewissen steuerbaren Voraussetzungen stellen diese eine Null-Hypothese dar, das heisst insbesondere hier, dass das Null-Hypothesen-Eigenwertspektrum der zugehörigen Korrelationsmatrix statistisch genau durch Mittelung über viele Realisierungen berechnet werden kann. Die zwei darauf folgenden Kapitel untersuchen den Einfluss der Wartezeiten des random walkers wie auch zwei Korrelationsschätzer, deren einer, eine neue mathematische Entwicklung, direkt auf asynchronen Datenpaaren funktioniert. Bis zum Schluss finden sich die Eigenwertspektren wieder. Ein Kapitel untersucht im Detail die Eigenschaften der Korrelationsmatrizen aus Zeitreihen wie auch der Spektren unter kontrolliert eingeführten Korrelationen. Gewisse Eigenschaften werden anscheinend bislang nicht beachtet, insbesondere in der Klassifizierung über relle Abstandsmasse. Ein Szenario sind microarray-Daten zur Aufdeckung funktioneller genetischer Gruppen. Im letzten Kapitel erweist sich das Eigenwert- und Eigenvektorspektrum ebenso nützlich in der Analyse diesmal nicht korrelierter, aber synchroner oszillatorischer Zeitreihen. Die Datenquelle sind hier elektroenzephalographische Aufnahmen aus dem Schlaflabor und die Matrix enthält Synchronisationskoeffizienten.

---

[1] Engl: stylised facts. Bislang scheint es keine andere deutsche Übersetzung zu geben als die wörtliche.

Chapter 1

Overview

Just the place for a Snark! I have said it twice:
That alone should encourage the crew.
Just the place for a Snark! I have said it thrice:
What I tell you three times is true.
L.C.

1.1 Complex system eclecticism

If the reader is really interested in certain parts and topics of this document a direct answer to the point can save a lot of time. I did not get many.

This document contains more than one topic, but they are all connected in either physical analogy, analytic/numerical resemblance or because one is a building block of another. The topics are anomalous diffusion, modelling of stylised facts based on an empirical random walker diffusion model and null-hypothesis tests in time series data-analysis reusing the same diffusion model. Inbetween these topics are interrupted by an introduction of new methods for fast production of random numbers and matrices of certain types. This interruption constitutes the entire chapter on random numbers that is purely algorithmic and was inspired by the need of fast random numbers of special types. The sequence of chapters is chronologically meaningful in the sense that fast random numbers are needed in the first topic dealing with continuous-time random walks (CTRWs) and their connection to fractional diffusion. The contents of the last four chapters were indeed produced in this sequence, but with some temporal overlap.

While the fast Monte Carlo solution of the time and space fractional diffusion equation is a nice application that sped-up hugely with our new method we were also interested in CTRWs as a model for certain stylised facts. Without knowing economists [80] reinvented what physicists had subconsciously used for decades already. It is the so called stylised fact for which another word can be empirical truth. A simple example: The diffusion equation gives a probability at a certain time to find a certain diffusive particle in some position or indicates concentration of a dye. It is debatable if probability is physical reality. Most importantly, it does not describe the physical system completely. Instead, the equation describes only a certain expectation value of interest, where it does not matter if it is of grains, prices or people which diffuse away. Reality is coded and “averaged” in the diffusion constant.

Interpreting a CTRW as an abstract microscopic particle motion model it can solve the time and space fractional diffusion equation. This type of diffusion equation mimics some types of anomalous diffusion, a name usually given to effects that cannot be explained by classic stochastic models. In particular not by the classic diffusion equation. It was recognised only recently, ca. in the mid 1990s, that the random walk model used here is the abstract particle based counterpart for the macroscopic time- and space-fractional diffusion equation, just like the “classic” random walk with regular jumps $\pm\Delta x$ solves the classic diffusion equation. Both equations can be solved in a Monte Carlo fashion with many realisations of walks. Interpreting the CTRW as a time series model it can serve as a possible null-hypothesis scenario in applications with measurements that behave similarly. It may be necessary to simulate many null-hypothesis realisations of the system to give a (probabilistic) answer to what the “outcome” is under the assumption that the particles, stocks, etc. are not correlated.

Another topic is (random) correlation matrices. These are partly built on the previously introduced continuous-time random walks and are important in null-hypothesis testing, data analysis and filtering. The main objects encountered in dealing with these matrices are eigenvalues and eigenvectors. The latter are carried over to the following topic of mode analysis and application in clustering. The presented properties of correlation matrices of correlated measurements seem to be wasted in contemporary methods of clustering with (dis-)similarity measures from time series. Most applications of spectral clustering ignores information and is not able to distinguish between certain cases. The suggested procedure is supposed to identify and separate out clusters by using additional information coded in the eigenvectors. In addition, random matrix theory can also serve to analyse microarray data for the extraction of functional genetic groups and it also suggests an error model. Finally, the last topic on synchronisation analysis of electroencephalogram (EEG) data resurrects the eigenvalues and eigenvectors as well as the mode analysis, but this time of matrices made of synchronisation coefficients of neurological activity.

Momentarily riding on the fashion wave the “topic” Complex Systems might soon be as outdated as chaos theory, the designated explanation for everything ca. 15 years ago. Surprisingly it was not the explanation for everything, while the surprise manifested in a many year long decay of attention due to persistent refusal of the theory to give answers. So far, research and in particular modelling is very data-driven while some recent models are said to produce some stylised facts. Or rather: stylised results. The content presented here is conservative. It is a collection of results and numerical methods which are partly spin-offs and came up while working on other issues related to the main project. Actually, the title is almost misleading and stems from the fact that some data treated here comes from such a complex system while the definition of the latter is quite open. One attempt could be: A complex system cannot be adequately described in terms of one or a few simple types of interacting objects, neither can we find a finite set of sensible variables to define a statistical macroscopic limit. In other words, it provides new effects across many different scales of time, space, number, energy, colour, money, etc.. This statement intentionally contradicts the intuitive definition of some contemporaries who

a) associate two things with complexity: chaos and a large number of “objects” or better degrees of freedom, and

b) associate (just) a large number of “steps” as a prerequisite of a suitable statistical description of a complex system.

In fact, a more or less ideal gas is the prime example of a non-chaotic, non-complex system with many degrees of freedom. Moreover, chaos in (as) many system degrees of freedom is contradictory by the definition of chaos, the latter defined for good reason as a low dimensional effect in far less than all available degrees of freedom, actually $\mathcal{O}(1)$.

The adequate description of complex systems mentioned above, however, is a rather gloomy concept. In fact it is exactly the loop-hole that allows us to do at least some calculation and define stylised facts as we please. The realism in dealing with such topics hindered the people involved in the work summed up here to come up with the explanation of everything. There are other researchers who are not as resignative. Instead we used, improved, developed or just explained down-to-earth scientific methods and mathematical objects and applied some to real world data in a conservative manner. This real world data can be considered “physical”. Unfortunately, or maybe fortunately, the complex systems are not friendlier than the not so complex ones and also produce the same kind of files with columns of numbers as in many other physical systems, at best.

To underline the scientific respectability the chapters start with a suitable quotation. Despite being abused many times in such cases I truly think that C. L. Dodgson is not really understood. It is worse with Kafka, although he would fit in many ways, too.

1.2 Papers and logistics

Some chapters correspond to the following published, submitted, drafted and hopefully to be published papers. They can be recognised by the title:

D. Fulger, E. Scalas, G. Germano, *Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation*, Physical Review E, **77**, 021122 (2008)

D. Fulger, E. Scalas, G. Germano, *Efficient generation of Lévy alpha-stable random numbers from the density tails*, submitted to Mathematics of Computing (2009)

D. Fulger, G. Germano, *Automatic generation of non-uniform random variates for arbitrary pointwise computable probability densities by tiling*, submitted to Mathematics of Computing (2009)

M. Politi, E. Scalas, D. Fulger, G. Germano, *Spectral densities of Wishart-Lévy free stable random matrices*, submitted to European Physical Journal B (2008)

U. Welling, G. Germano, F. Demmel, D. Fulger, W.-C. Pilgrim, *Collective dynamics in molten alkali halides*, to be submitted to Physical Review B

The last paper is missing in this document. The remaining ca. 50% of the pages consists of newer material that did not go through any review process in many ways. The material was also presented at some conferences, either as talks or just posters:

D. Fulger, *Random matrix analysis of an artificial stock market modelled with continuous-time random walks*, START Workshop Stochastic Analysis and Related Topics, Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, July 2006

D. Fulger, E. Scalas, G. Germano, *Correlation Matrices of Continuous-time Random Walks*, Jahrestagung der Deutschen Physikalischen Gesellschaft (DPG), Regensburg, March 2007

D. Fulger, E. Scalas, G. Germano, *Correlation matrices of synthetic continuous-time random walks*, CEF 2007, 13th International Conference on Computing in Economics and Finance of the Society of Computational Economics, Montréal, March 2007

G. Germano, D. Fulger, E. Scalas, *Efficient Monte Carlo simulation of high-frequency financial time series modelled with anomalous diffusion*, CEF 2007, 13th International Conference on Computing in Economics and Finance of the Society of Computational Economics, Montreal, March 2007

D. Fulger, *Correlation matrices of continuous-time random walks*, Institute for Scientific Interchange, July 2007, Torino, Italy (invited talk)

D. Fulger, E. Scalas, G. Germano, *Efficient Monte Carlo simulation of high-frequency financial time series and correlation matrix analysis*, ECCS 07 (European Conference on Complex Systems), Dresden, October 2007

D. Fulger, *Data analysis and modeling paradigms*, DAPHNet consortium review meeting, Bruxelles, February 2008 (invited talk)

G. Germano, D. Fulger, E. Scalas, *Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation*, workshop "Modelling anomalous diffusion and relaxation: From single molecules to the flight of albatrosses.", Jerusalem, March 2008

D. Fulger, *Lotka-Volterra equations for stylised fact modelling in physiological measurements*, DAPHNet consortium ATR review, Madrid, May 2008

G. Germano, D. Fulger, E. Scalas, *Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation*, Materialforschungstag Mittelhessen 2008, Marburg, June 2008 (poster)

U. Welling, F. Demmel, W.-C. Pilgrim, D. Fulger, G. Germano, *Collective dynamics in molten alkali halides*, Materialforschungstag Mittelhessen 2008, Marburg, June 2008 (poster)

G. Germano, D. Fulger, M. Politi, E. Scalas, *Synthetic high-frequency financial time series: Monte Carlo simulation with uncoupled CTRWs or GARCH-ACD processes, pricing of short to maturity options, and correlation analysis with random matrix theory*, CEF 2008, 14th International Conference on Computing in Economics and Finance of the Society of Computational Economics, Paris, June 2008

G. Germano, D. Fulger, E. Scalas, *Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation*, 3rd Rhein-Main Modelling Meeting (R3M), Frankfurt, July 2008 (poster)

U. Welling, F. Demmel, W.-C. Pilgrim, D. Fulger, G. Germano, *Collective dynamics in molten alkali halides*, 3rd Rhein-Main Modelling Meeting (R3M), Frankfurt, July 2008 (poster)

G. Germano, D. Fulger, E. Scalas, *Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation*, 1st Annual Conference of the EPSRC Network “Mathematical Challenges of Molecular Dynamics”, Warwick, July 2008 (poster)

U. Welling, F. Demmel, W.-C. Pilgrim, D. Fulger, G. Germano, *Collective dynamics in molten alkali halides*, 1st Annual Conference of the EPSRC Network “Mathematical Challenges of Molecular Dynamics”, Warwick, July 2008 (poster)

D. Fulger, *Synchronisation networks of EEG signals*, Racach Institute of Physics, Hebrew University of Jerusalem, August 2008 (invited talk)

E. Scalas, G. Germano, M. Politi, D. Fulger, *Synthetic high-frequency financial time series: Numerical study of free random Wishart-Lévy matrices*, Econophysics Colloquium 2008, Kiel, August 2008

D. Fulger, *Networks of synchronisation EEG activity*, DAPHNet consortium meeting, Berlin, January 2008

D. Fulger, E. Scalas, G. Iori, M. Politi, G. Germano, *A numerical analysis of eigenvalues and eigenvectors of covariance matrices*, Jahrestagung der Deutschen

Physikalischen Gesellschaft (DPG), Dresden, March 2009

The work involved several people, mostly the coauthors of the papers created on the way listed above. Also directly involved in addition to these are (in random order):

- Dr. Ronny Bartsch from the Department of Physics at the Bar-Ilan University in Tel Aviv (within DAPHNet),
- Prof. Jan Kantelhardt from the Department of Physics at the Martin-Luther University Halle (within DAPHNet),
- Prof. emeritus David Breé from the Department of Computer Science at the University of Manchester and from the Institute for Scientific Interchange in Torino (within DAPHNet),
- Prof. Giulia Iori from the Department of Economics at the City University London (within a British Council Researcher Exchange Program grant).

Fig. 1.1 gives a satellite overview of the places where the pieces of work where done.

Figure 1.1: Where the work was done. On a logarithmic scale the font size is an estimate of the respective amount.

The contributions of this document and related papers to make the world a better place are:

- The introduced method for the Monte Carlo solution of the time- and space-fractional diffusion equation speeds up the existing methods by at least a factor of 500. The bold inquiry by the author at the local stochastic mathematicians gave us the right hint. Scream and you get all.

- The two methods on non-uniform uni-variate random number production are new as well as the interpretation of transforms as maps that can be used for tail sampling in the Ziggurat method.
- The proposed method to generate directly rotationally invariant (free in the limit of large size) matrices for power-tailed distribution of elements with the potential of saving orders of magnitudes of time.
- A comparison of the Pearson and Fourier correlation estimators in the context of spectra of correlation matrices of continuous-time random walks was not available up to now.
- We are not aware of some other published reasoning leading to Eq. (7.9) and to the statements on the influence of CTRW-waiting-times on the correlation matrix spectrum.
- The numerical validation of the analytic spectrum of the free Wishart-Lévy ensemble.
- There is by far no such didactic explanation of the main elements of free probability theory around the Wishart matrix ensemble.
- In the context of correlation matrix analysis of time series we show that the random bulk of eigenvalues is less random than usually claimed. Furthermore, the example with empirical data questions the assumptions on the market mode.
- The properties of the Wishart and sample correlation matrix ensemble presented here seem to have gone unnoticed and question certain results in the existing field of genetic profiling with microarrays.
- Synchronisation analysis of electroencephalographic activity using eigenvalues and eigenvectors of the matrix of synchronisation coefficients for principal component selection seems a new idea and we do find sleep stage significant structures.

Chapter 2

Continuous-time random walks and anomalous diffusion – two birds with one stone

> *But I was thinking of a way To multiply by ten, And always, in the answer, get The question back again.*
>
> *L.C.*

This chapter presents a numerical method for the Monte Carlo simulation of uncoupled continuous-time random walks with a Lévy α-stable distribution of jumps in space and a Mittag-Leffler distribution of waiting-times. We apply it to the stochastic solution of the Cauchy problem for a partial differential equation with fractional derivatives both in space and time. The one-parameter Mittag-Leffler function is the natural survival probability leading to time-fractional diffusion equations. Transformation methods for Mittag-Leffler random variables were found later than the well-known transformation method by Chambers, Mallows and Stuck for Lévy α-stable random variables and so far have not received as much attention, nor have they been used together with the latter in spite of their mathematical relationship due to the geometric stability of the Mittag-Leffler distribution. Combining the two methods, we obtain an accurate approximation of space- and time-fractional diffusion processes almost as easy and fast to compute as for standard diffusion processes.

2.1 From random walks to the macroscopic diffusion equation

The well-known standard diffusion equation

$$\begin{aligned} \frac{\partial}{\partial t}u(x,t) &= D\frac{\partial^2}{\partial x^2}u(x,t), \qquad (2.1)\\ u(x,0^+) &= \delta(x), \quad x \in \mathbb{R}, \quad t \in \mathbb{R}_+, \end{aligned}$$

can be used to mimic the change of concentration of a substance or other physically meaningful variables, for example like heat diffusing in some medium. The interpretation in the particle diffusion context in particular allows u to be interpreted as probability density. This equation can be arrived at as a consequence of mass conservation and is analogous to the conservation and flux of probability. The assumption that the one-dimensional flux of a medium is proportional to its space gradient is stated equivalently as:

$$J = -D\frac{\partial u}{\partial x}, \tag{2.2}$$

with a real positive constant of proportionality D. The change of concentration in time must also be proportional to the change of flux in space. In terms of equations this means:

$$\frac{\partial u}{\partial t} = -\frac{\partial J}{\partial x}. \tag{2.3}$$

Put together this gives:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}. \tag{2.4}$$

In three dimensions the derivative operators are replaced by their respective generalisations giving with $\vec{J} = -D\nabla u$ the equation:

$$\frac{\partial u}{\partial t} = D\nabla^2 u. \tag{2.5}$$

The following examples are nevertheless already sufficiently complicated in one dimension. The complete formulation of the “task” includes the initial or boundary condition in time and space. In the example above the initial situation consists of an infinitely narrow conglomeration of all mass in one point at $x = 0$. Likewise, the probability to find the particle there at $t = 0$ is one. Note that time is not symmetric and not reversible. This is a typical asymmetry where the underlying, possibly Newtonian, mechanics of a diffusing particle is reversible but not so the macroscopic substitute equation. The judgement whether such a model is meaningful must always be based on human understanding and cannot be replaced by purely mathematical reasoning.

Continuous-time random walks (CTRWs) and fractional diffusion equations (FDEs), or fractional Fokker-Planck equations respectively are generalised concepts of the regular-time random walk and the classical diffusion equation. The relevance of fractional calculus in the phenomenological description of anomalous diffusion has been discussed within applications of statistical mechanics in physics, chemistry and biology [12, 19, 44, 50, 51, 138, 163] as well as finance [35, 36, 107, 108, 156]; even human travel and the spreading of epidemics were modelled with fractional diffusion [23]. Likewise, it is not uncommon that a good part of the anomalous diffusion community is situated in departments of Chemistry or Sociology. Metzler and Klafter reviewed analytic and numerical methods to solve fractional equations of diffusive type [123]. In Refs. [8, 57, 90, 120, 124, 154, 164, 181, 182] applications and enhancements of these techniques were presented. It is only in the past 20 years that non-Gaussian statistics has received increasing attention. This may be due to a change in the scientific community that began to accept that not everything is Gaussian. The "Gaussian attitude" or better the determination that everything has finite moments keeps things much more simple and was a natural assumption just like the Euclidean space had to be the right one. But the louder knocking and rising number of phenomena that could not be described by Gaussian statistics could not be overheard anymore. It might be of interest that the first theoretical treatment of processes with jumps and trapping times that did not fit Gaussian statistics was produced by Montroll and Scher working partly for Xerox [158]. The problem were electrons in the photoconductor plate being trapped in "holes" for a certain time until released again. Before their work these phenomena were more or less ignored and considered as something "complicated", too difficult to be treated in terms of probability models. Of particular nuisance are processes whose characteristic statistical measurements diverge, typically called super-diffusive. The variables are usually a function of position of moving or scattered objects. Super-diffusion in materials, for example, was observed only very recently in an artificial liquid [10] where the scattering statistics of light could be tuned in such a way that the mean squared displacement with scattering event number n grows as

$$\langle x^2(n)\rangle = \langle |x(N+n)-x(n)|^2\rangle_N = 2dDn^h \tag{2.6}$$

with exponent $1 < h \leq 2$ while for Brownian motion we have $h = 1$. The subscript N at the brackets indicates the average over many starting points which is in principle equivalent to the average over time. Eq. (2.6) is valid for all structure functions. The constant d depends on the dimensionality of the system and D is the diffusion constant with a fractional dimension that depends on h. Sub-diffusion is given if $h < 1$. It becomes apparent that the underlying diffusion model is a step or jump process where the distribution of jumps determines whether we have sub- or super-diffusion. We will see soon that random walk models are equivalent to macroscopic diffusion models in the continuum limit and that a direct Monte Carlo approach to fractional Fokker-Planck dynamics (one of several continuum-limit models for fractional dynamics) via a random walk model requires random waiting-times drawn from the Mittag-Leffler distribution. Since sampling the latter was considered troublesome, different schemes to avoid it were proposed. One possibility consists in replacing it with the Pareto distribution, i.e. its asymptotic power-law approximation for $t \to \infty$ [72]; however, this is limited to long times and the distribution index β not close to 1. A more general alternative is based on subordination [69, 105, 106]. On the contrary, we present a straightforward Monte Carlo method for the efficient simulation of uncoupled CTRWs using an inversion formula for the Mittag-Leffler distribution, and apply it to compute approximate solutions of the Cauchy problem for a generalised diffusion equation that has fractional space and time derivatives.

The most simple step process in one dimension: A point moving back and forth at regular times i a constant distance with equal chance has the following probability to be at position j at the next time step $i+1$:

$$p(j,i+1) = 1/2p(j-1,i) + 1/2p(j+1,i). \tag{2.7}$$

This equation is even intuitive, it controls probability in time and is sometimes called the master equation, though a very simple example. Note that this process is abstract and dimension free. Once we associate i with real time t ($i+1 \equiv t+\Delta t$) and position j with space x ($j+1 \equiv x+\Delta x$) in some unit of choice we can describe something real. Only then a diffusion constant makes sense to be introduced. The factors 1/2 are only to normalise both jump directions to probability one overall, that the object *will* move, and that the direction is fifty-fifty. In the limit of small jumps and short waiting-times between the jumps we can expand p in Taylor manner around x and t:

$$p(x \pm \Delta x, t) = p(x,t) \pm \Delta x \frac{\partial p(x,t)}{\partial x} + \frac{(\Delta x)^2}{2}\frac{\partial^2 p(x,t)}{\partial x^2} + \text{higher orders}, \tag{2.8}$$

$$p(x, t+\Delta t) = p(x,t) + \Delta t \frac{\partial p(x,t)}{\partial t} + \text{higher orders}. \tag{2.9}$$

The series end at the powers that are needed to show that once the above equations are equated we obtain

$$\frac{\partial}{\partial t} p(x,t) = \frac{(\Delta x)^2}{2\Delta t}\frac{\partial^2}{\partial x^2} p(x,t), \tag{2.10}$$

the classical diffusion equation. In the continuum limit we can abbreviate the constant

$$D = \lim_{\Delta t \to 0, \Delta x \to 0} \frac{(\Delta x)^2}{2\Delta t}, \tag{2.11}$$

and call it diffusion constant with the unit $[D] = m^2/s$. This can also be argued from the "other end": The central limit theorem ensures that a sufficient number of jumps (actually, it is around 12, a number to reappear as well as the central limit theorem in a different context) makes the resulting distribution of particles indistinguishable from a Gaussian if they start all at $x = 0$. The solution in time of Eq. (2.10) is the Gaussian:

$$p(x,t) = \frac{1}{2\sqrt{\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{2.12}$$

Fourier was the first to solve the (equivalent) heat diffusion equation more generally for non-periodic boundary conditions while having in mind the expansion of non-periodic functions in orthogonal harmonics. Up to then, only solutions for simple cases like annuli with sine or cosine initial conditions were known. It is helpful to note that the characteristic function of Eq. (2.12)

$$\hat{p}(k,t) = e^{-Dk^2 t} \tag{2.13}$$

is an eigenfunction of the time derivative operator in Fourier space:

$$\frac{\partial}{\partial t}\hat{p}(k,t) = -Dk^2\hat{p}(k,t). \tag{2.14}$$

The most straight forward choice of jump lengths is a Gaussian probability density. This is a choice in so far as the convergence of a random walk diffusion model to the solution of Eq. (2.10) is fastest in the sense that the smallest number of jumps is needed. Any choice of distribution with finite moments produces an equivalent result in the hydrodynamic limit. The situation is different if the moments are not finite. A typical distribution with (some) non-existent moments, i.e. diverging variance due to the divergence of the respective integrals, is the symmetric Lévy distribution whose characteristic function is given by

$$\hat{\lambda}(k) = e^{-(\gamma_x|k|)^\alpha} \approx 1 - (\gamma_x|k|)^\alpha \tag{2.15}$$

corresponding to the asymptotic power-tail behaviour in real space:

$$\lambda(x) \approx \frac{\gamma_x^\alpha}{|x|^{1+\alpha}}. \tag{2.16}$$

A scaling factor γ_x alows to adapt the distribution to the physical situation and the scaling consequentially appears in the diffusion constant. The distribution contains the Gaussian distribution as a special case with $\alpha = 2$. The focus on the tails is at the same time a statement that assigns a lower importance to the body of the distribution. The outcome in the hydrodynamic limit in time and space is entirely determined by the tail behaviour, an issue that is repeated in other contexts in later chapters. However, the choice of Lévy alpha-stable random numbers is less random than it appears because the Lévy distribution defines the most general stable law. Details follow in subsequent chapters. A good impression of some of the properties of a random walk with this distribution of jumps can be achieved in two dimensions. Fig. 2.1 shows a realisation of a Brownian motion with Gaussian jumps compared to a realisation of a Lévy flight, as the process with instantaneous jumps would then be called. The Brownian particle needs a finite time to paint any region completely black, whereas the Lévy particle cannot guarantee this even after an infinite time. With an increasing number of jumps the appearance does not change. All processes are self-similar, while the fractal dimension α of the Brownian motion is 2 in accordance with area, the dimension of the Lévy flight is lower, in this case $\alpha = 1.5$. The random walk in two dimensions was first described 1905 by Pearson as a model for mosquito infestation [142].

The first order approximation in Eq. (2.15) becomes useful if we notice through it that the Fourier transformation seems to provide a vehicle to define a fractional order derivative in terms of known mathematical objects. We can write

$$\frac{\mathrm{d}^\alpha}{\mathrm{d}|x|^\alpha} f(x) = \mathcal{F}_x^{-1}\left[-|k|^\alpha \hat{f}(k)\right](k), \quad 1 < \alpha \leq 2, \tag{2.17}$$

where $\mathcal{F}_x^{-1}$ denotes the inverse Fourier transformation

$$\mathcal{F}_x[f(x)](k) = \int_{-\infty}^{+\infty} f(x) e^{ikx} \,\mathrm{d}x = \hat{f}(k). \tag{2.18}$$

It is a matter of definition whether to use the direct or the inverse Fourier transformation. This definition of a fractional derivative is somewhat empirical but leads to a suggestive version of the diffusion equation which may model the Lévy flight in the hydrodynamic limit:

$$\frac{\partial}{\partial t} p(x,t) = D \frac{\partial^\alpha}{\partial |x|^\alpha} p(x,t) \tag{2.19}$$

with $D = \gamma_x^\alpha / 1$ where 1 corresponds to the abstract time index. The dimension is $[D] = m^\alpha / s$ if the distribution of jumps is chosen with physical meaning, i.e. if the jumps are measured in units of time. The above fractional space derivative operator is called the Riesz-Feller derivative. For $\alpha = 2$ this reduces to the usual second order derivative. For $\alpha < 2$ the following equation holds:

$$\frac{\mathrm{d}^\alpha}{\mathrm{d}|x|^\alpha} f(x) = \frac{\Gamma(\alpha+1)}{\pi} \sin \frac{\alpha\pi}{2} \int_0^\infty \frac{f(x+\xi) - 2f(x) + f(x-\xi)}{\xi^{\alpha+1}} \,\mathrm{d}\xi. \tag{2.20}$$

The symmetric derivative with respect to $|x|$ and $\alpha \neq 1$ is generally given by

$$\frac{\mathrm{d}^\alpha}{\mathrm{d}|x|^\alpha} g(x) = -\frac{1}{2\cos(\pi\alpha/2)} \left(\left[\frac{\mathrm{d}^\alpha}{\mathrm{d}x^\alpha} \right]_+ + \left[\frac{\mathrm{d}^\alpha}{\mathrm{d}x^\alpha} \right]_- \right) g(x). \tag{2.21}$$

For the special case $\alpha = 1$ the space-fractional derivative is given via the Hilbert transform:

$$\frac{\mathrm{d}^\alpha}{\mathrm{d}|x|^\alpha} g(x) = -\frac{\mathrm{d}}{\mathrm{d}x} H[g(x)], \quad \text{where} \quad H[g(x)] = \frac{1}{\pi} \int_{-\infty}^\infty \frac{g(t)}{x-t} \,\mathrm{d}t. \tag{2.22}$$

The Hilbert transform reappears in the last chapter on the synchronisation analysis of signals. We can also check backwards that we recover the stable law Eq. (2.15) we used as an ansatz: Transforming Eq. (2.19) via Fourier yields

$$\frac{\partial}{\partial t} \hat{p}(k,t) = -|k|^\alpha D \, \hat{p}(k,t), \tag{2.23}$$

and then integrated in time gives

$$\hat{p}(k,t) = \exp(-|k|^\alpha D t), \tag{2.24}$$

which is the characteristic function of the Lévy stable distribution. Note again that any distribution would be fine if the hydrodynamic limit is taken sufficiently far. The above choice, however, simplifies the reasoning.

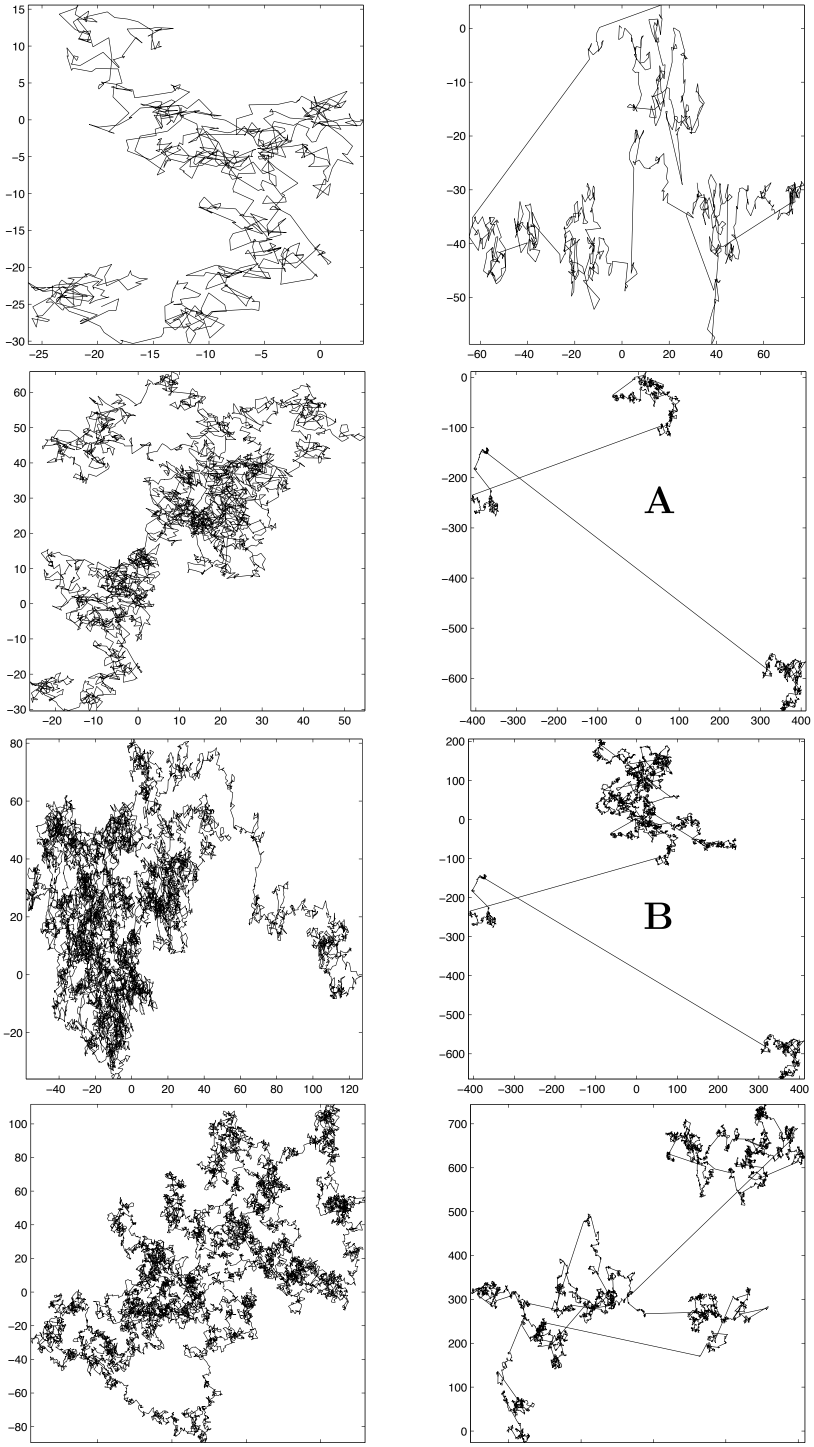


Figure 2.1: Brownian motion (left) and a Lévy flight (right) of a particle with $\alpha = 1.5$ whose position for each time index is connected to the previous with a straight line. The realisations are shown for 1000, 5000, 1000 and 20000 jumps. The difference between A and B is small. During the 5000 jumps the walker spent its time in the upper cluster and behaved Brownian. This alteration of domains is typical for Lévy flights.

2.2 Continuous-time random walks

So far, time is still meaningless and the diffusion constant unit is somewhat undefined. If a jump occurs instantly then time is still only an index i. The real world process under consideration may exhibit jumps (possibly fast but with finite speed) at regular times or according to some other distribution. If the latter has finite moments with scaling factors that are physically meaningful the result is equivalent to Brownian motion. A typical distribution of waiting-times is the exponential distribution, possibly scaled with a factor γ_t to fit reality:

$$\psi(t) = \frac{1}{\gamma_t} \exp(t/\gamma_t). \tag{2.25}$$

Now the diffusion constant is $D = \gamma_x^\alpha/\gamma_t$ such that $[D] = [\gamma_x^\alpha/\gamma_t] = cm^\alpha/s$. Exponentially distributed waiting-times define a Poisson process in which the probability of n events within a fixed period is given by the Poisson distribution. This defines the Normal Compound Poisson Process (NCPP). The exponential density has, by definition, an exponential tail, i.e. the moments are finite, and the dimension of the time unit is still one. Below we will see that a power-tailed waiting-time distribution gives rise to a fractional unit of time.

Prior to the proper introduction of waiting-times the CTRW should be defined properly. A CTRW [128] is a pure jump process; it consists of a sequence of independent identically distributed (iid) random jumps (events) ξ_i separated by iid random waiting-times τ_i,

$$t_n = \sum_{i=1}^{n} \tau_i\,, \quad \tau_i \in \mathbb{R}_+, \tag{2.26}$$

so that the position at time $t \in [t_n, t_{n+1})$ is given by

$$x(t) = \sum_{i=1}^{n} \xi_i, \quad \xi_i \in \mathbb{R}. \tag{2.27}$$

A realisation of the process is a piecewise constant function resulting from a sequence of up or down steps with different height and depth; see Fig. 2.2 for a schematic picture and Fig. 2.3 for example realisations. Jumps are assumed to happen instantaneously or at least within negligible time. In general, jumps and waiting-times depend on each other and they can be described by a joint probability density $\varphi(\xi, \tau)$. The latter appears in the integral equation giving the probability density $p(x, t)$ for the process being in position x at time t, conditioned on the fact that it was in position $x = 0$ at time $t = 0$:

$$p(x,t) = \delta(x)\,\Psi(t) + \int_{-\infty}^{+\infty} \mathrm{d}\xi \int_0^t \mathrm{d}\tau\, \varphi(\xi,\tau)\, p(x-\xi, t-\tau). \tag{2.28}$$

Here the initial condition $x(0) = 0$ is contained implicitly in the first term $\delta(x)\Psi(t)$, where we find the complementary cumulative distribution function (survival function)

$$\Psi(t) = 1 - \int_{-\infty}^{+\infty} \mathrm{d}\xi \int_0^t \mathrm{d}\tau \, \varphi(\xi, \tau). \tag{2.29}$$

Recently, Enrico Scalas, one of the authors of the respective paper in which this material was presented, found an analytic solution of the integral equation in the uncoupled case, i.e. when $\varphi(\xi, \tau) = \lambda(\xi)\psi(\tau)$, where $\lambda(\xi)$ is the jump marginal density and $\psi(\tau)$ is the waiting time marginal density [157]. The master equation then reads

$$p(x, t) = \delta(x)\Psi(t) + \int_0^t \mathrm{d}\tau \; \psi(\tau) \int_{-\infty}^{+\infty} \mathrm{d}\xi \; \lambda(\xi) p(x - \xi, t - \tau). \tag{2.30}$$

It might be instructive to compare Eq. (2.28) and Eq. (2.7). The latter is the most simple version of the master equation for the lattice walk model introduced there. Both equations take all necessary history and surrounding space to give the desired probability for a particular location. While the simple lattice model includes only the neighbouring sites, the integral equation Eq. (2.28) requires the entire space. This provoked some criticism when used as a macroscopic model since it is non-causal or at least non-local. However, one point of view can be that such "microscopic details" are irrelevant if the required stylised fact is represented well, in which case this fact is the probability density of the particle.

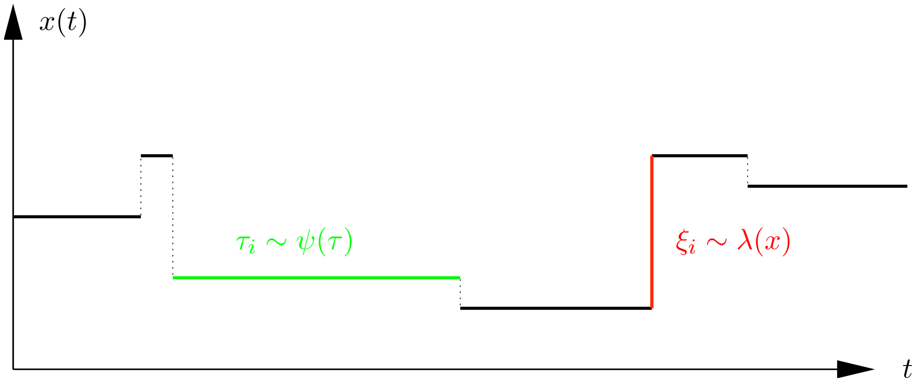


Figure 2.2: Schematic picture of a continuous-time random walk. Waiting-times and jumps are distributed according to the densities $\psi(\tau)$ and $\lambda(x)$ functions.

2.3 The time and space fractional diffusion equation

In the previous section we deduced the "right" jump density to solve the space-fractional derivative in the diffusion equation via an analogy in Fourier space. In this section we deduce the "right" waiting-time density to solve the time-fractional part via an analogy in Laplace space.

The classical diffusion equation can be generalised to the space-time-fractional

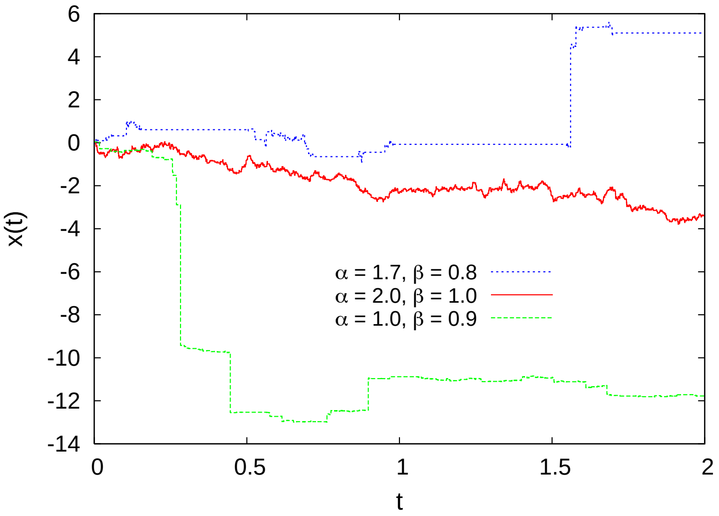


Figure 2.3: Sample paths of CTRWs with scale parameters $\gamma_t = 0.001$, $\gamma_x = \gamma_t^{\beta/\alpha}$ and different choices of α and β. With smaller α the jumps become larger; with smaller β the waiting times become longer.

diffusion equation

$$\begin{aligned} \frac{\partial^\beta}{\partial t^\beta} u(x,t) &= D \frac{\partial^\alpha}{\partial |x|^\alpha} u(x,t) \\ u(x,0^+) &= \delta(x), \quad x \in \mathbb{R}, \quad t \in \mathbb{R}_+. \end{aligned} \tag{2.31}$$

The time and space dependent variable is now again $u(x,t)$ to make clear that it does not necessarily refer to probability. For $0 < \alpha \leq 2$, $\partial^\alpha / \partial |x|^\alpha$ denotes the symmetric Riesz-Feller operator. For $0 < \beta \leq 1$, $\partial^\beta / \partial t^\beta$ is the time-fractional Caputo derivative:

$$\frac{\mathrm{d}^\beta}{\mathrm{d}t^\beta} f(t) = \mathcal{L}_s^{-1} \left[s^\beta \widetilde{f}(s) - s^{\beta-1} f(0^+) \right] (t). \tag{2.32}$$

$\mathcal{L}$ denotes the Laplace transform:

$$\widetilde{f}(s) = \mathcal{L}_t[f(t)](s) = \int_0^\infty f(t) e^{-st} \, \mathrm{d}t, \quad s \in \mathbb{C}. \tag{2.33}$$

The Laplace transform is suggestive as a vehicle to implement the fractional time derivative because time is not symmetric as compared to space. The presence of the initial condition in Eq. (2.32) originates from this reasoning. For more information on the fractional derivatives see Refs. [34, 68, 144, 153, 157]. For $\beta = 1$ this reduces to the usual first order derivative. For $\beta < 1$ the following equation holds:

$$\frac{\mathrm{d}^\beta}{\mathrm{d}t^\beta} f(t) = \frac{1}{\Gamma(1-\beta)} \left[\frac{\mathrm{d}}{\mathrm{d}t} \int_0^t \frac{f(\tau)}{(t-\tau)^\beta} \, \mathrm{d}\tau - \frac{f(0^+)}{t^\beta} \right], \tag{2.34}$$

where $f(0^+)$ is the initial condition. For $\alpha = 2$ and $\beta = 1$, the standard diffusion equation, Eq. (2.1), is recovered. Without loss of generality, we assume $D = 1$; a different value would just mean a scale transformation of space and/or time units. $u(x,t) \geq 0$ is the Green function of the FDE, that has the general form

$$u(x,t) = t^{-\beta/\alpha} \, W(x/t^{\beta/\alpha};\, \alpha, \beta), \tag{2.35}$$

with the scaling function

$$W(\xi;\, \alpha, \beta) = \mathcal{F}_k^{-1} \left[E_\beta(-|k|^\alpha) \right](\xi). \tag{2.36}$$

$E_\beta(z)$ is the one-parameter Mittag-Leffler function [74],

$$E_\beta(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\beta n + 1)}, \quad z \in \mathbb{C}, \tag{2.37}$$

with

$$E_\beta(-t^\beta) = \mathcal{L}_s^{-1} \left[\frac{s^{\beta-1}}{1+s^\beta} \right](t), \quad t \in \mathbb{R}_+. \tag{2.38}$$

The Mittag-Leffler function naturally appears once we note that the survival probability distribution function satisfies the ordinary relaxation equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} \Psi(t) = -\Psi(t), \quad t > 0, \quad \Psi(0^+) = 1. \tag{2.39}$$

Eq. (2.32) represents the fractional generalisation of Eq. (2.39) by writing $\mathrm{d}^\beta/\mathrm{d}t^\beta$. Taking the Laplace transform of the Caputo derivative

$$\mathcal{L}_t \left[\frac{\mathrm{d}^\beta}{\mathrm{d}t^\beta} f(t) \right](s) = s^\beta \widetilde{f}(s) - s^{\beta-1} f(0^+) \tag{2.40}$$

and the latter applied to equation (2.39) gives

$$\widetilde{\Psi}(s) = \frac{s^{\beta-1}}{1+s^\beta} \tag{2.41}$$

which turns out to be the Laplace image of the Mittag-Leffler function. The transform back from Laplace space yields indeed the solution

$$\Psi(t) = E_\beta(-t^\beta) \tag{2.42}$$

in terms of the Mittag-Leffler function $E_\beta(z)$.

For $t \in \mathbb{R}$ and $\beta = 1$, the Mittag-Leffler function with argument $-t^{\beta}$ reduces to a standard exponential decay, e^{-t}; when $0 < \beta < 1$, the Mittag-Leffler function is approximated for small values of t by a stretched exponential decay (Weibull function), $\exp(-t^{\beta}/a)$, where $a = \Gamma(\beta + 1)$, and for large values of t by a power law, $bt^{-\beta}$, where $b = \Gamma(\beta)\sin(\beta\pi)/\pi$; see Fig. 2.5. The Mittag-Leffler distribution is an important example of fat-tailed waiting-times that arises as the natural survival probability leading to time-fractional diffusion equations. There is increasing evidence for physical phenomena [122, 160, 177] and human activities [7, 150, 155] that do not follow neither exponential nor, equivalently, Poissonian statistics. The analytic solution of the fractional diffusion equation for uncoupled CTRWs is derived in Ref. [108] and also shown in Fig. 2.6.

Together with power-tailed waiting-times the diffusion constant obtains the unit $[D] = m^{\alpha}/s^{\beta}$. The exponent of anomalous diffusion h in Eq. (2.6) controls in this model of CTRWs whether we have sub- or super-diffusion as a function of the two fractional derivative orders and is given by:

$$h = \frac{2\beta}{\alpha}. \tag{2.43}$$

Fig. 2.4 shows the range and diffusion types.

It is inevitable to numerically solve a FDE in the most general case, also known as fractional Fokker-Planck equation, that may include space- and time-dependent diffusion and drift terms. Possible approaches are the direct calculation of the integrals in Equations (2.20) and (2.34) [58], finite difference methods [43, 121, 167] and stochastic methods [72, 105, 106, 120, 182]. All of them are complicated, the latter ones mainly because of the supposedly cumbersome generation of Mittag-Leffler random numbers. While this problem has been often worked around in the past, we show how to overcome it, obtaining a fast and accurate method for the Monte Carlo solution of FDEs via uncoupled CTRWs. As a benchmark we focus our attention on the Cauchy problem defined in Eq. (2.31), for which an analytic solution given by Equations (2.35) and (2.36) is available.

2.4 Monte Carlo solution of the fractional diffusion equation

The link between CTRWs and time-fractional diffusion was discussed rigorously in Ref. [73] in terms of the generalised Mittag-Leffler function $E_{\beta,\beta}(-\tau^{\beta})$.

In order to approximate the Green function in Eq. (2.35), it is sufficient to simulate CTRWs whose jumps are distributed according to the symmetric Lévy α-stable probability density (that reduces to a Gaussian for $\alpha = 2$)

$$L_{\alpha}(\xi) = \mathcal{F}_k^{-1}\left[\exp\left(-|\gamma_x k|^{\alpha}\right)\right](\xi) \tag{2.44}$$

and whose waiting-times have the probability density

$$\psi_{\beta}(\tau) = -\frac{\mathrm{d}}{\mathrm{d}\tau} E_{\beta}\left(-(\tau/\gamma_t)^{\beta}\right), \tag{2.45}$$

where $E_\beta(z)$ is the one-parameter Mittag-Leffler function given by Eq. (2.37). Then a weak-limit approximation of the Green function is obtained by rescaling waiting-times by a constant γ_t and jumps by a constant $\gamma_x = \gamma_t^{\beta/\alpha}$, letting γ_t (and as a consequence γ_x) vanish, and plotting the histogram for the probability density $p_{\gamma_x,\gamma_t}(x, t;\, \alpha, \beta)$ of finding position x at time t for the rescaled process. This probability density weakly converges to the Green function $u(x, t;\, \alpha, \beta)$. Weak convergence means that for $x = 0$ a singularity is always present in $p_{\gamma_x,\gamma_t}(x, t;\, \alpha, \beta)$ at $x = 0$ for any finite value of γ_t and γ_x. This singularity is the term $\delta(x)\Psi(t)$ in Eq. (2.28) with $\Psi(t) = E_\beta(-t^\beta)$. In the case $\alpha = 2$ and $\beta = 1$ the CTRWs are an NCPP and, in the diffusive limit, one recovers the Green function for the standard diffusion equation, Eq. (2.1), i.e. the Wiener process. This procedure is justified in Refs. [154] and [157]. In the latter reference, one can also find a theoretical justification for the Monte Carlo procedure where waiting-times are generated according to a power-law distribution; a more complete treatment has been given in Ref. [69].

2.5 Isotropic random walks

Fig. 2.1 shows a random walker who chooses a random direction for each jump. It is a well known technique how to choose a direction at random, or equivalently, random points on a circle, sphere or hyper-sphere. Nevertheless, it should be mentioned here, because we will learn in Chap. 3 that random vectors with Gaussian distribution of elements are isotropically distributed. This concept will be generalised to matrices and be called rotational invariance. Vectors whose elements are independent but do not have a finite variance do not form an isotropic ensemble.

The Matlab code for producing pictures like in Fig. 2.1 is very short:

```
x=[0 ;0];
for i=1:N
   rx = randn(2,1); % vector of normal random numbers
   rx = rx/sqrt(sum(rx.^2)); %normalise to the unit circle
   xp1 = x + levyrnd(alpha,beta,1,0,1,1)* rx; %multiply unit vector with Levy variate
   plot([x(1) xp1(1)],[x(2) xp1(2)],'k'); hold on
   x = xp1;
end
```

The method to produce Lévy alpha-stable random numbers is described later in this chapter and can also be downloaded as piece of Matlab code [64].

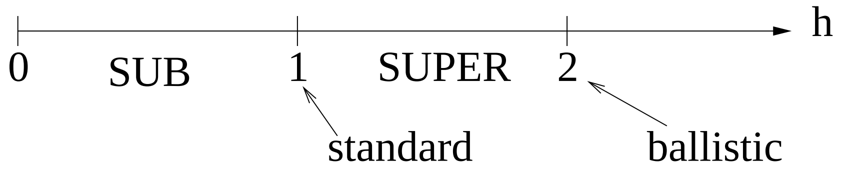


Figure 2.4: The exponent of the time dependent mean squared displacement controls two types of anomalous diffusion. For $h = 1$ we get normal diffusion. In the range $0 < h < 1$ it is sub-diffusion, in the range $1 < h < 2$ super-diffusion. Bigger values $h \geq 2$ are not in the realm of diffusion anymore.

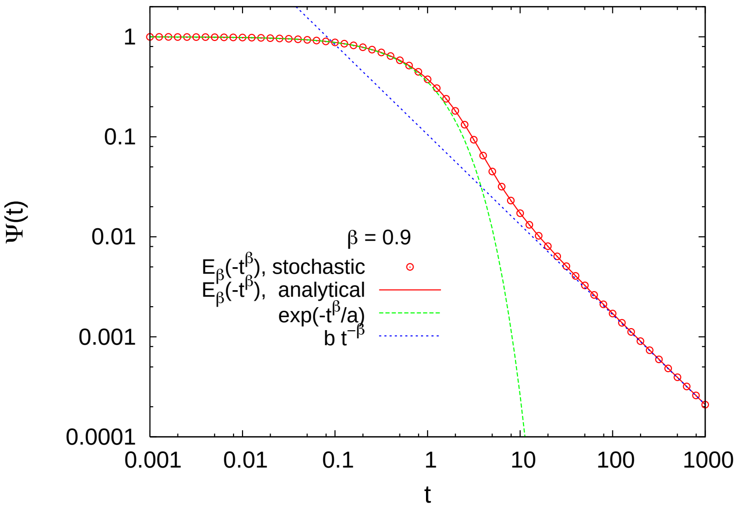


Figure 2.5: The Mittag-Leffler complementary cumulative distribution function sampled from Eq. (2.47) (circles) and computed analytically (solid line) [145], as well as its approximations for $t \to 0$ (Weibull function, long dashes) and $t \to \infty$ (power law, short dashes).

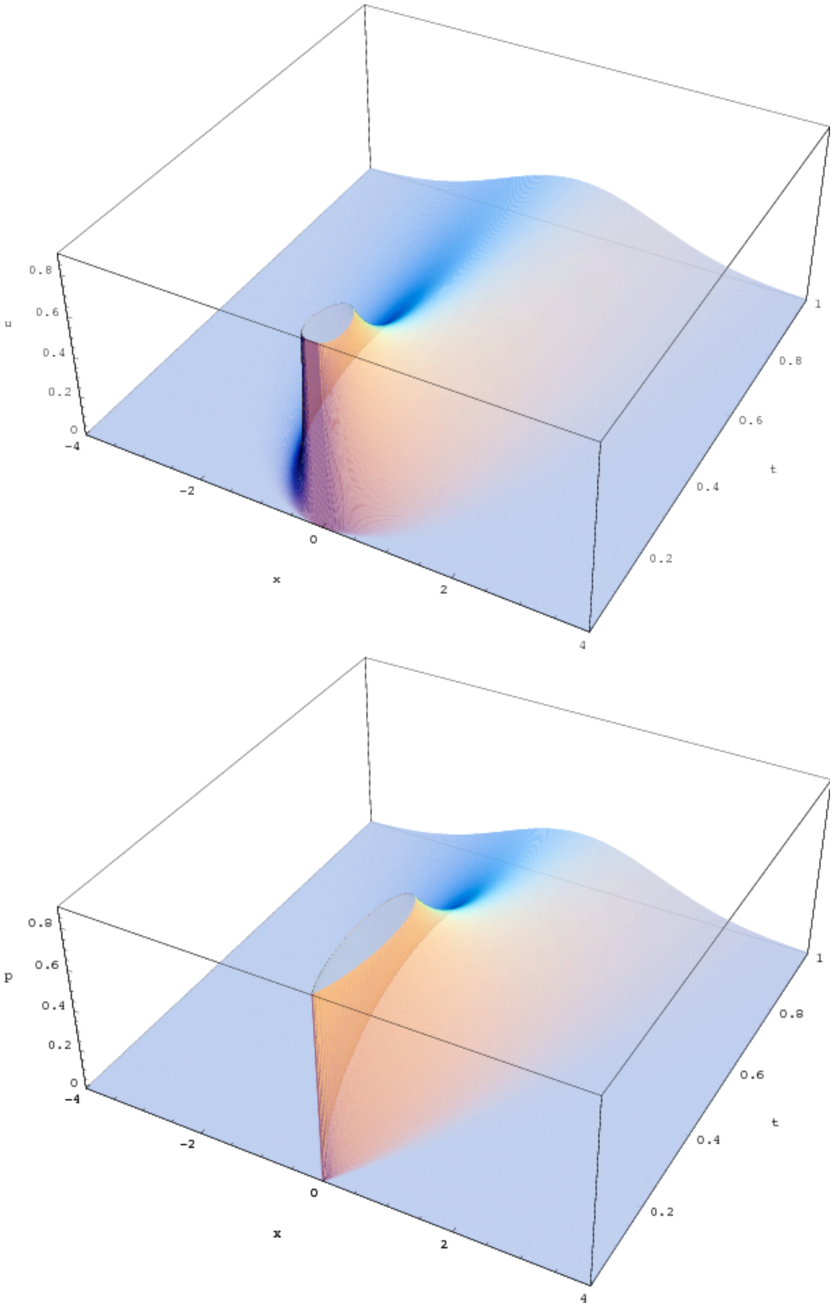


Figure 2.6: Analytic solution of the fractional diffusion equation. Top: $\alpha = 1.7$, $\beta = 0.9$. Bottom: $\alpha = 1.0$, $\beta = 0.9$.

2.6 Transformation formulas for non-uniform random numbers

The usual methods for generating random numbers with a specific probability density are transformation, also called inversion because it requires the inverse cumulative distribution function, [56] and von Neumann rejection [176]. While the latter is more general, the former is usually faster when it is available.

2.6.1 Symmetric Lévy α-stable probability distribution

The symmetric Lévy α-stable probability density $L_\alpha(\xi)$ for the jumps, Eq. (2.44), can be calculated by series expansion, that we do not report here, by direct integration [135, 136] or by numerical Fourier transform [126]. A practical guide to do the latter is given in Appendix A. The exact definition of the general Lévy α-stable probability density is given in Sec. 4.2.2 These methods produce a pointwise representation of the density on a finite interval that can be used for rejection, most efficiently with a look-up table and interpolation. Most convenient is the following transformation method by Chambers, Mallows and Stuck [37]:

$$\xi_\alpha = \gamma_x \left(\frac{-\log u \cos\phi}{\cos((1-\alpha)\phi)} \right)^{1-\frac{1}{\alpha}} \frac{\sin(\alpha\phi)}{\cos\phi}, \tag{2.46}$$

where $\phi = \pi(v - 1/2)$, $u, v \in (0, 1)$ are independent uniform random numbers, γ_x is the scale parameter, and ξ_α is a symmetric Lévy α-stable random number. For $\alpha = 2$ Eq. (2.46) reduces to $\xi_2 = 2\gamma_x\sqrt{-\log u}\sin\phi$, i.e. the Box-Muller method for Gaussian deviates. The other two notable limit cases are the Cauchy distribution, with $\alpha = 1$ and $\xi_1 = \gamma_x \tan\phi$, and the Lévy distribution, with $\alpha = 1/2$ and $\xi_{1/2} = -\gamma_x \tan\phi/(2\log u \cos\phi)$.

2.6.2 One-parameter Mittag-Leffler probability distribution

The probability density $\psi_\beta(\tau)$ for the waiting-times, Eq. (2.45), can be computed as a power series from the definition of the one-parameter Mittag-Leffler function, Eq. (2.37), leading to a pointwise representation on a finite interval; random numbers can then be produced by rejection, again with a look-up table and interpolation. Though CTRW sample paths with a Mittag-Leffler waiting time distribution have appeared in the literature [69, 70, 106, 105], so far it has not been recognized in this context that inversion formulas analogous to Eq. (2.46) are available [46, 139, 84, 87, 85, 86, 79, 63]. The most convenient expression is due to Kozubowski and Rachev [87]:

$$\tau_\beta = -\gamma_t \log u \left(\frac{\sin(\beta\pi)}{\tan(\beta\pi v)} - \cos(\beta\pi) \right)^{\frac{1}{\beta}}, \tag{2.47}$$

where $u, v \in (0, 1)$ are independent uniform random numbers, γ_t is the scale parameter, and τ_β is a Mittag-Leffler random number. For $\beta = 1$, Eq. (2.47) reduces to the inversion formula for the exponential distribution: $\tau_1 = -\gamma_t \log u$. Eq. (2.47) and equivalent forms stem from mixture representations of a Mittag-Leffler random variable through an exponential and a stable random variable. The oldest representation is [46, 79]

$$\tau_\beta = \tau_1^{1/\beta}\,\xi_{\beta,1}, \tag{2.48}$$

where $\xi_{\beta,1}$ is a skew Lévy α-stable random number independent of τ_1, with index $\alpha = \beta$, skewness parameter 1 and scale factor $\gamma_x = 1/8$. A more recent representation is [139, 84]

$$\tau_\beta = \tau_1\,\xi_{1+}^{\pm 1/\beta}, \tag{2.49}$$

where ξ_{1+} is a positive random number distributed according to a Cauchy distribution $L_{1+}(\xi)$ with scale parameter $\gamma_x = \sin(\beta\pi)$, location parameter $\delta = -\cos(\beta\pi)$ and normalization on $\mathbb{R}_+$: $L_{1+}(\xi) = L_1(\xi)/\beta$ for $\xi > 0$.

The connection of Mittag-Leffler to stable random variables can be obtained in the framework of the theory of geometric stable distributions. A random variable ξ is stable if and only if, for all $n \in \mathbb{N}$ iid copies of it, $\xi_1, \ldots, \xi_n$, there exist constants $a_n \in \mathbb{R}_+$ and $b_n \in \mathbb{R}$ such that the scaled and shifted sum $a_n(\xi_1 + \ldots + \xi_n) + b_n$ has the same distribution as ξ. A Mittag-Leffler random variable is not stable, but it is geometric stable [83], i.e. it is the weak limit for $p \to 0$ of the appropriately scaled and shifted geometric random sum $a(p)[\tau_1 + \ldots + \tau_{\nu(p)}] + b(p)$ of suitable iid random variables τ_i, where $\nu(p)$ is a geometric random variable independent of each τ_i, with mean $1/p,\ p \in (0,1)$, and a geometric probability distribution

$$P(\nu(p) = n) = p(1-p)^{n-1}, \quad n \in \mathbb{N}. \tag{2.50}$$

A random variable is geometric stable if and only if its characteristic function $\widehat{\psi}(k)$ is related to the characteristic function $\widehat{\lambda}(k)$ of a stable random variable by the equation [127]

$$\widehat{\psi}(k) = \frac{1}{1 - \log \widehat{\lambda}(k)}. \tag{2.51}$$

With this one-to-one correspondence, a parametrization of a geometric stable probability density $\psi(x)$ can be established from a parametrization of the corresponding stable probability density $\lambda(x)$. Geometric random sums of symmetric τ_i yield the class of Linnik distributions (a generalisation of the Laplace distribution $\frac{1}{2}e^{-|t|}$), while positive τ_i yield the class of Mittag-Leffler distributions (as already seen, a generalisation of the exponential distribution $e^{-t},\ t \geq 0$). In particular, the Mittag-Leffler distribution can be written as a mixture of exponential distributions [68, 86]:

$$E_\beta(-t^\beta) = \int_0^\infty \exp(-\mu t) g(\mu)\, d\mu, \tag{2.52}$$

with a weight

$$g(\mu) = \frac{1}{\pi} \frac{\sin(\beta\pi)}{\mu^{1+\beta} + 2\cos(\beta\pi)\mu + \mu^{1-\beta}} \tag{2.53}$$

given by $g(\mu)d\mu = L_{1+}(\mu^\beta)d\mu^\beta$, where $L_{1+}(\xi)$ is the probability density of ξ_{1+} in Eq. (2.49) introduced before. Equations (2.52–2.53) express Eq. (2.49) in terms of density functions. The inverse cumulative distribution of $L_{1+}(\xi)$ yields the transformation formula for ξ_{1+} appearing as the argument of the power function in Eq. (2.47) [85, 87]. Alternatively, the inversion formula $\xi_1 = \gamma_x \tan\phi + \delta$ for $L_1(\xi)$, see Eq. (2.46), can be substituted into Eq. (2.49), provided negative values of ξ_1 are discarded.

An older equivalent form of Eq. (2.47) was obtained substituting an inversion formula for $\xi_{\beta,1}$ [81] into Eq. (2.48) [46, 79]. A similar result can be reached using a general transformation formula for skew Lévy α-stable random numbers [37], of which Eq. (2.46) is a special case with skewness parameter 0. Both ways require three independent uniform random numbers and more transcendent functions than Eq. (2.47), making the latter slightly more appealing from a numerical point of view.

2.7 Numerical results

Examples of CTRWs generated according to the described procedure, i.e. Equations (2.26), (2.27), (2.46) and (2.47), are shown in Fig. 2.3. The complementary cumulative distribution function (survival function) of random numbers obtained through Eq. (2.47) is checked against its analytic value [145] and its approximations for $t \to 0$ and $t \to \infty$ in Fig. 2.5, where a log-log scale and logarithmic binning [132] is used. Timings are reported in Table 2.1 and Ref. [63].

The advantage of Eq. (2.47) is that Mittag-Leffler deviates are generated with a simple and elegant procedure and no accuracy losses due to truncation of the power series in Eq. (2.37) or truncation of the density function to a finite interval as necessary in the rejection method. The effects of the truncation of the jump density in Lévy flights are analyzed in Ref. [110], whereas no study is available for truncation effects on Mittag-Leffler deviates. Together with Eq. (2.46), a scheme is obtained that yields sample paths for a CTRW with a Lévy jump marginal density and a Mittag-Leffler waiting time marginal density at a speed comparable to that of a NCPP: Though each point for a generic CTRW takes about 3.6 times more than for a NCPP, fewer points are necessary (see $\bar{n}$ in Table 2.1) because the waiting-times are longer. The latter reference reports also that if Lévy and Mittag-Leffler random numbers are produced by rejection, computing the values of the probability density functions simple-mindedly with a series expansion every time they are needed, rather than just once at the beginning to set up a look-up table, for Lévy deviates the procedure takes 400 times longer than with Eq. (2.46), and for Mittag-Leffler deviates it takes 5000 times longer than with Eq. (2.47). Because of the slow convergence of the power series in Eq. (2.37), up to 200 terms are necessary to achieve an acceptable accuracy, and each term is computationally expensive because of the Γ function. Of course these are extreme figures on the

α	β	γ_t	$\bar{n}$	t_{CPU}/sec
2.0	1.0	0.010	200	337
2.0	1.0	0.001	2000	3362
1.7	0.8	0.010	74	437
1.7	0.8	0.001	470	2895

Table 2.1: Average number $\bar{n}$ of jumps per run and total CPU time t_{CPU} in seconds for 10^7 runs with $t \in [0, 2]$ on a 2.2 GHz AMD Athlon 64 X2 Dual-Core with Fedora Core 4 Linux, using the `ran1` uniform random number generator [148] and the Intel C++ compiler version 9.1 with the -O3 -static optimization options.

other end of the efficiency scale meant to show how wide the latter can be; there are smarter ways to compute both the Lévy and Mittag-Leffler [144, 145] probability densities.

Using many CTRW realisations, histograms can be built that give the evolution of $p(x, t)$ with initial condition $p(x, 0) = \delta(x)$, as displayed in Fig. 2.7. According to Eq. (2.28), the initial condition evolves as $\delta(x)\Psi(t)$, i.e. it is visible as a spike at $x = 0$ that decays as t evolves. The mass of the spike is $\Psi(t) = E_\beta\left(-(t/\gamma_t)^\beta\right)$. In Fig. 2.7 this feature appears as a crest. Fig. 2.9 shows how histograms built with CTRWs converge to the Green function, Eq. (2.35), of the FDE for decreasing values of the scale parameters γ_t and $\gamma_x = \gamma_t^{\beta/\alpha}$. To evaluate the scaling function in Eq. (2.36) needed for Eq. (2.35), we used standard FFT methods and algorithms for $E_\beta(-t^\beta)$ [144, 150, 145]. In Fig. 2.8 we plot $\max_{x\neq 0} |p_{\gamma_x,\gamma_t}(x, t;\, \alpha, \beta) - u(x, t;\, \alpha, \beta)|$ as a function of vanishing γ_t with $\gamma_x = \gamma_t^{\beta/\alpha}$. A rigorous analysis of convergence bounds is beyond the scope and resources.

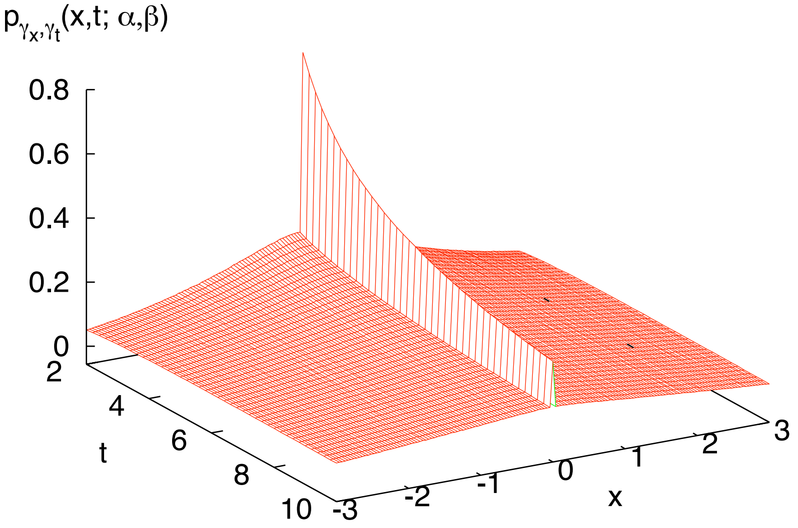


Figure 2.7: Decay of the probability density $p_{\gamma_x,\gamma_t}(x,t;\,\alpha,\beta)$ with $\alpha = 1.7,\ \beta = 0.8,\ \gamma_t = 0.1,\ \gamma_x = \gamma_t^{\beta/\alpha}$. The crest at $x = 0$ is the survival function $\Psi(t) = E_\beta\left(-(t/\gamma_t)^\beta\right) = P(0^+,t) - P(0^-,t)$, where $P(x,t) = \int_{-\infty}^{x} p(u,t)\,du$.

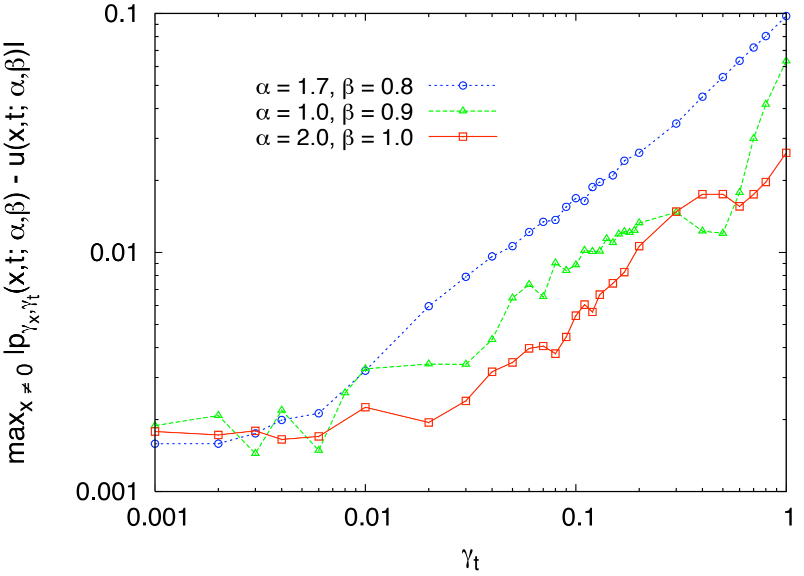


Figure 2.8: Convergence of $\max_{x\neq 0}|p_{\gamma_x,\gamma_t}(x,t;\ \alpha,\beta) - u(x,t;\alpha,\beta)|$ for selected values of α and β when $\gamma_x,\ \gamma_t \to 0$ with $\gamma_x^\alpha = \gamma_t^\beta$.

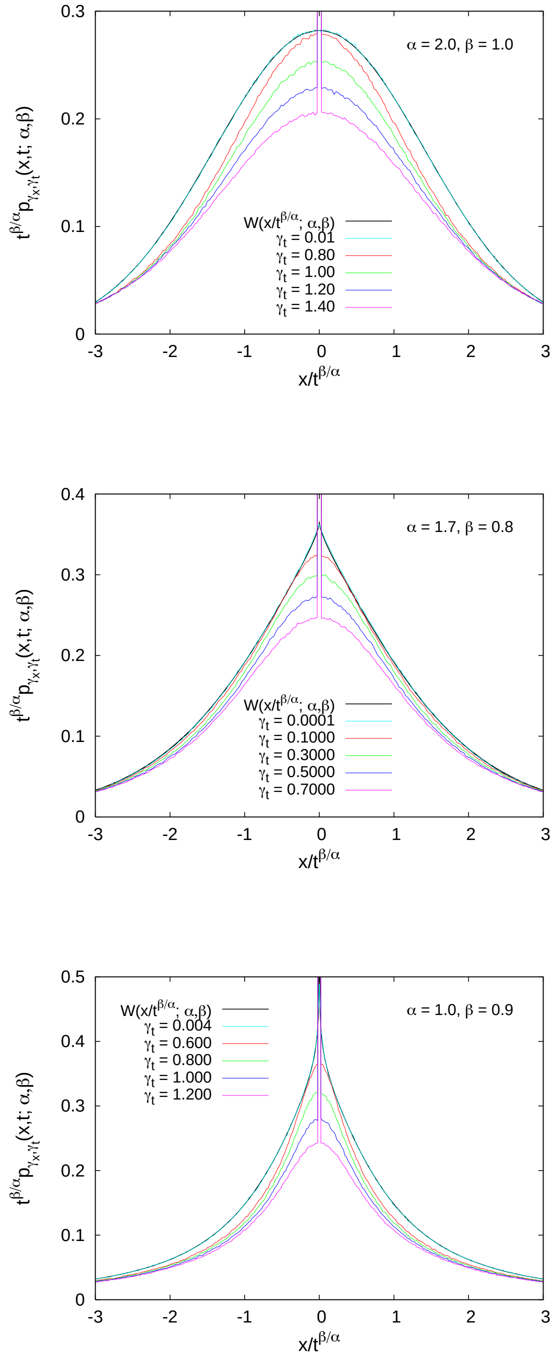


Figure 2.9: Convergence of $t^{\beta/\alpha} p_{\gamma_x,\gamma_t}(x, t; \alpha, \beta)$ to the scaling function $W(x/t^{\beta/\alpha}; \alpha, \beta)$, Eq. (2.36), at $t = 2$ for selected values of α and β. The curves are shown in a time-independent way as scaling plots, and appear in the same order from bottom to top as reported in the legend, i.e. with decreasing γ_t. The curve with the smallest γ_t is almost indistinguishable from its theoretical limit W (solid black line). However, in spite of the impression that may arise from the few terms and the ranges chosen here, in general the function sequences are not monotonic. The scale parameters γ_x and γ_t tend to 0 as $\gamma_x^\alpha = \gamma_t^\beta$. The central peak decreases when the ratio t/γ_t becomes larger, as is evident in Fig. 2.7.

2.8 Conclusions

The use of Mittag-Leffler random numbers generated according to Eq. (2.47) in combination with Lévy random numbers generated according to Eq. (2.46) is very useful in the Monte Carlo simulation of uncoupled continuous-time random walks. In the hydrodynamic limit, appropriately rescaled uncoupled continuous-time random walks with a one-parameter Mittag-Leffler distribution of waiting-times and a symmetric Lévy α-stable distribution of jumps in space yield the Green function of the Cauchy problem for a space-time fractional diffusion equation; we verified this for Eq. (2.31), that has an analytic solution, Eq. (2.35), as a benchmark for more difficult cases where the diffusion and drift terms depend on space and time. We have shown that the computational effort for a fractional diffusion process is almost as small as for a standard diffusion process. It is true that in the same fluid limit the Green function can be obtained too by Monte Carlo sampling of just the asymptotic power-law tail approximations of the Lévy and Mittag-Leffler probability distributions, at least when the indices α and β are not close to 2 and 1 respectively. However, the neat transformation formulas given by Equations (2.46) and (2.47) are numerically so convenient that there is no good reason for resorting to the asymptotic approximations. Moreover we think that, in applications, continuous-time random walks are seen as a more fundamental model than fractional diffusion equations, and sample paths will be generated without taking the scale parameters γ_x and γ_t to the diffusive limit, by using the approach presented here.

Special acknowledgments

We thank Björn Böttcher and René Schilling for help with the literature search, Rudolf Gorenflo and Francesco Mainardi for illuminating discussions, and Tom Kozubowski for useful comments.

Chapter 3

Spectral densities of Wishart-Lévy free stable random matrices

> *'Can you do Addition?' the White Queen said. 'What's one and one and one and one and one and one and one and one and one and one?' 'I don't know', said Alice. 'I lost count'. 'She can't do Addition', the Red Queen interrupted.*
>
> *L.C.*

In correlation analysis the theory of random matrices is used to assess the significance of weak correlations. The theory is well established for Gaussian statistics. However, many complex systems, with stock markets as a prominent example, exhibit statistics with power-law tails, that can be modeled with Lévy stable distributions. We review comprehensively the derivation of an analytic expression for the spectra of covariance matrices approximated by free Lévy stable random variables and validate it by Monte Carlo simulation. It remains to be seen how useful free Lévy stable random variables are in this context since the realistic situation cannot be characterised by a single value of the Lévy parameter α or power-tail of the data increments.

3.1 Introduction

The classical ensembles of random matrices play an important role in the modelling of physical systems, in time series analysis and in other fields. The first notion of a matrix ensemble in statistics was given in the 1920s by Wishart for the purpose of correlation analysis [180]. Physicists began to be interested in random matrices in the 1950s when Wigner presented a model of nuclear energy levels as eigenvalues of symmetric random matrices with Gaussian entries and also with equi-probable ±1 elements [179]. In 1998 Guhr et al. wrote a review on the application of random matrix theory in physics with more than 800 references [71].

Random matrices are used in other fields too, e.g. operations research, for as diverse problems as bandwidth efficiency in wireless communication [59, 129] or optimal aircraft boarding [5, 166]. In correlation analysis the theory of random matrices can be used to assess whether weak correlations are significant or just noise. The mathematical link between correlation matrices of time series and random matrices is the Wishart matrix ensemble, that, together with the Wigner ensemble, is one of the standard tools in the theory of random matrices. The Wigner ensemble also goes under the label Gaussian Orthogonal Ensemble (GOE) that forms together with the Gaussian Hermitian Ensemble (GUE), the Gaussian Symplectic Ensemble (GSE) and the Wishart ensemble the four classical matrix ensembles. A recent introduction and overview including numerical aspects can be found in Ref. [53]. Since the 1990s econophysicists have employed random matrix theory for the analysis of correlation in financial time series [18, 21, 41, 88, 143, 171], with portfolio theory [111, 159] as one of the motivations. Recently, random matrix theory was also used for a correlation analysis of macroeconomic time series [137].

Consider $i = 1, \ldots, N$ stochastic time series x_{ij} observed at synchronous times $t_j,\ j = 0, \ldots, T$. The data can be arranged in a $N \times T$ matrix $\mathbf{M}$ of increments $\Delta x_{ij} = x_{ij} - x_{i,j-1}$, where each row corresponds to a time series and each column to a sampling time. Assuming that the average of the increments is zero, the Pearson estimator for the covariance of two time series i and j is

$$c_{ij} = \frac{1}{T} \sum_{k=1}^{T} \Delta x_{ik} \Delta x_{jk}. \tag{3.1}$$

The covariances of all pairs can be collected in a $N \times N$ symmetric matrix

$$\mathbf{C} = \frac{1}{T} \mathbf{M}\mathbf{M}^{\mathsf{T}}. \tag{3.2}$$

The covariance matrix $\mathbf{C}$ is also called Wishart matrix as it was studied by him. One is often interested in testing the hypothesis that there are no significant correlations. This can be done comparing the eigenvalue spectrum of an empirical correlation matrix with the spectrum of a reference matrix built with synthetic uncorrelated time series. If the matrix rows are random walks whose increments are independent and identically distributed (iid) normal deviates with standard deviation σ, the spectrum describing the above null hypothesis in the limit for $N, T \to \infty$ with $m = N/T$ is given analytically by the Marčenko-Pastur law [118]:

$$\begin{aligned} \rho_{\mathbf{C}}(\lambda) &= \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi m \sigma^2 \lambda}, \\ \lambda_\pm &= \sigma^2 \left(1 \pm \sqrt{m}\right)^2. \end{aligned} \tag{3.3}$$

This result has been rediscovered a few times [6, 52, 53]. Indeed, for a sufficiently large matrix the exact distribution of m_{it} becomes less and less relevant, and the Marčenko-Pastur law can be obtained for iid increments drawn from any distribution with a finite second moment. This effect was also evident in Wigner's studies of matrices whose entries are binary random variables assuming the values ± 1

with equal probability. In both the Wigner and Wishart ensembles the spectra of large matrices converge to that of an infinite matrix (respectively the semi-circle law and the Marčenko-Pastur law) as a consequence of a generalised central limit theorem.

A practical use of Eq. (3.3) is that if the empirical spectrum of data shows significant differences from the theoretical curve, then it may be justified to reject the null hypothesis of no true correlations. The details of the latter are then a separate issue. In principle it is possible to test not only the absence of true correlations, but also any kind of suitable assumption leading to a given shape of the expected spectrum, both theoretically or numerically. Depending on the specific case one chooses a suitable null hypothesis. For example, if the considered time series are the log-prices of traded stocks, in a first approximation it is reasonable to test the absence of true correlation with normally distributed log-returns [88, 143, 39]. Another powerful approach requiring less knowledge of the distribution of the increments is a bootstrap scheme that consists in re-sampling the covariance matrix after random permutations of the empirical time series. Since the reshuffling of the rows of **M** destroys any possible correlation, an absence of correlation among the original time series requires that the eigenvalue spectrum of **C** does not change [31].

So far, the result given by Eq. (3.3) lies within classical random matrix theory and requires iid matrix elements with finite moments. In this work we are concerned with the Wishart-Lévy ensemble as a natural extension of the Wishart-Gaussian ensemble treated by the Marčenko-Pastur theory. The situation becomes more complicated if the elements of **M** are distributed with power-law tails, as happens in numerous physical, biological and economic data [39]. Stock markets as well as many other complex systems exhibit a dynamics that results in power-law tailed statistics. The Marčenko-Pastur theory is not valid any more when the second moment is not finite, and the corresponding spectral densities cannot be obtained from a simple extension of Gaussian random matrix theory. As a consequence of the central limit theorem for scale-free processes the distribution of many of the above phenomena is usually assumed to be a symmetric Lévy α-stable distribution, whose pdf is given most suitably as the inverse Fourier (cosine) transform of its characteristic function:

$$L_\alpha(x) = \mathcal{F}_k^{-1}\left[e^{-\gamma|k|^\alpha}\right](x) = \frac{1}{\pi}\int_0^\infty e^{-(\gamma k)^\alpha}\cos(xk)\,\mathrm{d}k. \tag{3.4}$$

The second and higher moments of $L_\alpha(x)$ diverge for $\alpha < 2$, and for $\alpha \leq 1$ even the first moment does not exist any more. If $\alpha = 2$ Eq. (3.4) gives a Gaussian. However, we shall see that the functional representation of this distribution is not required in the derivation of the spectrum.

A matrix whose elements are iid samples from a stable density is called a Lévy matrix. A square and symmetric Lévy matrix is called a Wigner-Lévy matrix. Sampling the elements from the probability density function

$$f_X(x) = N^{2/\alpha} L_\alpha(N^{2/\alpha} x), \tag{3.5}$$

the limiting spectrum becomes independent of the matrix size N [38]. It turns out that the spectrum has no longer a finite support as in the Marčenko-Pastur case and is dominated by the behaviour of its power-law tail.

It was proposed to use the theory of free probability with its convenient machinery leading to analytic results that could be obtained otherwise only by means of a painful use of combinatorics. A free Lévy stable random matrix has a spectrum belonging to the class of free stable laws. The contemporary physical and mathematical literature on correlation matrix analysis with power-law tailed uncorrelated noise is very active also in the context of free probability. Limiting the list to physics journals, the reader can consult Refs. [15, 16, 20, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 172]. For a review of free probability theory see Ref. [134]. The Marčenko-Pastur spectrum can be obtained as a special case of this more general theory.

Our aim in this chapter is to review comprehensively the analytic derivation of the spectral density of free stable Wishart-Lévy random matrices already solved by Burda et al. [24, 25, 26, 27, 28, 29, 30, 31, 32, 33] and, as a further step, to validate numerically the analytic result by Monte Carlo simulation. The rest of this chapter is organised as follows. Sec. 3.2 introduces the mathematical background of free probability theory, whose objects are elements of an algebra, usually an operator algebra, and may enjoy the property of freeness. Sec. 3.3 explains free stability and presents an approximation for the Wishart-Lèvy covariance matrix of time series using free stable random variables. An explanation of free stability is provided too. Sec. 3.4 derives in detail a transcendental equation, due to Burda et al., whose solution gives the spectral density for the approximated covariance matrix. Sec. 3.5 shows numerically the validity of this equation comparing analytic and Monte Carlo results. A summary and an Appendix with computer code conclude this material.

3.2 Mathematical background

A symmetric $N \times N$ matrix $\mathbf{X}$ has real eigenvalues $\lambda_1, \ldots, \lambda_N$. The spectral density of $\mathbf{X}$ can be written as

$$\rho_{\mathbf{X}}(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i), \tag{3.6}$$

where it is assumed that the weight of each eigenvalue is the same and each eigenvalue is counted as many times as its multiplicity. The resolvent matrix [42] is defined as

$$\mathbf{G}_{\mathbf{X}}(z) = (z\mathbf{1} - \mathbf{X})^{-1}, \quad z \in \mathbb{C}, \tag{3.7}$$

where $\mathbf{1}$ is the $N \times N$ identity matrix. The Green function is defined as

$$G_{\mathbf{X}}(z) = \frac{1}{N} \operatorname{tr} \mathbf{G}_{\mathbf{X}}(z), \tag{3.8}$$

where the trace tr of a square matrix is defined as the sum of its diagonal elements. If $\mathbf{X}$ is a random matrix, the above definition is generalised including an expectation operator $\mathbb{E}$:

$$G_{\mathbf{X}}(z) = \frac{1}{N}\,\mathbb{E}[\operatorname{tr}\mathbf{G}_{\mathbf{X}}(z)]. \tag{3.9}$$

The Green function contains the same information as the eigenvalues and the eigenvalue density of $\mathbf{X}$ [17]. The Green function can be written in terms of the eigenvalues of $\mathbf{X}$:

$$G_{\mathbf{X}}(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_i}. \tag{3.10}$$

This is a special case of the definition through the Cauchy transform of a generic spectral density:

$$G_{\mathbf{X}}(z) = \int_{-\infty}^{+\infty}\frac{1}{z-\lambda}\rho_{\mathbf{X}}(\lambda)\,\mathrm{d}\lambda. \tag{3.11}$$

By using the following representation of Dirac's δ-function,

$$\frac{1}{x \pm i\epsilon} = \mathrm{PV}\left(\frac{1}{x}\right) \mp i\pi\delta(x), \tag{3.12}$$

where PV denotes the principal value, the spectral density can be obtained from the Green function:

$$\rho_{\mathbf{A}}(\lambda) = \lim_{\epsilon\to 0^+}\frac{1}{\pi}\mathrm{Im}[G_{\mathbf{X}}(\lambda - i\epsilon)]. \tag{3.13}$$

This means that the eigenvalues follow from the discontinuities of $G_{\mathbf{X}}(z)$ on the real axis.

Non-commutativity of matrices and, in general, of operators makes it difficult to extend standard probability theory to matrix as well as operators spaces. Among the possible extensions of probability theory to operator spaces the so-called free probability theory has the advantage that many results can be deduced from well-known theorems on analytic functions [25].

In order to explain the framework of free probability, let us start from conventional classical probability. A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a measure space, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathrm{P} : \mathcal{F} \to [0,1] \in \mathbb{R}$ is a non-negative measure on sets in $\mathcal{F}$ obeying Kolmogorov's axioms; $\omega \in \Omega$ is called an elementary event, $A \in \mathcal{F}$ is called an event. A random variable $X : \Omega \to \mathbb{R}$ is a measurable function that maps elements from the sample space to the real numbers, and thus elements from $\mathcal{F}$ to a Borel σ-algebra Σ on $\mathbb{R}$. The probability distribution of X with respect to P is described by a measure μ_X on $(\mathbb{R}, \Sigma)$ defined as the image measure of P: $\mu_X(B) = \mathrm{P}[X^{-1}(B)]$, where B is any Borel set and $X^{-1}(B) \subset \mathcal{F}$ is the counter-image of B. The cumulative distribution function of X is $F_X(x) = \mu_X(X \le x)$. The expectation value for any bounded Borel function

$g : \mathbb{R} \to \mathbb{R}$ is

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\mu_X(\mathrm{d}x) = \int_{\mathbb{R}} g(x)\,\mathrm{d}F_X(x). \tag{3.14}$$

If $F_X(s)$ is differentiable, the probability density function (pdf) of X is $f_X(x) = \mathrm{d}F_X(x)/\mathrm{d}x$.

This construction can be extended to non-commutative variables, e.g. matrices or more in general operators. Let $\mathcal{A}$ denote a unital algebra over a field $\mathbb{F}$, i.e. a vector space equipped with a bilinear product $\circ : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ that has an identity element $\mathbf{I}$. A tracial state on $\mathcal{A}$ is a positive linear function $\tau : \mathcal{A} \to \mathbb{F}$ with the properties $\tau(\mathbf{I}) = 1$ and $\tau(\mathbf{XY}) = \tau(\mathbf{YX})$ for every $\mathbf{X}, \mathbf{Y} \in \mathcal{A}$. The couple $(\mathcal{A}, \tau)$ is called a non-commutative probability space.

For our purposes $\mathcal{A} = \mathcal{B}(\mathcal{H})$, where $\mathcal{B}(\mathcal{H})$ denotes the Banach algebra of linear operators on a real separable Hilbert space $\mathcal{H}$. This is a $*$-algebra, as it is equipped with an involution (the adjoint operation) $\mathbf{X} \mapsto \mathbf{X}^* : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$. Considering a self-adjoint operator $\mathbf{X} \in \mathcal{B}(\mathcal{H})$, it is possible to associate a (spectral) distribution to $\mathbf{X}$ as in classical probability. Thanks to the Riesz representation theorem and the Stone-Weierstrass theorem, there is a unique measure $\mu_{\mathbf{X}}$ on $(\mathbb{R}, \Sigma)$ satisfying

$$\int_{\mathbb{R}} g(x)\mu_{\mathbf{X}}(\mathrm{d}x) = \tau(g(\mathbf{X})) \tag{3.15}$$

where $g : \mathbb{R} \to \mathbb{R}$ is any bounded Borel function [134]. Therefore we say that the distribution of $\mathbf{X}$ is described by the measure $\mu_{\mathbf{X}}$. For our purposes this measure is equal to the spectral density $\rho_{\mathbf{X}}$ defined in Eq. (3.13).

Classically, independence between two random variables X and Y can be defined requiring that for any couple of bounded Borel functions f, g

$$\mathbb{E}[(f(X) - \mathbb{E}[f(X)])(g(Y) - \mathbb{E}[g(Y)])] = 0. \tag{3.16}$$

Analogously, two elements $\mathbf{X}$ and $\mathbf{Y}$ in a non-commutative probability space are defined as free (of freely) independent with respect to τ, if for any couple of bounded Borel functions f, g

$$\tau[(f(\mathbf{X}) - \tau[f(\mathbf{X})])(g(\mathbf{Y}) - \tau[g(\mathbf{Y})])] = 0. \tag{3.17}$$

Defining freeness between more than two elements is a non-trivial extension [9].

Generally, square $N \times N$ random matrices $\mathbf{X}$ are non-commutative variables with respect to the function $\tau(\mathbf{X}) = (1/N)\,\mathbb{E}[\operatorname{tr}\mathbf{X}]$, see Eq. (3.9), but for any given N no pair of random matrices is free. Nevertheless two random matrices $\mathbf{X}, \mathbf{Y}$ can reach freeness asymptotically if for any integer $n > 0$ and any set of non-negative integers $(\gamma_1, \ldots, \gamma_n)$ and $(\beta_1, \ldots, \beta_n)$ for which in the limit $N \to \infty$

$$\tau(\mathbf{X}^{\gamma_1}) = \ldots = \tau(\mathbf{X}^{\gamma_n}) = \tau(\mathbf{Y}^{\beta_1}) = \ldots = \tau(\mathbf{Y}^{\beta_n}) = 0 \tag{3.18}$$

we have

$$\tau(\mathbf{X}^{\gamma_1}\,\mathbf{Y}^{\beta_1}\,\dots\,\mathbf{X}^{\gamma_n}\,\mathbf{Y}^{\beta_n}) = 0. \tag{3.19}$$

This means that large random matrices can be good approximations of free non-commutative variables.

Given an operator $\mathbf{X} \in \mathcal{B}(\mathcal{H})$, the following functions are useful in deriving its spectral distribution $\mu_{\mathbf{X}}$:

1. *Moment generating function,* defined as

   $$w_{\mathbf{X}}(z) = zG_{\mathbf{X}}(z) - 1. \tag{3.20}$$

   The name stems from the fact that, if the distribution of $\mathbf{X}$ has finite moments $m_{\mathbf{X},k} = \tau[\mathbf{X}^k]$,

   $$w_{\mathbf{X}}(z) = \sum_{k=1}^{\infty} \frac{m_{\mathbf{X},k}}{z^k}. \tag{3.21}$$

2. *R-transform.* In classical probability the pdf of the sum of two independent random variables $X + Y$ is equal to the convolution of the individual pdfs, i.e.

   $$f_{X+Y}(x) = (f_X * f_Y)(x). \tag{3.22}$$

   The convolution is done conveniently in Fourier space, where it becomes a multiplication: the characteristic function

   $$\hat{f}_{X+Y}(k) = \int_{\mathbb{R}} f_{X+Y}(x) e^{ikx}\, \mathrm{d}x \tag{3.23}$$

   of $X + Y$ is the product of the characteristic functions of X and Y,

   $$\hat{f}_{X+Y}(k) = \hat{f}_X(k)\hat{f}_Y(k), \tag{3.24}$$

   and the cumulant generating function of $X + Y$ is the sum of the cumulant generating functions of X and Y:

   $$\log \hat{f}_{X+Y}(k) = \log \hat{f}_X(k) + \log \hat{f}_Y(k). \tag{3.25}$$

   The free analogue of the cumulant generating function is the R-transform invented by Voiculescu [14, 134, 173] as part of the functional inverse of the Green function:

   $$G_{\mathbf{X}}\left(R_{\mathbf{X}}(z) + \frac{1}{z}\right) = z. \tag{3.26}$$

   The R-transform for the sum of two free operators is the sum of their R-

transforms:

$$R_{\mathbf{X}+\mathbf{Y}}(z) = R_{\mathbf{X}}(z) + R_{\mathbf{Y}}(z). \tag{3.27}$$

The free analogue of convolution is indicated with the symbol $\boxplus$:

$$\mu_{\mathbf{X}+\mathbf{Y}} = \mu_{\mathbf{X}} \boxplus \mu_{\mathbf{Y}}. \tag{3.28}$$

This is computed through $R_{\mathbf{X}}$, given the connection between the Green function $G_{\mathbf{X}}$ and the spectral distribution $\mu_{\mathbf{X}}$. Other definitions of the R-transform were proposed later.

3. *Blue function.* It is convenient to introduce also an inverse of the Green function $G_{\mathbf{X}}(z)$, called Blue function as a pun [78]:

$$G_{\mathbf{X}}(B_{\mathbf{X}}(z)) = B_{\mathbf{X}}(G_{\mathbf{X}}(z)) = z. \tag{3.29}$$

The Blue function is related to the R-transform by

$$B_{\mathbf{X}}(z) = R_{\mathbf{X}}(z) + \frac{1}{z}. \tag{3.30}$$

4. *S-transform.* In the same fashion as the R-transform for the sum, another transform allows to compute the spectral distribution of the product of two operators from their individual spectral distributions:

$$S_{\mathbf{X}}(z) = \frac{1+z}{z}\chi_{\mathbf{X}}(z), \tag{3.31}$$

where $\chi_{\mathbf{X}}(z)$ is defined through

$$\chi_{\mathbf{X}}(zG_{\mathbf{X}}(z) - 1) = \frac{1}{z}. \tag{3.32}$$

For $\mathbf{X} \neq \mathbf{Y}$ the S-transform of the product is the product of the individual S-transforms:

$$S_{\mathbf{XY}}(z) = S_{\mathbf{X}}(z)S_{\mathbf{Y}}(z). \tag{3.33}$$

3.3 Free stable random variables and the Wishart-Lévy ensemble

Let $\mathbf{P}$ be the matrix projector of size $T \times T$, with N ones in arbitrary positions on the diagonal and all the other elements zero, e.g.:

$$\mathbf{P} = \mathrm{diag}(\dots, 1, 1, \dots, 0, 1, 0, 0, 1, \dots, 1, 0, \dots). \tag{3.34}$$

Let $\mathbf{\Lambda}$ be a (large) $T \times T$ matrix with a free stable spectral distribution. This property is the analogue of classical stability. The sum of two free non-commutative

μ-distributed variables results in a new μ-distributed variable. The Wishart matrix ensemble of size $N \times N$ defined in Eq. (3.2) can be approximated using the $N \times T$ matrix $\mathbf{M}$ obtained from $\mathbf{P\Lambda}$ if only the non-zero entries are considered [26, 27, 28, 29, 30, 31, 32, 33]. Indicating this operation with curly braces, the approximation reads

$$\mathbf{C} = \frac{1}{T^{2/\alpha}}\mathbf{M}\mathbf{M}^{\mathsf{T}} \simeq \{\mathbf{P\Lambda}\}\{\mathbf{\Lambda}^{\mathsf{T}}\mathbf{P}\}. \tag{3.35}$$

Notice that the normalisation factor has been generalised with respect to Eq. (3.2) to take into account Lévy α-stable statistics. The former equation is justified by very good results in a similar approach for Wigner-Lévy matrices [33]. In what follows, $\mathbf{W}$ will denote a generic Wigner-Lévy matrix.

Once we know the domain of attraction for one specific classical stable distribution, we can expect that a sum of iid random numbers, e.g. $Z = (1/\mathcal{N}_n)\sum_{i=1}^{n} Z_i$ with some suitable normalisation $\mathcal{N}_n$, converges to their attractor for large n. If Z_i are independent elements of random matrices, as in Ref. [88], each of them tends to a stable law under matrix addition. However, for free stability we must consider random matrices as a whole, and a different procedure is needed. A fundamental point is the property discussed extensively by Bercovici and Pata [13], that can be summarized as follows. If $\mathcal{D}_{\mathrm{c}}(\mu_{\mathrm{c}})$ and $\mathcal{D}_{\mathrm{f}}(\mu_{\mathrm{f}})$ are the domains of attraction of the stable laws μ_{c} and μ_{f} in classical and free probability respectively, a distribution $\nu \in \mathcal{D}_{\mathrm{c}}(\mu_{\mathrm{c}}) \Leftrightarrow \nu \in \mathcal{D}_{\mathrm{f}}(\mu_{\mathrm{f}})$. In other words, if we are able to recognise the classical attractor $\mathcal{D}_{\mathrm{c}}$ of a distribution ν, we also know its free attractor $\mathcal{D}_{\mathrm{f}}$. Moreover, one and only one free stable distribution corresponds to any set of parameter values characterising a classically stable distribution. The spectrum of $\mathbf{W}$ is symmetric with the same tail index α of its entries, i.e. it belongs to the domain of attraction of a well-recognised classical stable law. This means that the sum of sufficiently many free non-commutative variables with this spectrum converges to a non-commutative variable with a stable distribution.

Another property largely discussed in Refs. [141, 134, 165] can be summarised for our purpose as follows. Considering two $N \times N$ matrices $\mathbf{W}_i$ and $\mathbf{W}_j$ with $i \neq j$ and two independent random orthogonal $N \times N$ matrices $\mathbf{O}_i$ and $\mathbf{O}_j$, the matrices $\mathbf{O}_i\mathbf{W}_i\mathbf{O}_i^{\mathsf{T}}$ and $\mathbf{O}_j\mathbf{W}_j\mathbf{O}_j^{\mathsf{T}}$ are free in the limit $N \to \infty$. These properties together with the observation that $\mathbf{W}_i$ and $\mathbf{O}_i\mathbf{W}_i\mathbf{O}_i^{\mathsf{T}}$ have the same spectrum justify the equation

$$\mathbf{\Lambda} \simeq \frac{1}{(TR)^{1/\alpha}}\sum_{i=1}^{R}\mathbf{O}_i\mathbf{W}_i\mathbf{O}_i^{\mathsf{T}}. \tag{3.36}$$

This means that a free stable non-commutative variable can be approximated by adding randomly rotated classical Lévy random matrices.

To generate Lévy matrices we use the Chambers-Mallows-Stuck algorithm [37, 119]: a random number X drawn from the symmetric Lévy α-stable pdf, Eq. (3.4), can be obtained from two independent uniform random numbers $U, V \in (0, 1)$

through the transformation

$$X = \gamma \left(\frac{-\log U \cos \Phi}{\cos((1-\alpha)\Phi)} \right)^{1-\frac{1}{\alpha}} \frac{\sin(\alpha\Phi)}{\cos\Phi}, \tag{3.37}$$

where $\Phi = \pi(V - 1/2)$. For $\alpha = 2$ Eq. (3.37) reduces to $X = 2\gamma_x\sqrt{-\log U}\sin\Phi$, i.e. the Box-Muller method for Gaussian deviates.

The QR-decomposition of a $T \times T$ matrix $\mathbf{H}$ with random Gaussian entries allows to write

$$\mathbf{H} = \mathbf{O}\,\mathbf{U}, \tag{3.38}$$

where $\mathbf{O}$ is random orthogonal and $\mathbf{U}$ is upper (or right) triangular. For alternative methods to obtain a random orthogonal matrix see Ref. [47] and references therein.

3.4 The analytic spectrum

The moment generating function of the $T \times T$ matrix $\mathbf{D} = \mathbf{\Lambda}\mathbf{P}\mathbf{\Lambda}^{\mathsf{T}}$ satisfies the transcendental equation [24, 25, 26, 30]

$$-\exp\left(i\frac{2\pi}{\alpha}\right) w_{\mathbf{D}}(z)^{2/\alpha}\, z = (w_{\mathbf{D}}(z)+1)(w_{\mathbf{D}}(z)+m), \tag{3.39}$$

which can be solved analytically for a few special values of $\alpha = 1/4, 1/3, 1/2, 2/3, 3/4, 1, 4/3, 3/2, 2$; the solution was published for $\alpha = 1$ [26]. The equation can be solved numerically for other values, see Sec. 3.7. Actually, we are interested in the spectrum of the approximation of $\mathbf{C}$ provided by the rhs of Eq. (3.35), but the Green functions of the matrices $\mathbf{D}$ and $\mathbf{C}$ are related by the equation [25]

$$G_{\mathbf{D}}(z) = m\, G_{\mathbf{C}}(z) + \frac{1-m}{z}, \tag{3.40}$$

whence

$$w_{\mathbf{D}}(z) = z\, G_{\mathbf{D}}(z) - 1 = m\, z\, G_{\mathbf{C}}(z) - m = m\, w_{\mathbf{C}}(z). \tag{3.41}$$

In the following we will explain in detail the route that leads to Eq. (3.39) and then to the desired spectral density $\rho_{\mathbf{C}}(\lambda)$.

As in classical probability stable laws have an analytic form for their Fourier transform, free stable laws have an analytic form for their Blue transform [9, 13, 33, 174]:

$$B_{\mathbf{\Lambda}}(z;\alpha) = a + bz^{\alpha-1} + \frac{1}{z}. \tag{3.42}$$

The parameter a accounts for a horizontal shift in the distribution of the matrix elements and can be set to zero without loss of generality. The parameter b depends on the distribution; for the symmetric Lévy α-stable pdf, Eq. (3.4), it has the value [24]

$$b = e^{i\pi(\alpha/2-1)}. \tag{3.43}$$

As discussed in the previous subsection, given an index $\alpha \in (0, 2]$, $B_{\mathbf{\Lambda}}(z; \alpha)$ indirectly but precisely defines the attractor law for the sum of free variables with α-tailed spectral distribution. Since free probability theory is exact only in the large size limit $T, N \to \infty,\ N/T = m$, the only variables that define the model are α and m.

Rewriting Eq. (3.42) with $G_{\mathbf{\Lambda}}(z)$ in place of z and using Eq. (3.29) yields

$$b\, G_{\mathbf{\Lambda}}^{\alpha-1}(z) + G_{\mathbf{\Lambda}}^{-1}(z) = z, \tag{3.44}$$

which is equivalent to

$$b\, G_{\mathbf{\Lambda}}^{\alpha}(z) + zG_{\mathbf{\Lambda}}(z) + 1 = 0, \quad G_{\mathbf{\Lambda}}(z) \neq 0. \tag{3.45}$$

In Sec. 3.2 we established calculation rules with the help of which the solution of our specific problem can be put together piece by piece. First notice that thanks to Eq. (3.33), if for simplicity $\mathbf{\Lambda} = \mathbf{\Lambda}^{\mathsf{T}}$,

$$S_{\mathbf{\Lambda P \Lambda}} = S_{\mathbf{\Lambda}} S_{\mathbf{P \Lambda}} = S_{\mathbf{\Lambda}} S_{\mathbf{\Lambda P}} = S_{\mathbf{\Lambda \Lambda P}} = S_{\mathbf{\Lambda}^2\mathbf{P}}. \tag{3.46}$$

For the S-transform of the matrix product $\mathbf{\Lambda}^2$ we also require the resolvent. The desired relation is a consequence of the fact that the spectral measure for free Lévy α-stable operators in the Wigner ensemble is symmetric [78]:

$$\rho_{\mathbf{\Lambda}}(\lambda) = \rho_{\mathbf{\Lambda}}(-\lambda) \tag{3.47}$$
$$G_{\mathbf{\Lambda}}(z) = G_{-\mathbf{\Lambda}}(z). \tag{3.48}$$

Via the Cauchy transform representation of the resolvent and exploiting the previous symmetry we can express the Green function of $\mathbf{\Lambda}^2$ in terms of the Green function of $\mathbf{\Lambda}$:

$$\begin{aligned} G_{\mathbf{\Lambda}^2}(z) &= \int_{-\infty}^{+\infty} \frac{1}{z - \lambda^2} \rho(\lambda)\, \mathrm{d}\lambda \\ &= \int_{-\infty}^{+\infty} \left[\frac{1}{2\sqrt{z}} \left(\frac{1}{\sqrt{z} - \lambda} + \frac{1}{\sqrt{z} + \lambda}\right)\right] \rho(\lambda)\, \mathrm{d}\lambda \\ &= \frac{1}{2\sqrt{z}} \left(G_{\mathbf{\Lambda}}(\sqrt{z}) + G_{-\mathbf{\Lambda}}(\sqrt{z})\right) \\ &= \frac{1}{\sqrt{z}} G_{\mathbf{\Lambda}}(\sqrt{z}). \end{aligned} \tag{3.49}$$

The next piece in the composition of the solution is the S-transform of the projector $\mathbf{P}$, which requires its Green function too. Inserting the spectral density of $\mathbf{P}$,

$$\rho_{\mathbf{P}}(\lambda) = m\delta(\lambda - 1) + (1 - m)\delta(\lambda), \tag{3.50}$$

into the definition of the Green function of $\mathbf{P}$ as a Cauchy transform yields

$$\begin{aligned} G_{\mathbf{P}}(z) &= \int \frac{1}{z-\lambda}\rho_{\mathbf{P}}(\lambda)\,\mathrm{d}\lambda \\ &= \int \frac{1}{z-\lambda}[m\delta(\lambda-1)+(1-m)\delta(\lambda)]\,\mathrm{d}\lambda \\ &= \frac{m}{z-1}+\frac{1-m}{z}. \end{aligned} \tag{3.51}$$

The moment generating function $w_{\mathbf{P}}(z) = zG_{\mathbf{P}}(z) - 1$ and the definition of the S-transform finally give

$$S_{\mathbf{P}}(z) = \frac{z+1}{z+m}. \tag{3.52}$$

Rewriting Eq. (3.45) with $\sqrt{z}$ in place of z,

$$b\,G^{\alpha}_{\mathbf{\Lambda}}(\sqrt{z}) - \sqrt{z}G^{2}_{\mathbf{\Lambda}}(\sqrt{z}) + 1 = 0, \tag{3.53}$$

and inserting Eq. (3.49) yields

$$b\,z^{\alpha/2}G^{\alpha}_{\mathbf{\Lambda}^2}(z) - zG_{\mathbf{\Lambda}^2}(z) + 1 = 0. \tag{3.54}$$

Observing that from Eq. (3.32)

$$z = \frac{1}{\chi_{\mathbf{\Lambda}^2}(zG_{\mathbf{\Lambda}^2}(z)-1)} \equiv \frac{1}{\chi_{\mathbf{\Lambda}^2}}, \tag{3.55}$$

Eq. (3.54) becomes

$$b\,\chi^{-\alpha/2}G^{\alpha}_{\mathbf{\Lambda}^2}\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}}\right) - \frac{1}{\chi_{\mathbf{\Lambda}^2}}G_{\mathbf{\Lambda}^2}\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}}\right) + 1 = 0. \tag{3.56}$$

Because from Eq. (3.31) it follows that

$$\frac{1}{\chi_{\mathbf{\Lambda}^2}}G_{\mathbf{\Lambda}^2}\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}}\right) - 1 = z, \tag{3.57}$$

Eq. (3.56) can be simplified to

$$b\chi^{-\alpha/2}_{\mathbf{\Lambda}^2}G^{\alpha}_{\mathbf{\Lambda}^2}\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}}\right) = z. \tag{3.58}$$

Multiplying both sides by $\chi^{-\alpha/2}_{\mathbf{\Lambda}^2}/b$ yields

$$\chi^{-\alpha}_{\mathbf{\Lambda}^2}G^{\alpha}_{\mathbf{\Lambda}^2}\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}}\right) = \frac{z}{b}\chi^{-\alpha/2}_{\mathbf{\Lambda}^2}; \tag{3.59}$$

then subtracting and adding 1,

$$\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}} G_{\mathbf{\Lambda}^2}\left(\frac{1}{\chi_{\mathbf{\Lambda}^2}}\right) - 1 + 1\right)^{\alpha} = \frac{z}{b}\chi_{\mathbf{\Lambda}^2}^{-\alpha/2}, \tag{3.60}$$

and inserting again Eq. (3.57) gives

$$(z+1)^{\alpha} = \frac{z}{b}\chi_{\mathbf{\Lambda}^2}^{-\alpha/2}, \tag{3.61}$$

which can be written as

$$\chi_{\mathbf{\Lambda}^2} = \frac{1}{(z+1)^2}\left(\frac{z}{b}\right)^{2/\alpha}. \tag{3.62}$$

Now, using the definition of the S-transform and the result

$$S_{\mathbf{\Lambda}^2} = \frac{1+z}{z}\chi_{\mathbf{\Lambda}^2} = \frac{1}{z(1+z)}\left(\frac{z}{b}\right)^{2/\alpha}, \tag{3.63}$$

which can be used to write $S_{\mathbf{D}}$, the S-transform of the Wishart matrix on the rhs of Eq. (3.35) is

$$S_{\mathbf{P}\mathbf{\Lambda}^2} = S_{\mathbf{P}} S_{\mathbf{\Lambda}^2} = \frac{1}{z(m+z)}\left(\frac{z}{b}\right)^{2/\alpha}. \tag{3.64}$$

This result is the starting point for the way back. Re-applying the definition of the S-transform we can write

$$\chi_{\mathbf{\Lambda}^2\mathbf{P}} = \frac{z}{z+1} S_{\mathbf{\Lambda}^2\mathbf{P}} = \frac{1}{(z+1)(z+m)}\left(\frac{z}{b}\right)^{2/\alpha} \tag{3.65}$$

and

$$\chi_{\mathbf{\Lambda}^2\mathbf{P}}^{-1} = (z+1)(z+m)\left(\frac{z}{b}\right)^{-2/\alpha}. \tag{3.66}$$

Together with $w_{\mathbf{D}}(z) = z\,G_{\mathbf{D}}(z) - 1$ this allows to substitute $\chi_{\mathbf{D}}(w_{\mathbf{D}}(z)) = 1/z$ and $w_{\mathbf{D}}(1/\chi_{\mathbf{D}}(z)) = z$. Notice that we changed the index $\mathbf{\Lambda}^2\mathbf{P}$ to $\mathbf{D}$ to emphasise our goal. So we can finally write

$$z = (w_{\mathbf{D}}(z)+1)(w_{\mathbf{D}}(z)+m)\left(\frac{w_{\mathbf{D}}(z)}{b}\right)^{-2/\alpha}. \tag{3.67}$$

Inserting Eq. (3.41) yields the corresponding equation for $\mathbf{C}$:

$$z = (m\,w_{\mathbf{C}}(z)+1)(m\,w_{\mathbf{C}}(z)+m)\left(\frac{m\,w_{\mathbf{C}}(z)}{b}\right)^{-2/\alpha}; \tag{3.68}$$

gathering m:

$$z = m^{2-2/\alpha}(w_{\mathbf{C}}(z) + 1/m)(w_{\mathbf{C}}(z) + 1)\left(\frac{w_{\mathbf{C}}(z)}{b}\right)^{-2/\alpha}. \tag{3.69}$$

From Eq. (3.20) and from the relation between the resolvent and the spectrum we finally obtain

$$\rho_{\mathbf{C}}(\lambda) = \frac{1}{\pi\lambda}\mathrm{Im}[w_{\mathbf{C}}(\lambda + i0^{+})]. \tag{3.70}$$

Inserting b from Eq. (3.43) and rearranging, Eq. (3.67) takes the form anticipated in Eq. (3.39). Returning to the motivation of the topic, the result described by Eq. (3.69) must be considered an approximation of the curve corresponding to the null hypothesis of absence of correlation in time series with fat-tailed increments.

3.5 Monte Carlo validation

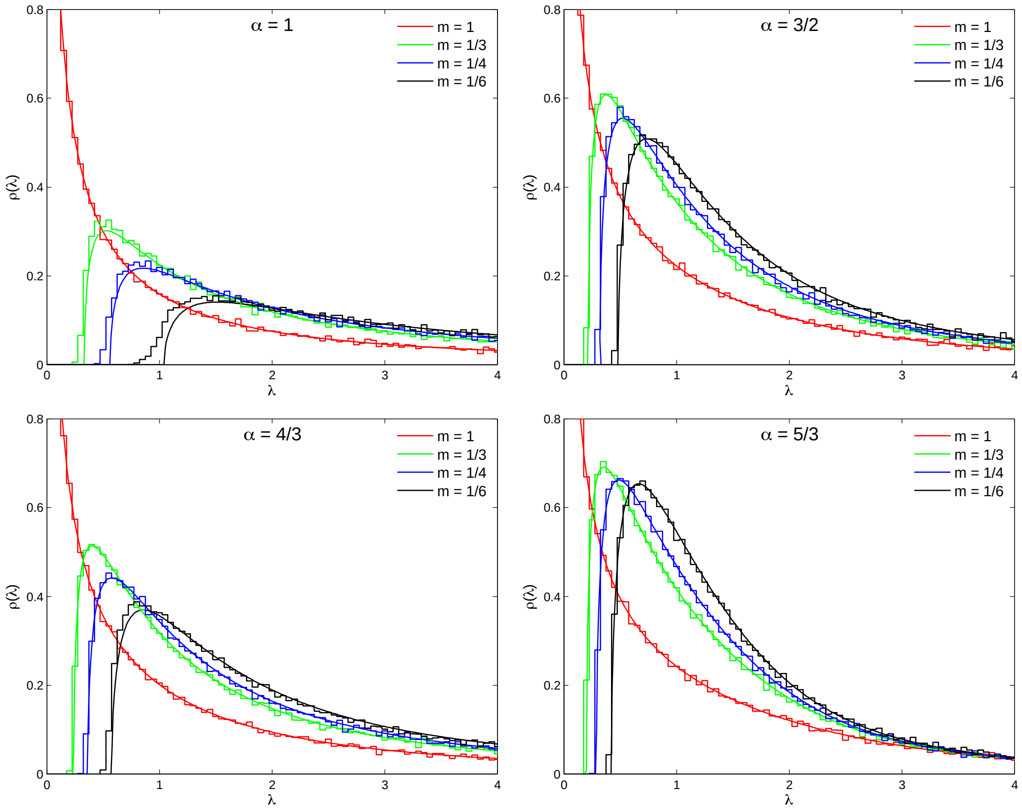


Figure 3.1: Spectral densities from the numerical solution of the analytic equation (solid lines) and from Monte Carlo simulation (stairs). $N = mT = 100$ for $m = 1/3, 1/4, 1/6$ and $N = mT = 400$ for $m = 1$. In each case 19 200 eigenvalues have been considered.

It has already been shown numerically that the theory works in the Wigner-Lévy case [33]. For the Wishart-Lévy case we produce free stable $T \times T$ random matrices via Eq. (3.36). An $N \times N$ Wishart matrix with the desired asymmetry ratio $m = N/T \leq 1$ can be obtained taking a principal minor $\mathbf{C}$ of size $N \times N$ from $\mathbf{\Lambda}$. This is achieved by the projector $\mathbf{P}$ in Eq. (3.35). This projection can be repeated, as there are $n = \lfloor T/N \rfloor$ non-overlapping principal minors. To obtain the minors it is computationally favourable to execute directly Eq. (3.35), i.e. the matrix multiplications $\{\mathbf{P}_i\mathbf{\Lambda}\}\{\mathbf{P}_i\mathbf{\Lambda}\}^{\mathsf{T}}$ where $i = 1, \ldots, n$ labels the projectors that give all non-overlapping $N \times T$ rectangular sub-matrices of $\mathbf{\Lambda}$ with rows 1 to N for $i = 1$, with rows $N + 1$ to $2N$ for $i = 2$, etc.. This is particularly useful if m is small. The averaging procedure can be repeated producing additional realisations of matrices $\mathbf{C}$ until the desired statistical accuracy is reached. The eigenvalue histograms for these minors can be calculated and averaged to give the final spectrum. All plots in Fig. 3.1 have been produced using an equal number of eigenvalues for the sake of comparability.

Free stable laws as defined by the Blue function in Eq. (3.42) and the empirical spectra have different normalisations. For the purpose of a comparison as in Fig. 3.1, this can be corrected dividing $\mathbf{M}$ by a factor $\Gamma(1 + \alpha)^{1/\alpha}$.

This procedure implements directly the definition of the Wishart correlation matrix based on a real random rectangular data matrix $\mathbf{M}$. In this chapter free probability theory has been presented for self-adjoint operators on the right hand side of Eq. (3.35) to provide an analytic equation for the spectrum of a Wishart matrix. Therefore, $\mathbf{\Lambda P \Lambda}^{\mathsf{T}} = \mathbf{\Lambda P \Lambda}$. But if $\mathbf{\Lambda}$ is symmetric, the left largest square minor of $\mathbf{M}$ will be symmetric too, which is not necessary in the definition of the Wishart ensemble. However, by following the pairwise correlation of rows of $\mathbf{M}$ element by element one can see that the properties of $\mathbf{M}\mathbf{M}^{\mathsf{T}}$ remains unchanged. In other words, the symmetrisation introduced for simplicity in the analytic derivation does not change the original numerical problem by introducing correlations. Actually, our Monte Carlo scheme does not use symmetric matrices $\mathbf{W}$ and $\mathbf{\Lambda}$ in Eq. (3.36).

Sec. 3.7 gives the code for the calculation of the spectral density by Monte Carlo as just described.

3.6 Summary and discussion

We have explained the mathematical basis as well as the justification with which free probability theory enters random matrix theory, in particular in the context of the Wishart matrix ensemble. Since the derivation of the analytic solution for the spectra of free stable random matrices has not been published in a self-contained way yet, we recollected it in detail. Then we validated numerically with Monte Carlo calculations the analytic prediction of the eigenvalue spectrum for free stable Wishart-Lévy matrices given in Refs. [24, 25, 26, 30, 33]. Overall we find an excellent consistency between theory and simulation.

We must be aware however, that a realistic situation in financial data analysis is not as homogeneous as assumed above, not to speak of the required rotational invariance. It might well be that the null-hypothesis to be used in practice will still be a numerical Monte Carlo calculation using more realistic assumptions. We present a simple version of such a toy market in Sec.6.

Special acknowledgments

We are thankful to Maciej A. Nowak, Jerzy Jurkiewicz and Giulia Iori for useful explanations.

3.7 Computer codes

The numerical solution of Eqs. (3.69–3.70) can be computed with MATHEMATICA [1] in almost one line. The constant *SOL* is a positive integer that indicates which of the possible solutions to pick. A value of α not expressed as a fraction of integers causes a dramatic increase in running time.

α *= 3/2;*
m = 1/3;
width = 0.01;
lmax = 5;
SOL := 2;
ρ *= Table[l, N[Im[w/.NSolve[-Exp[I2*π/α*]*$w^{2/\alpha}l$ *==* $m^{2-2/\alpha}$
(w+1)(w+1/m),w]][[SOL]]/(π*l)], l, width, lmax, width];*
ListPlot[Abs[ρ*]]*

The Monte Carlo approximation of a free stable random matrix described in Sec. 3.3, the statistical averaging described in Sec. 3.5, and the numerical computation of the eigenvalue spectrum were carried out with MATLAB [2].

```
alpha = 3/2; % index of Levy stable distribution
gam   = 1;   % scale parameter of Levy stable distribution
width = 0.1; % bin width of eigenvalue histogram
N     = 200; % number of time series
T     = 600; % points in each time series; must be >= N.
R     = 20;  % random rotations
S   = 19200; % number of sampled eigenvalues

psi = (T*R*gamma(1+alpha))^(2/alpha); % normalisation factor
rho = []; % set up array of eigenvalues
iS = 0; % initialise normalisation counter

while (iS < S)

    % approximation of a free stable matrix
    L = stabrnd(alpha,0,gam,0,T,T);
    for iR = 2:R
        [O,U] = qr(randn(T,T)); % O is a random orthogonal matrix
        L = L + O*stabrnd(alpha,0,gam,0,T,T)*O';
    end

    % average over covariance matrices
    for i = 1:N:T-N+1
        Li = L(i:i+N-1,:); % choose N out of T rows from M
        Ci = Mi*Mi'/psi; % normalisation
        rho = [rho eig(Ci)']; % collect the eigenvalues
```

```
        iS = iS + N;
        if (iS >= S)
            break;
        end
    end

end

[histrho lrho] = hist(rho,0:width:100); % build the histrogram
histrho = histrho/(length(rho)*width) % normalisation
% lrho contains the abscissa and histrho the ordinate
```

Chapter 4

Random numbers

> *"Begin at the beginning,", the King said, very gravely, "and go on till you come to the end: then stop*
>
> *L.C.*

We present a rejection method based on recursive covering of the probability density function with equal tiles. The concept works for any probability density function that is pointwise computable or representable by tabular data. By the implicit construction of piecewise constant majorizing and minorizing functions that are arbitrarily close to the density function the production of random variates is arbitrarily independent of the computation of the density function and extremely fast. The method works unattended for probability densities with discontinuities (jumps, poles, cusps). The setup time is short, marginally independent of the shape of the probability density and linear in table size. Recently formulated requirements to a general and automatic non-uniform random number generator are topped. We give benchmarks together with a similar rejection method and with a transformation method.

The speed of many one-line transform methods and their contemporary implementations for the production of, for example, Lévy alpha-stable random numbers (Chambers, Mallows and Stuck) and Mittag-Leffler random numbers (Kozubowski and Rachev) is very high and satisfactory for most purposes. However, for the class of strictly decreasing probability densities fast rejection methods like the Ziggurat implementation by Marsaglia and Tsang promise a significant speed-up if it is possible to find a sampling method that handles the tails of the infinite support. This requires the fast generation of random numbers greater or smaller than a certain value. This chapter presents a method to achieve this, and also to generate random numbers within any arbitrary interval. We demonstrate the method showing the properties of the transform maps of the above mentioned distributions as examples of stable and geometric stable random numbers used for the stochastic solution of the space-time fractional diffusion equation.

4.1 Non-uniform variates for arbitrary densities with finite support

4.1.1 Introduction and background

This article introduces a setup method for a rejection algorithm for the production of random numbers with an arbitrary probability density functions (PDF) with finite support and that is at least pointwise computable. The key feature is fast production of random variates in computational applications and simple applicability to any probability density with any number of modes. The principle consists of covering the surface under the PDF with *equal* tiles for which no a priori information is required. The speed of random number production is arbitrarily independent of the shape and computational cost to evaluate the PDF. Prior to the introduction of the method we give a brief review of the subject, some existing methods and terminology.

The two topics, uniform and non-uniform random number generation, are rather disjoint. The literature and associated communities do not overlap a lot. This is not surprising as the respective problems are quite distinct. The generation of non-uniform random numbers can often be considered as a subsequent task, i.e. it requires a uniform generator, usually employed as a black box. The quality of non-uniform numbers depends strongly on the quality of the uniform numbers. Contrary to expectations, the past twenty years have seen a considerable development of uniform random number generators. Well into the eighties simple linear congruential uniform random number generators were standard. Due to imprecision in mathematical definition and programming language, hardware requirements, pitfalls but mostly due to the questionable results the efforts to produce the "perfect" random uniform numbers did not cease. As a result of this development we can say that some types of uniform generators are not only outdated but so flawed that one has to advise strongly against their use. This advice points mainly to *all* (multiplicative) linear congruential generators. Finally, and as a general rule, any trade-off with respect to the quality of random numbers is not acceptable if the computation time of random variates is insignificant to the overall application. Thorough reviews are given in Refs. [45, 92, 93, 149] and references therein.

Fast generation of non-uniform random numbers is important in e.g. Monte Carlo simulations [3, 45, 75, 77, 93, 100, 101, 116]. Statistical theory shows how one can produce random variates for any meaningful distribution. Nevertheless, intelligent mathematical but also purely computational methodology was developed to achieve speed for well known analytic and invertible distributions, for non-analytic but transformable distributions and also for empirical PDFs that only exist as tabular data. In the earlier days of computing any progress was taken very seriously [112, 113, 114, 115] in applied mathematics but also more recently new perspectives on seemingly converged methodology on, for example, Gaussian distributed random numbers can be found [96, 152, 170].

A plethora of mathematically involved publications was inspired by the practitioner's need to increase the speed and quality of non-uniform random number production in applications of statistical computing. Another big driving force is simplicity of application. Special requirements of initialisation for example are a nuisance. Each context and application provides different and often opposite challenges. For example the famous Ziggurat method by Marsaglia [116] and its implementation by Marsaglia and Tsang [117] is a non-truncating method within the narrow class of symmetric, strictly decreasing, analytic and invertible densities. Other algorithms existed long before, but the method's appeal is that the specific *implementation* is even faster than any other that is specialized entirely on exponential or normal distributed numbers. The required initial data structure setup depends on parameters that so far have been published only for the exponential and normal distributions [117] and are difficult to derive automatically [152]. A more general approach is taken by Ahrens [3]. This well-known method is able to process any tabular data that fulfills few restrictions on smoothness, but the setup and production of random numbers is slower. The latter two methods are related to more general strip or slice methods — already existing for a long time [45, 114, 115, 131, 140] — but are of much higher importance in computational applications. It should be kept in mind that information on the speed of a method is only meaningful with respect to a particular implementation and hardware. Some famous methods are actually specialized implementations that rely on the cache memory of contemporary processors.

There are many collections of specialized non-uniform random number generators, usually more than one for a particular class of PDFs, each of them consisting of tailored code. The generators can be categorized into two classes: a) A setup of some data structure is carried out before the first random variate is drawn, and b) a setup is not needed, e.g. with PDFs for which inversion methods exist. Clearly, any "universal" method will need a setup, as explained well for instance in Ref. [102]. In statistical computing the user taps from these collections of software choosing a particular generator. Ideally one can also choose gradually between large setup time and fast generation of random numbers or fast setup and slow generation. The drawbacks of such collections can be huge codes, each bug prone, and the increasingly intractable specialties of the requirements for the setup. Furthermore, the classes of available distributions are limited in the end. The ultimate goal is a universal, easy to use and also fast black box generator [102]. The definition and limitations of a *universal* random number generator, however, is often imprecise in many publications. For example, it does not have any built-in knowledge of the PDF, except that a meaningful PDF can be known only in as much as the number of modes is not infinite [4] and that the modes are not infinitely thin, i.e. meaningless delta-like functions. For universal applicability all transformation methods drop out because they cannot be found automatically by a general algorithm. Therefore, in the general case only some approximation like pointwise data will be available to represent the PDF, implying that the support of the distribution cannot be infinite either. Any computational approach subdues to this restriction. If the PDF can only be evaluated at horrendous costs and no inversion method exists, then no procedure will be able to produce usefully random numbers. Thus, we can assume that the evaluation cost of the PDF for the appropriate number of points is within a similar order of magnitude as the overall task within which the random numbers are used, e.g. a Monte Carlo calculation.

In some cases of PDFs with infinite support truncation of the probability density can be justified with statistical negligibility of the tails. In few cases of distributions with infinite support and where the PDF is at best pointwise computable as with the Lévy distribution [135, 136], which we use as a benchmark, transformation methods that do sample the infinite range were found [37], at least within overflow limitations. Such heavy-tailed distributions deserve attention if truncated early unless justified (or required) by the application [35, 36, 110]; however, this aspect is largely left un-discussed in the literature. State of the art methods, e.g. Refs. [45, 75, 77, 93, 100, 101], construct in the setup phase a majorizing (or envelope or comparison) function and usually also a minorizing (or squeeze) function (see Sec. 4.1.2.2 for an introduction) with secants or other segmentations to be used within a rejection technique with look-up tables. In addition to truncation, in many methods it is often also required to know the approximate or even exact location of the mode (of which mostly only one is allowed), while the method is still declared to be suitable for "arbitrary PDFs" [54, 98]. More limitations to "general methods" are explained in Ref. [54], where it is argued that the modes of the PDF must be known beforehand for certain techniques to be suitable for "arbitrary densities". The same authors also construct "general algorithms" that depend on concavity properties and analyticity of the density. It is common in several methods, e.g. adaptive procedures for log-concave distributions to use the value of maximum density explicitly in the setup of the approximation of the comparison function [45]. Another example is the transformed density method [99], that employs a strictly monotonically increasing differentiable transform such that the transformed PDF is concave. This method is also considered universal. In the improved ratio of uniforms method [98] the setup is restricted to certain classes of distributions if the required transform of variables must yield a region that can be sampled efficiently. In some cases one often resorts to a rejection technique and the concept of squeeze functions, as also employed in the methods presented here, to improve the situation. Yet another general and adaptive approach constructs a polynomial approximation of the inverse distribution function in $X = F^{-1}(U)$ that is stored in tables to be used for interpolation during production [76]. For the class of log-concave distributions Ref. [65] introduces piecewise exponentials for the approximation of the majorizing and minorizing functions using previously sampled points of the density function. This method however is dedicated to the context of Gibbs sampling where each variate is usually drawn from different densities. Finally, closing the topic of piecewise approximation, we mention the approximation of arbitrary densities via a mixture of simpler densities. An interesting example is the triangular approximation giving a piecewise linear approximation of the target density via many overlapping triangle densities, which recently was also implemented in hardware [169]. The most generally applicable method published so far that is also fast is due to Ahrens [3, 4]. It can deal with more than one mode without a priori information on the modes.

Overall we find inevitably that methods titled "universal", "automatic", "black box", "out of the box" and combinations thereof clearly cannot sample an infinite support of the PDF, are restricted to certain classes of density functions or are not automatic out of the box. But this verdict is much less restrictive in realistic applications of statistical computing where the requested distributions can always be represented for example in terms of (interpolated) pointwise data or other approximations to any computationally sensible accuracy and finite support limits. We stress that approximations of the density function via previous sampling of points or the approximation of the inverse distribution function F^{-1} as mentioned above are like other similar concepts a common approach in the field of non-uniform random numbers [45, 76]. It is accepted to truncate the originally infinite support, if statistically justified. Moreover, if the PDF is given as an arbitrarily accurate approximation, the location of extrema can always be determined within any required accuracy in finite time. Our method is applicable in this context and belongs to the type of rejection and segmentation methods as the ones by Ahrens [3] and Marsaglia [117]. In this context and for our claim we can therefore continue to speak of arbitrary PDFs since this nomenclature is widely accepted within the literature. Therefore, in the quest for a universal random number generator, a method can be called universal if it can process *arbitrary finite density function data* without any further information. Techniques that take finite samples of the desired distribution and then try to match this distribution [98] are not discussed here.

Lately easy applicability has become more important. The initial motivation for this work was to overcome the setup difficulty of the Ziggurat method for the general symmetric monotonic decreasing case. Eventually, we developed a simpler method for a significantly more general class of PDFs, at the cost of a moderate performance penalty due to larger memory requirement as compared to the original Ziggurat implementation of Ref. [117] and to specialized transformation methods, were available.

With the intention to provide a quintessence of recent research and demands of statistical computing a wish-list of requirements to a universal random number generator was presented in Ref. [102], which we quote here:

1. Only one piece of code, debugged only once.
2. By a simple parameter choose between fast setup and slow generation or long setup time and fast generation.
3. It can sample from truncated distributions.
4. The rejection rate can be made as close to zero as desired, i.e. as close to inversion as one wants.
5. The setup time is independent of the density function and is faster than many specialized generators.
6. The quality of the non-uniform random numbers is as good as the underlying uniform random numbers.

Point 5 should be made more precise. It refers to the independence of the shape of the density function, but a density with complicated shape usually requires more information, in particular in regions of high curvature. If the input size increases, the setup time is indeed allowed to grow. There is no obvious universal measure that gives the minimum input size required for the suitable representation of a function. This is responsibility of the scientist. Several of the methods mentioned in the above overview are considered to meet these requirements. This is also the case with the method presented here plus additional relaxations with respect to the properties of the PDF.

In Sec. 4.1.2 the tiling is introduced along with some numerical considerations, an explanation of the role of the squeeze function, and a proof of correctness. Sec. 4.1.3 shows that the tiling procedure is capable of dealing *unattended* with poles and other discontinuities. Sec. 4.1.4 gives benchmarks of typical and a-typical situations. Sec. 4.4 summarizes and provides a short discussion. We chose for comparison well-known methods for specialized distributions (Gaussian and exponential) but also a difficult non-analytic distribution (Appendix) for which a transformation method is available. We explain computational issues that are usually ignored but are decisive for performance. This serves in placing the tiling method into the right context among other methods with respect to speed and applicability.

4.1.2 The tiling and numerical considerations

4.1.2.1 The tiling procedure

For any computational task a PDF with finite support, even one with a non-invertible distribution for which no specialized method exists, can be represented either as a) a sufficiently good approximation that can be evaluated sufficiently fast, e.g. by series expansion or polynomials, or b) as tabular data for interpolation. Any feature of a meaningful PDF $f(x)$ can be represented in the latter case by varying the sampling density of the tabular data that represents the PDF in the form $(x_1, f(x_1)), (x_2, f(x_2)), \dots$ [3]. However, since the tiling is completely independent of such considerations, we will simply speak of "evaluating $f(x)$". Furthermore, it can safely be assumed that the PDF can be evaluated in a finite time comparable to the duration of the application within which the random variates are to be used. With these prerequisites the determination of local extrema is achievable in $\mathcal{O}(N)$ where N is the number of data points.

For the following considerations the integral over the density function is required only up to a constant factor $C = \int_a^b f(x)dx$. The rejection method does not require $C = 1$. The tiling concept is simple, see Fig. 4.1: The area under the PDF $f(x),\ x \in [a, b]$, is covered with rectangular tiles of equal area. The procedure starts from one single tile $b - a$ wide and $\max(f(x))$ high. Choosing an initial tile larger than required by the support and the maximum did not show significant influence on the outcome in all cases we tested. The initial tile is split into four equal tiles, and so on recursively. At each refinement cycle *all* tiles are split. Those that lie entirely above the PDF are discarded in each cycle. The splitting can be stopped once a given accuracy of the covering is reached; Sec. 4.1.2.2 explains the details of the calculation of this condition. Fig. 4.1 shows a truncated asymmetric Lévy PDF with parameters $\alpha = 1, \beta = 0.7, \gamma = 1, \delta = 0$ according to the S_0-parametrization convention [135, 136]. We use this distribution as an arbitrary example for comparisons that provides fat tails and for which a fast transformation method is available; for details see Sec. 4.2.2. The support is chosen small to produce deliberately a visible truncation. Fig. 4.2 shows the tiling of a bimodal PDF.

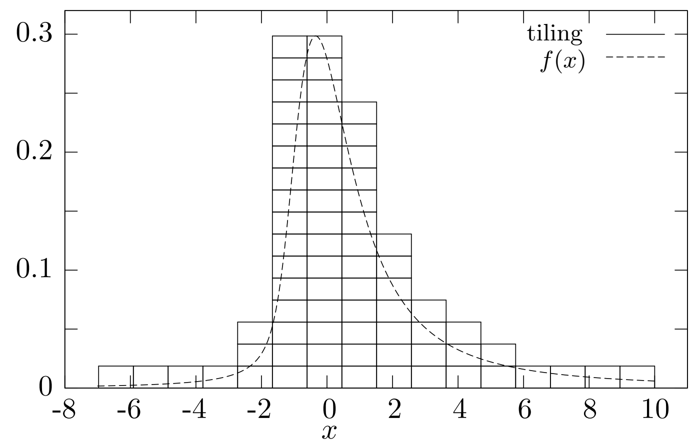


Figure 4.1: For the intuitive introduction of the tiling procedure the plot shows an early refinement stage in the tiling of a truncated asymmetric Lévy PDF with parameters used as an arbitrary example.

The above *recursive* procedure may be considered the most elegant and simple way to construct the tiling. For the subsequent production stage it is irrelevant however if the tiling was constructed, for example, by plastering, i.e. starting from a small initial tile somewhere within the support.

Thus, the tiling constructs a piecewise constant majorizing function $g(x)$ of the PDF $f(x)$, with $g(x) \geq f(x)\ \forall x \in [a,b]$. The closer $g(x)$ to $f(x)$, the better. The universal von Neumann rejection method has two main steps:

a) Generate a random $X \in [a,b] \sim g(x)$ and a random uniform $Y \in [0, g(X)]$.

b) Accept X if $Y < f(X)$, otherwise reject it and repeat the procedure.

The rejection rate is given by the ratio R of the areas under the PDF and the comparison function:

$$R = 1 - \frac{\int_a^b f(x)dx}{\int_a^b g(x)dx} = 1 - \frac{1}{NS}\int_a^b f(x)dx. \tag{4.1}$$

The denominators correspond to the sum over all N tile surfaces S which are equal.

At this point one could think that an adaptive scheme would be more appropriate, e.g. only tiles intersected by the PDF are split or even deformed to fit the boundary better, or tiles lying below the PDF are merged. Indeed this is common practice in computer graphics and some approximation methods. However, this measure to save memory is not recommended here; actually, it is to be avoided for the sake of simplicity and speed. In the production stage the probability of random selection of a tile would have to be proportional to its area to guarantee uniform probing. This is more complicated and slower especially if the area ratios are not integer. Moreover, a uniform random coordinate is more expensive to produce in shapes other than rectangles. Additional details to why the segmentation into

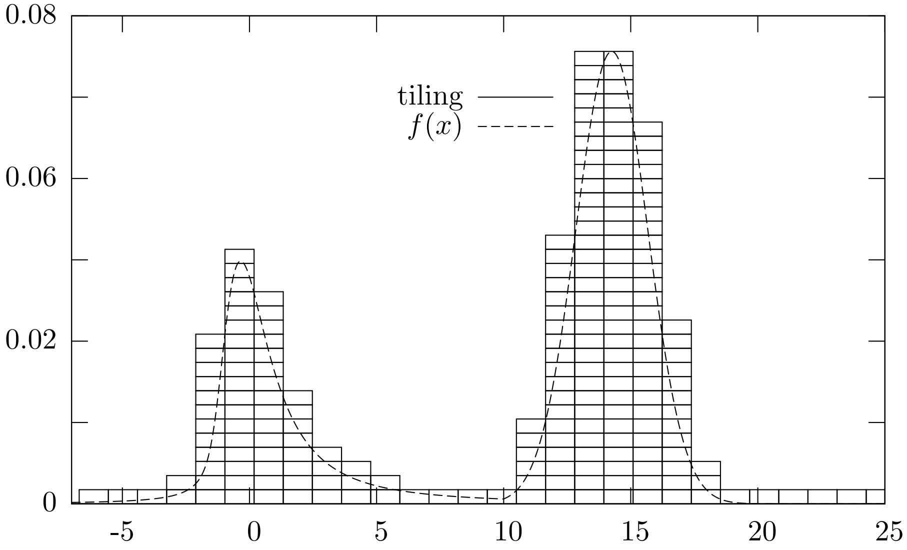


Figure 4.2: Tiling of a bimodal probability density function. $f(x)$ is composed of two Lévy density functions with parameters $\alpha = 1$, $\beta = 0.7$, $\gamma = 1$ (left part) and $\alpha = 2$, $\beta = 1$, $\gamma = 1$ (right part); the heights are adjusted to fit the curves seamlessly at $x = 10$.

equal areas is crucial are given in numerous publications [3, 45, 116, 117]. Instead of using strips of different height and width or even different shape as in other methods, here we suggest *equal* tiles as the clearly simplest and fastest approach. This is the key idea in this work.

Now the von Neumann rejection can be implemented with a modified step a:

a) Generate a random tile index $i = 1, \ldots, N$; generate a random coordinate (X, Y) within tile i.

b) Accept X if $Y < f(X)$, otherwise reject it and repeat the procedure.

This way we are able to sample efficiently the majorizing function $g(x)$. Moreover, the evaluation of the condition in b) is hugely sped up using the implicitly constructed minorizing function as explained in more detail in the following Sec. 4.1.2.2.

Although the sampling with tiles seems sufficiently intuitive and equivalent to analogous methods of this kind [116, 117] we give nevertheless a reasoning on the correctness.

THEOREM The introduced sampling of the comparison function is equivalent to the standard von Neumann sampling, i.e. $g(x)$ is sampled uniformly within all tiles generated on the support $x \in [a, b]$.

PROOF Define $I = \{i_1, i_2, i_3, \ldots, i_N\}$ the set of tile indices and $I_j \subset I$ the subset of all indices $i_1^j, i_2^j, \ldots$ corresponding to a particular tile column j with width $\Delta x = (b - a)/r$, where r is the number of columns. Thus $\bigcup_j I_j = I$. Construct an bijective mapping $i_k^j \to n_l^j$ with $n_l < n_{l+1}$. The mapping is purely a renaming of indices in column j. So we have $n_l^j = 1, \dots, n_{\max}^j$, $n_{\max}^j = g(x)/\Delta y$ where Δy is the height of the tile. Note that within column j the function $g(x)$ is constant. Now define a random number $Y^j = n_l^j \; u \; \Delta y$ with uniform random $u \in [0, 1)$. The index n_l^j is random by the random choice of i_k^j and the subsequent mapping. Then $Y^j \in [0, g(x))$ is a uniform random number in column j and we arrive at the standard situation of the rejection method for the interval $x \in \Delta x_j$: Generate a uniform coordinate (X^j, Y^j) with uniform $X^j \in \Delta x_j$ and reject X^j if $Y^j > f(x)$. The sampling of j is implicitly proportional to the size of I_j, i.e. the height of column j, due to the uniform sampling of tile indices $i \in I$ and $\bigcup_j I_J = I$. Therefore X^j is sampled as desired according to $g(x)$ and the sampling of pairs (X, Y) is achieved with $X \in \bigcup_j \{X^j\} \sim g(x)$. □

The correctness of the standard rejection method can be taken for granted since the seminal paper by John von Neumann [176].

4.1.2.2 Implicit squeeze function

The tiling also constructs implicitly a so-called squeeze function $q(x)$ that fulfills the condition $q(x) \leq f(x) \leq g(x) \leq g(x)$ within the required interval $[a, b]$. This is the usual definition of the squeeze and comparison functions, see for example Ref. [93]. $q(x)$ is the upper edge of the top tiles lying completely underneath $f(x)$, or equivalently the bottom edge of the tiles intersected by $f(x)$. The role of the squeeze function is to reduce the number of evaluations of $f(x)$ if $q(x)$ can be evaluated faster: In the setup all tiles below $f(x)$ are labelled and the test $Y \leq q(X)$ involves no computation — just one label look-up. Actually Y must not be generated at all for tiles that are not intersected by $f(x)$. The latter is the key advantage of the squeeze function. Thus the following modified steps implement the von Neumann rejection:

a) Generate a random tile with index $i = 1, \dots, N$; generate a random X within tile i.

b) Look up if tile i is labelled as "$< f(x)$". If yes, accept X. Otherwise generate Y within tile i and compare $Y < f(X)$. If yes, accept X. Otherwise reject it and repeat the procedure.

With dense tiling most X are accepted in b) by one table look-up only without the generation of a second real coordinate Y. The PDF itself is hardly ever evaluated. The relative number of evaluations of $f(x)$ per non-uniform variate is given by

$$E = 1 - \frac{1}{NS} \int_a^b q(x) dx. \tag{4.2}$$

The integral over the squeeze function is given by the sum of all tile surfaces not intersected by $f(x)$. Thus, the number of evaluations of $f(x)$ can be greatly reduced and is equal to the area fraction of the border tiles. Both numbers R and E are cheaply calculated on the fly, so that the resulting rejection rate can be pre-imposed as a condition for the tile refinement. The latter results will be reconsidered in Sec. 4.1.2.3 on the distribution cutoff.

To have a better measure of the "quality" of $g(x)$ and $q(x)$ we estimate an upper limit for the probability density p_E that $f(x)$ must be evaluated for one non-uniform random number. Define $\Delta x := (b-a)/n$ where n is the number of columns, so Δx is simply the final width of the tiles. For $\Delta x \ll b-a$, i.e n sufficiently high, $f(x)$ can be assumed linear in the interval Δx. Then

$$p_E(x, \Delta x) \propto \frac{b-a}{r} \frac{\mathrm{d} \log f(x)}{\mathrm{d}x}. \tag{4.3}$$

This expression is deduced from the ratio of areas contained in a tile column corresponding to $Y \leq q(x)$ and $q(x) < Y \leq g(x)$ respectively.

4.1.2.3 Distribution cutoffs

In the introduction and thereafter we explained that all procedures that are not specialized to particular analytic and thus invertible distributions will never sample an infinite support. Considerations on the appropriate cutoff apply only to special distributions [45]. If the support of the PDF $f(x)$ is infinite, a general algorithm will inevitably reduce it to a reasonable finite interval $x \in [a, b]$. It is the scientist's responsibility to control appropriately these support limits.

However, the period length L of the $[0,1]$-uniform generator used in the sampling along the abscissa must satisfy the condition $f(x) < 1/L$ at both limits a, b [4]. This situation appears for example in the standard rejection method or the Ziggurat method. In the latter, the $[0,1]$-uniform generator must sample the whole bottom strip. The sampling procedure in our method lifts this limitation by the number of columns $n = 2^{r-1}$, where $r = 1, 2, \ldots$ is the refinement level: A random integer is generated to sample a tile and a subsequent uniform X is generated *within* the tile. In practically relevant cases the number of tiles will always be exceedingly smaller than the period length of any sensible random integer generator. Fat (or somehow long) tailed distributions deserve attention for the above reason.

4.1.3 Discontinuous probability densities

The literature also considers density functions which contain a pole (which numerically is indistinguishable from a cusp) [4], i.e. $f(x) \to \infty$ as $x \to c^+$ or $x \to c^-$ in the range of interest $[a, b]$. Within the standard von Neumann rejection method [176] a pole is dealt with as follows: Choose $\epsilon \ll 1$ and assign the cumulative probability

$$P_c = \int_{c-\epsilon}^{c+\epsilon} f(x) dx \tag{4.4}$$

to the interval $[c-\epsilon, c+\epsilon]$, a so-called mass point. If the [0,1]-uniform deviate is smaller than P_c return c. Otherwise sample from $[a, b] \backslash [c-\epsilon, c+\epsilon]$. If c and $c \pm \epsilon$ have the same numerical representation then no better method exists to sample from $f(x)$. Usually this situation must be treated computationally as a special case in the setup and production phases.

The tiling procedure and subsequent production works unchanged with an appropriately approximated (or modified) density function as follows. Fig. 4.3 shows the situation of a density function with a pole at $x = c$. Figure dimensions, especially the vertical scale, are exaggerated to convey intuitively the geometry. Choose $[c - \epsilon, c + \epsilon]$ and modify $f(x)$ yielding $\bar{f}(x)$ such that the cumulative probability P_c according to Eq. (4.4) is preserved (hatched area in Fig. 4.3):

$$\int_{c-\epsilon}^{c+\epsilon} f(x)dx = \int_{c-\epsilon}^{c+\epsilon} \bar{f}(x)dx \tag{4.5}$$

which gives the implicit condition for $\max(\bar{f}(x))$:

$$\max(\bar{f}(x)) = \frac{1}{2\epsilon}\int_{c-\epsilon}^{c+\epsilon} f(x)dx. \tag{4.6}$$

This is the minimum value of the height of the initial tile. If ϵ is chosen sufficiently small with numerical or/and statistical reasoning, the result will be identical to the procedure in the standard von Neumann rejection described above.

One has to be aware that the choice of ϵ as the smallest representable “distance” from the position of the pole is unnecessarily restrictive. Any statistical verification requires a significant number of deviates to fall in the region of the pole to reveal a possibly too large value for ϵ. Depending on the error norm and test method it is likely to turn out that ϵ can safely be chosen magnitudes larger than the initial numerical consideration. The statistical needs of the application must be considered in any case. Thus, there is no generally obvious upper limit for ϵ.

An example application is the scaled symmetric modified Bessel function of the second kind $K_0(|x|)/\pi$, which is the density of the product XY, where X and Y are independent normal distributed random numbers. $K_0(|x|)$ diverges at $x = 0$. A possible setting could be the following. Restricting the support to $x \in [-15, 15]$ accounts for over 99.99999% of all mass. With $\epsilon = 0.00001$ we get a fraction of 8.03978×10^{-5} of the mass contained in the interval $[0 - \epsilon, 0 + \epsilon]$. The number of recursive refinements is given by $\lceil \log_2(30/(2\epsilon)) \rceil = 21$. About 235000 tiles are retained to cover the density function, corresponding to two megabytes memory in our data format. Benchmarks on the setup are presented in the next section. Although it is quicker to multiply directly normal random variates, this example demonstrates the applicability of the method to densities with no simple alternative.

PDFs with first order discontinuities or jumps are implicitly contained in the above case. With a suitable interpolation scheme one can simply use tabular data to model the jump from one data point $(x_i, f(x_i))$ to the next $(x_{i+1}, f(x_{i+1}))$ and fix $(x_{i+1} - x_i) \approx \epsilon$ as close as possible. There will be no deviates falling in $[x_i, x_{i+1}]$. This situation is contained schematically in Fig. 4.3 as well, showing two consecutive but very near data points for the left flank of a jump.

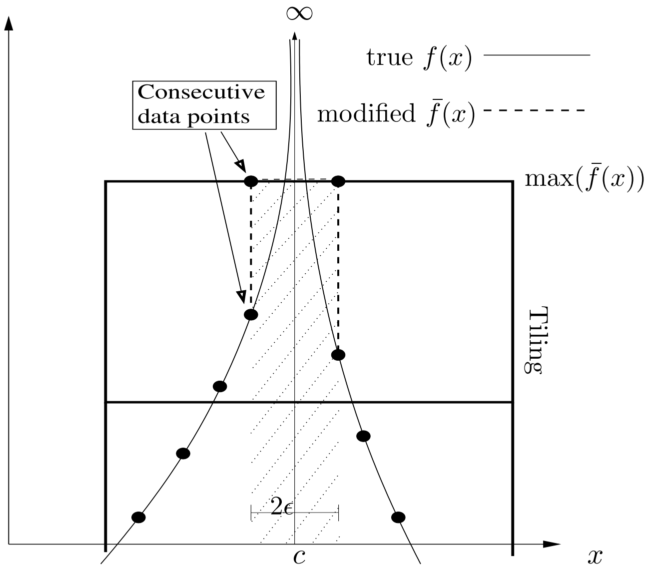


Figure 4.3: Probability density containing a pole at $x = c$ (schematic). The returned deviates are numerically correct if the modified density function $\bar{f}(x)$ fulfills the condition of equal area (hatched region) for $x \in [c - \epsilon, c + \epsilon]$ and $\epsilon \ll 1$ sufficiently small. The initial tile is chosen $\max(\bar{f}(x))$ high. The true or modified PDF can be approximated via data points as shown or by any other method that implements this condition. The mathematical jumps in the modified PDF can be modeled numerically via two very close consecutive data points along x.

4.1.4 Measurements and comparisons

In general, speed comparisons are not obvious to do and interpret and only meaningful with respect to a particular software and hardware implementation. With increasing optimization of code, the mathematical description of a method and its implementation become inseparable. We must stress this and point to technical aspects that are responsible for a speed difference of two orders of magnitude, even though the mathematical/algorithmic description is identical. The focus of this material is not pure speed, but portability and easy applicability in compromise with speed.

All measurements were performed on a desktop PC with a 2.4 GHz Intel Pentium 4 processor using the GNU C++ compiler version 3.2.2 on Red Hat Linux. We explain below the importance of using a multi-tasking operating system in its standard operation mode with a typical process time slice during the measurements. This will almost always be the case with applications in statistical computing.

4.1.4.1 Memory requirements

At the start the graph of the PDF is embedded in one tile. The memory requirements for a Gaussian-like density function and rejection rate below 0.02 [117]

Refinement level r	Number of tiles N	Rejection rate R	Evaluation rate E
1	1	0.813	1
2	3	0.750	1
3	8	0.627	1
4	24	0.502	0.910
5	70	0.317	0.650
6	238	0.196	0.390
7	857	0.108	0.220
8	3 246	0.058	0.110
9	12 609	0.029	0.058
10	49 685	0.015	0.029
11	197 233	0.007	0.013
12	785 936	0.002	0.005

Table 4.1: Number of tiles, rejection rate and evaluation rate for the uni-modal PDF shown in Fig. 4.1, but with a larger cutoff at $x = \pm 64$. Refinement level 5 is shown in Fig. 4.1. The memory needed to store 49 685 tiles (refinement level 10) is ca. 0.4 megabytes. The evaluation rate tells how often $f(x)$ must be evaluated per non-uniform random number.

is never more than a few megabytes. For details on a uni-modal example as in Fig. 4.1 see Table 4.1, for the bimodal case in Fig. 4.2 see Table 4.2. Only obnoxious density functions with fat tails and many sharp peaks require more memory. In the tested variations of such extreme cases using multiple peaks and the support truncated very far out the memory needed to achieve a rejection rate below 0.02 did not exceed 10 megabytes (about one million tiles). This happens using two numbers to store the coordinates of one tile and is more than acceptable for contemporary desktop computers. We skipped entirely memory optimization and removal of redundancy since we preferred a clear class structure and simple data management. Setup time can be reduced via speed optimized data structures, but that typically increases computation time and storage. Just a decade ago the above memory requirements were large for a standard desktop computer with a few megabytes memory. This may explain why this fairly straightforward method has not been proposed before.

4.1.4.2 Speed of random variate production

With the SHR3 uniform RNG [117] on the above mentioned configuration our method produces 2.6 million non-uniform random numbers per second *independently* of all tested PDFs. In the following we discuss a few pitfalls of speed measurement and code execution, and we compare to other methods. The benchmarks refer to methods and implementations that appear most similar or useful in judging the tiling method. In any case, the comparisons cannot be entirely fair since each method has different specialities.

In rejection methods the speed of random variate production is arbitrarily in-

Refinement level r	Number of tiles N	Rejection rate R	Evaluation rate E
1	1	0.858	1
2	4	0.858	1
3	9	0.747	1
4	23	0.605	0.956
5	70	0.481	0.871
6	213	0.317	0.582
7	718	0.189	0.356
8	2 602	0.106	0.195
9	9 859	0.056	0.104
10	38 324	0.029	0.053
11	151 068	0.014	0.025
12	599 819	0.007	0.011

Table 4.2: Statistics for the bimodal PDF shown in Fig. 4.2, where refinement level 6 is plotted. The memory needed to store 151 068 tiles for refinement level 11 is ca. 1.2 megabytes.

dependent of the PDF and its representation, whether by data points or a closed formula. The speed depends only on the properties of the comparison and squeeze functions. In all our tested examples with tabular data or simple explicit density functions the evaluations representing $f(x)$ are negligible at a rejection rate below 0.02. Since interpolation or evaluation of density functions is not the topic here, we only give as a rule of thumb that evaluations for 1% of the produced random numbers is sufficiently low for almost all practically relevant densities. The production of one random variate with the desired distribution requires at least two uniform random variates as in most methods. Recently a method was published that can provide non-uniform variates with $1 + s,\ s \in [0, 1]$, uniform variates where s can be made arbitrarily small [102]. However, it turns out that in almost all applications the generation of uniform random numbers is not the major sink of computer time. It is up to the scientist to evaluate the trade-off between a few percent gain in overall speed and quality of the obtained variates. The use of less than two uniform random variates per non-uniform variates in the context of a rejection technique but also the importance of uniform random number quality, in particular in the Ziggurat implementation by Marsaglia and Tsang, is commented in Refs. [22, 49, 95, 140]. Some constructive remarks on the Ziggurat implementation in Ref. [117] can be found in Ref. [130].

We chose as one of the benchmarks the symmetric Lévy α-stable distribution. It is a generalisation of the Gaussian distribution, that is recovered for $\alpha = 2$; see Appendix. The transformation method by Chambers et al. [37] is the contemporary method of choice. As opposed to other published methods, it has no accuracy deficiency, it does not truncate the support and is sufficiently fast for most applications. Moreover, it is applicable to asymmetric Lévy α-stable deviates too. We use an implementation in C++ for the purpose of this comparison. It is about 3 times faster than our method on the above mentioned test configuration.

We also compared to the most efficient implementation of the Ziggurat method [117] for $\exp(-x)$ and $\exp(-x^2)$ distributed variates. This implementation is considered the fastest for these two distributions. The exponential and normal densities could be wired into the code exploiting their mathematical properties and using inline coding. In the limit of a negligible rejection rate, this Ziggurat implementation could produce 232 million variates per second. This means one variate per 10 CPU clock cycles! It is important to note that this number could only be achieved if executed alone without any other code, for example within a Monte Carlo application. This speed may be surprising at first sight since the rejection principle is quite similar to the tiling method. Actually there are profound differences. First and most obviously, the number of tiles is not a power of two. Choosing randomly between exactly 2^8 or 2^7 objects is faster if one uses 8 bits of the 32 bit XOR shift RNG as in Ref. [117]. Secondly, it is stated self-evidently in Ref. [117] that small code is important. This purely technical issue is hardly ever explained in the literature on random numbers despite being highly technical on several occasions. Numerical literature [148, Chap. 7] finally picks up this issue and also more recently in Ref. [152], but only briefly say *why* small code is important. We outline the situation.

CPUs use hierarchical memory to speed up computation. The access to the internal cache memory is magnitudes faster than to the external main memory. However, the code and data fitting into this cache is not the only condition for faster execution. An algorithm hard-wired in the CPU transfers repeatedly and frequently used sections of memory into the cache and also considers the size and distribution of the data over the memory banks. A good implementation (and compiler) therefore tries to minimize cache misses by arranging data of subsequent memory accesses into the same cache line. The latter are sequences of bytes transferred into the cache with each memory access. This statistics is disrupted by cache misses that are also provoked by a process switch of the operating system at built-in time intervals or other events. Small code might therefore end up in the cache for a significant time. Very large code that accesses its data in random fashion as it is the case in the sampling with tiles will not be able to exploit properly cache memory. We can therefore say that the execution of code is subject to decisive factors of hardware, compilation and operating system that can usually not be controlled entirely. It is also known that CPU-specific compilers are able to produce code that can be several times faster than a more generic compiler.

On our typical configuration of operating system and compiler the execution of the Ziggurat code [117] is the fastest by far. The speed factor of ca. 100 to our code is in fact consistent to the latency of low-level memory as compared to second-level cache of contemporary hardware. This speed difference is leveled out considerably if the code and tables of the Ziggurat implementation is forced to leave the cache by executing some arbitrary and larger code alternatingly with calls to the Ziggurat generator. This measure creates a more realistic use case and reduces the execution speed of the Ziggurat code by a factor of ca. 50. A more rigorous analysis of code and hardware interplay would require the *exact* reproduction of the original test environment which is not readily available anymore.

Finally we make a few more technical remarks and comparisons. The implementation of the tiling method is only moderately optimized, but completely portable and uses throughout Standard Template Library arrays. The period of the XOR shift RNG is considered short with 32 bit arithmetic but modification to higher models is possible. Following the results in Refs. [49, 130, 140] on quality, resolution and portability we recommend a slower and also portable uniform RNG. Refs [22, 49, 95, 130, 140] also comment other problems of the XOR shift RNG in conjunction with the Ziggurat method. The Ziggurat method requires for the decision whether to evaluate the density function one coordinate comparison for each attempt to draw a non-uniform number. Our method requires one table look-up only. But this advantage is not enough to compensate the disadvantage of a large table and resulting slow memory access.

For accelerated production of random variates to make sense, their part must take up a significant proportion of the overall CPU time. But there is hardly anything do-able within the order of 10 clock cycles. Moreover, fast production of variates imply that enormous amounts are required. This poses very high demands on their quality. The findings above as well as the critical publications on the Ziggurat implementation Ref. [117] encourage to analyse the appropriateness of extremely fast but medium quality variates. A detailed analysis of this issue can be found in Ref. [170].

4.1.4.3 Speed measurements of the setup

The setup part in our implementation is not speed-optimized but turned out to be sufficiently fast for the production of ca. one million variates and above. This includes the extreme examples with more than one mode and a very large support. To provide a meaningful time measurement for the setup we subtract the cumulative time for the evaluations of $f(x)$. For the presented examples we used a standard polynomial interpolation with 7 data points. The setup for a typical uni-modal PDF (Gaussian or Lévy, the latter with sufficiently wide support) with 2^{15} data points takes ca. 0.2 seconds plus cumulative 2.1 seconds for all evaluations of $f(x)$. The calculation time of the Lévy PDF via fast Fourier transform for 2^{15} points is negligible with only 0.2 seconds. Thus, as a rule of thumb, the overall total speed of the setup depends almost entirely on the number of evaluations of $f(x)$. With a constant number of data points the total speed of the setup increases noticeably only for very unusual multi-modal PDFs with many sharp peaks and long tails.

The setup of the Ziggurat for general symmetric, strictly decreasing, non-analytic and safely truncatable PDFs was attempted in Ref. [91]. This setup, our C++ version of the Matlab code from Ref. [91] as well as our own generalised iterative C++ code along the original Ziggurat setup formula [117] is sensitive and computationally expensive. For example, a numerical error in the flat regions of the tail or in the inversion of the PDF can cause a disturbance which often causes a breakdown of the procedure. Precautions to mend this are possible but complicate the code further and do not guarantee unattended functionality. The empirical parameters needed for the setup of the Ziggurat method are an additional difficulty for making the method truly automatic. The setup time depends strongly on the given data and the above mentioned empirical parameters, and is

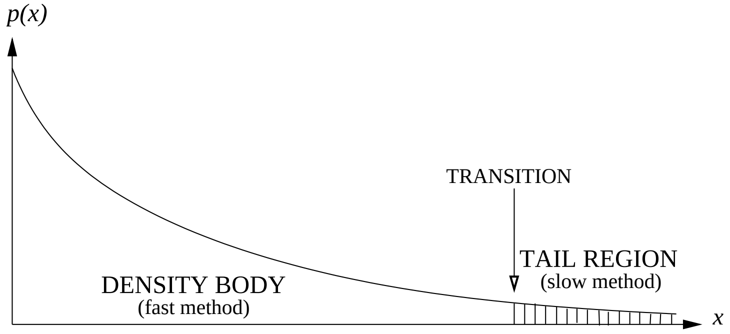


Figure 4.4: Schematic illustration of using two methods for sampling a distribution. A fast method may available only for the body while the tails can be sampled with a slow but accurate method.

at least one order of magnitude slower than the tiling.

4.2 Random numbers from the distribution tails using the transformation and the tiling methods

4.2.1 Introduction

Many numerical methods for the generation of random numbers represent the main body of the probability density using a fast method and the tails using an alternative method. A famous example is the Ziggurat method by Marsaglia and Tsang [116]. Fig. 4.4 depicts the situation schematically. A reason for this apparent complication is that the method for the main body works best and fastest on a finite support or is specially designed for the main body in terms of accuracy or speed. Handling the tails efficiently is often more involved, especially with difficult non-invertible densities with infinite support. Since rarely needed, variates from the tail can safely be generated by a slower method [45, 116, 117]. Overall, a significant speed-up can be achieved. In this chapter we show how to sample directly and efficiently via a rejection technique a random number X such that $X > x$ or $X < x$ where $x \in (-\infty, +\infty)$ at least within the limits of the numerical representation. This is achieved by using properties of the transform representation of the distributions. The examples we use for demonstration are the Lévy α-stable [97, 135, 136] and the Mittag-Leffler one-parameter probability densities [74]. A transform formula for the former is well known [37, 178], while the transform representation of the latter was discovered [46, 79, 84, 85, 86, 87, 139] and applied [62, 63, 64] only recently. The two distributions are generalisations of the Gaussian and exponential distribution respectively and play an important role together for the solution of the space-time fractional diffusion equation.

Our rejection concept is general for any distribution that provides a transform representation. It can sample efficiently from arbitrary finite or infinite intervals as opposed to other existing methods that are designed especially for certain densities. In this work we do not consider the technical details of a speed-optimized implementation, but explain the basis of the algorithm and show example applications. The method is based on properties of the two-dimensional transform maps that seem unnoticed yet.

The assumption for using the method introduced here is that the tail region requires high accuracy due to high demands on statistics as well as speed. The transformation formula by Chambers, Mallows and Stuck [37] for example is exact and for most applications the recommended method for the production of Lévy α-stable random numbers [178]. The replacement of the tails by a simple invertible Pareto function is not totally appropriate because this is only an asymptotic approximation; moreover it introduces a transition region. The more sophisticated and smooth this transition, the more complicated and slower the overall procedure. Such a replacement of the tail contrasts the initial goal of speed. But the most demanding contemporary applications of random numbers [120, 178], as of the two suitable examples we treat here, will require large amounts and therefore fast production. The tails should be accurate without an approximated transition region from the density body to its tails. In some cases fast series expansion methods can be used but with a compromise in accuracy [45]. A more detailed analysis of such considerations can be found in Ref. [170], where most known algorithms for the Gaussian distribution (as a simple and special case of the Lévy α-stable distribution) are analyzed in the context of contemporary statistical applications as well as expectations of future demands. It is argued extensively how speed of production implies the demand for very many random numbers, which in turn requires greater accuracy of the resulting distribution.

Consider the Ziggurat rejection method by Marsaglia et al. [114, 115, 116] that was introduced to produce Gaussian and exponential random numbers. It is an exact method up to the numerical limits of floating point representation. In principle it is applicable to all decreasing or symmetric densities, provided a suitable tail sampling method is available [113]. In particular the implementation by Marsaglia and Tsang [117] and a recent version by Rubin and Johnson [152] are about two orders of magnitude faster on contemporary processors than other dedicated methods for Gaussian and exponential random variates; therefore it is likely to outrun any non-trivial transformation method by at least the same factor. The hurdles to apply the Ziggurat method to other densities with infinite support, with additional parameters and for which no closed form or simple transform exist are: a) the costly setup of the look-up table, b) the necessity of equal areas of the rectangles covering the density as well as the area under the tail and finally c) a reasonably fast and accurate tail sampling method. Difficulty a) must be evaluated in relation to the required number of variates if it is possible to predict the setup costs as a function of the density parameters. The meaning of "fast" in c) is defined by the ratio of tail variates versus body variates and the speed of the body sampling method. A slow tail sampling can always be balanced by sufficiently infrequent calls to the latter.

Provided the complementary cumulative density function $\int_x^\infty f(x')\,\mathrm{d}x'$ can be computed sufficiently exact on demand, then any required value of the tail surface, and thus any relative frequency of calls to the tail sampling function, can be achieved in the setup of the Ziggurat by an iterative process. For the details of the setup refer to Ref. [117] and for alternative concepts to Ref. [152]. Independently of such considerations the production of Lévy α-stable random numbers in the tails, but also in arbitrary finite intervals, are themselves examples where the method introduced in this paper is suitable. Of course the Ziggurat method is applicable to non-symmetric decreasing densities by representing two halves with separate generators which have to be called alternatingly in a ratio that corresponds to the ratio of respective areas covered by each halves.

In Sec. 4.2.2 we introduce the Lévy α-stable probability density on the basis of which Sec. 4.2.3 explains our method. In Sec. 4.2.4 the Mittag-Leffler distribution, its transform representation and transform map are presented.

4.2.2 The Lévy α-stable probability density and its transform map

A convenient representation of the Lévy probability density function in its most popular parametrization [135, 136, 178] is via the inverse Fourier transform of its characteristic function:

$$L_{\alpha\beta\gamma\delta}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_{\alpha\beta\gamma\delta}(k)\ \exp(-ikx)\,\mathrm{d}k \tag{4.7}$$

where

$$\log\phi_{\alpha\beta\gamma\delta}(k) = \begin{cases} -\gamma^{\alpha}|k|^{\alpha}\left(1 - i\beta\operatorname{sign}(k)\tan(\frac{2}{\pi}\alpha)\right) + i\delta k & \text{for} \quad \alpha \neq 1, \\ \\ -\gamma|k|\left(1 + i\beta\operatorname{sign}(k)\frac{2}{\pi}\log|k|\right) + i\delta k & \text{for} \quad \alpha = 1. \end{cases} \tag{4.8}$$

The index or order $\alpha \in (0, 2]$ determines the exponent of the power-law tail. The parameter $\beta \in [-1, 1]$ governs the skewness, $\gamma \in (0, \infty)$ the horizontal scale and $\delta \in (-\infty, \infty)$ the location. The advantage of this parametrization is that the density and the distribution function are jointly continuous in all four parameters, the same applies to the convergence to the power-law tail. The last two parameters can safely be set to 1 and 0 without loss of generality. Other values can be obtained through

$$X_{\alpha\beta\gamma\delta} = \gamma X_{\alpha\beta 10} + \delta. \tag{4.9}$$

We therefore omit γ and δ in the subscripts and also β if equal to zero. The symmetric case with $\beta = 0$ has the simpler form of an inverse cosine transformation

$$L_{\alpha}(x) = \frac{1}{\pi}\int_{0}^{\infty} \exp(-k^{\alpha})\ \cos(kx)\,\mathrm{d}k. \tag{4.10}$$

Rejection methods for Lévy α-stable random numbers that use asymptotic series representations of the density function are sometimes used if speed has highest priority [45]. However, the known types of series expansions for the Lévy density tend to become inaccurate especially in the tails and also account for a certain fraction of uniform random numbers to be lost (rejected) in the sampling. To achieve best performance (minimum rejection rate and maximum accuracy) one must use different versions of the algorithms and expansions depending on the combination of parameter values and their range. This is in particular the case for

$\beta \neq 0$. A review on these methods and their deficiencies can be found in Ref. [45].

A transformation method for Lévy α-stable random numbers by Chambers, Mallows and Stuck has been available for over 30 years [37]. Two independent uniform random numbers $U, V \in (0,1)$ are mapped via a transform $F_{\alpha\beta}(U,V)$ such that $X = F_{\alpha\beta}(U,V)$ is distributed correctly according to $L_{\alpha\beta}(x)$. The general case for $\alpha \neq 1$ is given by

$$X = F_{\alpha\beta}(U,V) = \frac{\sin(\alpha(\Phi+\Phi_0))}{\cos\Phi}\left(\frac{-\log U \cos\Phi}{\cos(\Phi-\alpha(\Phi+\Phi_0))}\right)^{1-1/\alpha}, \tag{4.11}$$

where $\Phi = \pi\left(V-\frac{1}{2}\right)$ and $\Phi_0 = \frac{1}{2}\pi\beta\frac{1-|1-\alpha|}{\alpha}$, while for $\alpha = 1$

$$X = F_{\alpha\beta}(U,V) = \left(1+\frac{2}{\pi}\beta\Phi\right)\tan\Phi - \frac{2}{\pi}\beta\log\left(\frac{-\log U\cos\Phi}{1+2\beta\Phi/\pi}\right). \tag{4.12}$$

The symmetric case with $\beta = 0$ simplifies to

$$X = F_{\alpha}(U,V) = \frac{\sin(\alpha\Phi)}{\cos\Phi}\left(\frac{-\log U\cos\Phi}{\cos((1-\alpha)\Phi)}\right)^{1-1/\alpha}. \tag{4.13}$$

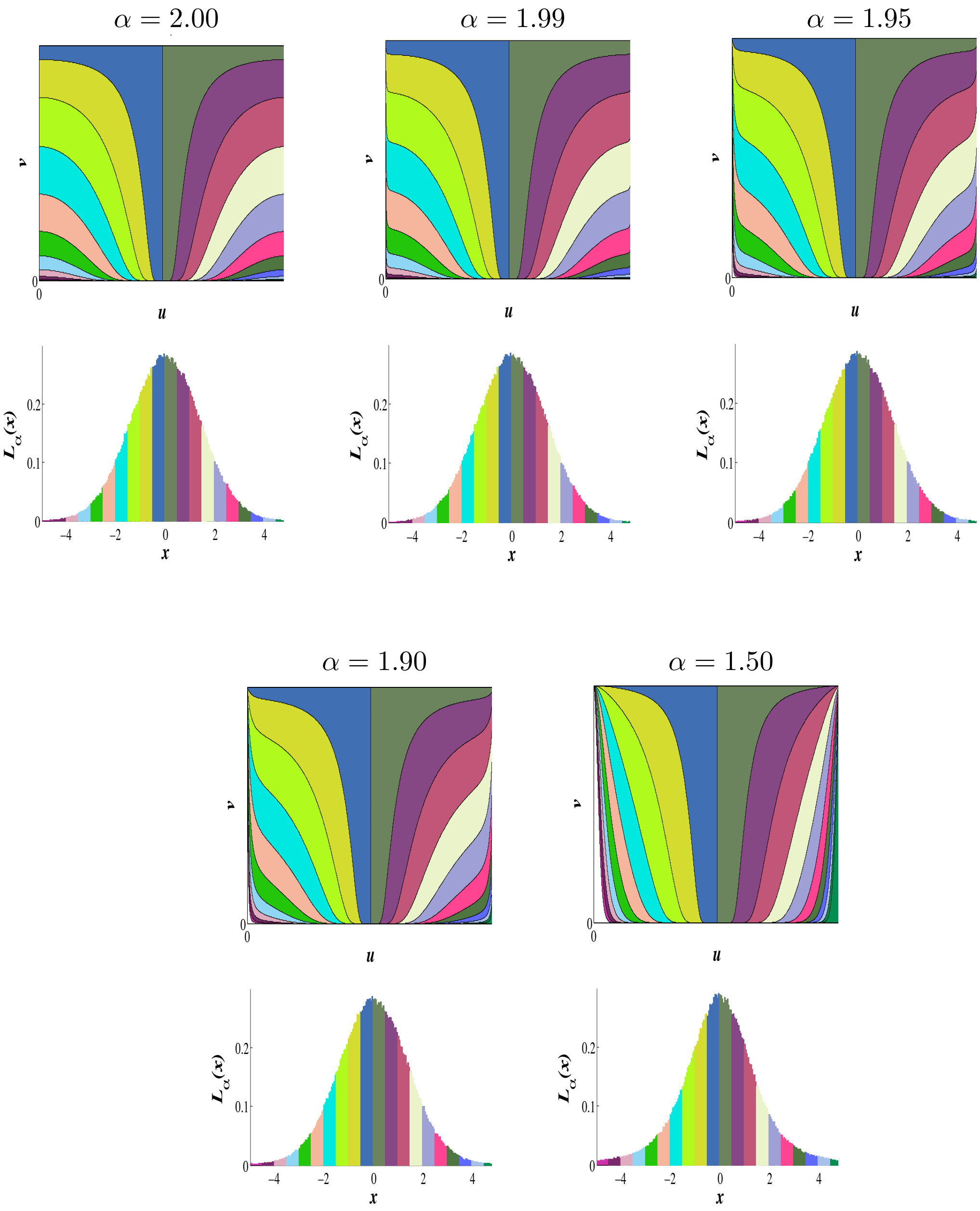
The variables $X_1, \ldots, X_N$ are stable as well as their normalized sum

$$X = \frac{1}{N^{1/\alpha}}\sum_{i=1}^{N} X_i. \tag{4.14}$$

This transform representation is a mixture of the form $g(V)W^{1-1/\alpha}$ where $g(V)$ is a real valued function, V is a uniform random number and W is exponentially distributed. Figs. 4.5 and 4.6 show symmetric and asymmetric examples of the mapping of the random number plane (U,V) to “quantiles” of the probability density via the map $X = F_{\alpha\beta}(U,V)$. Colors are used to designate the respective regions $x_1 < X < x_{i+1}$ separated by isolines defined by $\mathrm{d}F_{\alpha\beta}(U,V) = 0$. The pictures show isolines as borders between colors for $x_i = 0, \pm 0.5, \pm 1, \pm 1.5, \ldots$. The colors in the map and in the respective histogram correspond to each other and all points (U,V) on the same isoline are mapped onto exactly one unique number. Fig. 4.7 shows the behaviour of the isolines only with further decreasing α. Notable is the analytic Cauchy case $\alpha = 1$ whose inversion formula depends only on one variable. This is expressed by perfectly vertical isolines. For values of $\alpha < 1$ the overall behaviour turns over and the slopes change sign in each half of the unit square. The pictures showing isolines are produced with MATLAB’s [2] `contourf` function on a grid of size 800×800.

We would like to remark that different solutions are thinkable of how to sample uniform random points in a specific region in the (U,V)-plane. A differential equation for the isolines can be obtained via the implicit function theorem by Ulisse Dini [48]:

$$0 = \mathrm{d}F(u,v) = \frac{\partial F(u,v)}{\partial u}\mathrm{d}u + \frac{\partial F(u,v)}{\partial v}\mathrm{d}v\ ; \tag{4.15}$$


$\alpha = 2.00$

$\alpha = 1.99$

$\alpha = 1.95$

$\alpha = 1.90$

$\alpha = 1.50$

Figure 4.5: The map $X = F_\alpha(U, V)$ with $U, V \in (0, 1)$ giving the symmetric Lévy distribution $L_\alpha(x)$ for different values of α. For $\alpha = 2$ the picture corresponds to the Box-Muller map for the generation of Gaussian random numbers. The bottom part of each map shows the respective histogram. Areas with equal colors correspond to each other. Note that the transition from $\alpha = 2$ to $\alpha < 2$ is discontinuous for $u = 0$ and $u = 1$ and the points (0,1) and (1,1) develop a singularity.

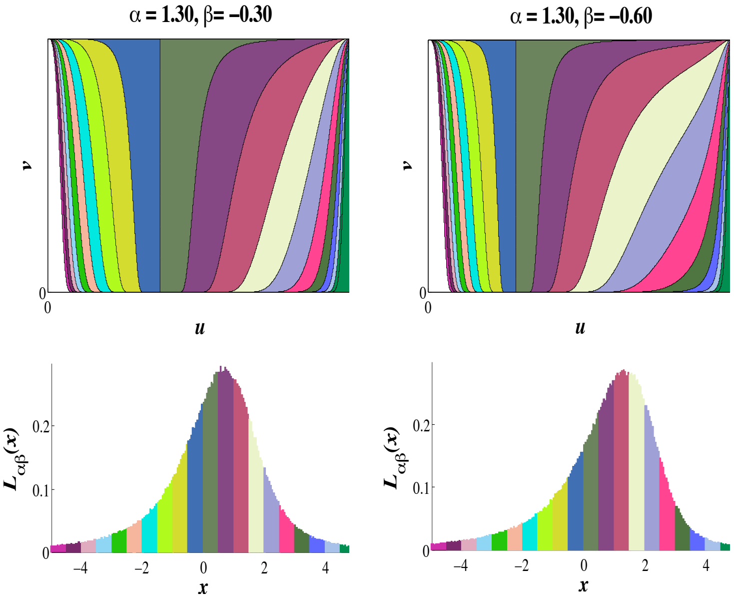


Figure 4.6: The map $X = F_{\alpha\beta}(u,v)$ giving the asymmetric Lévy distribution $L_{\alpha\beta}(x)$ for two values of β.

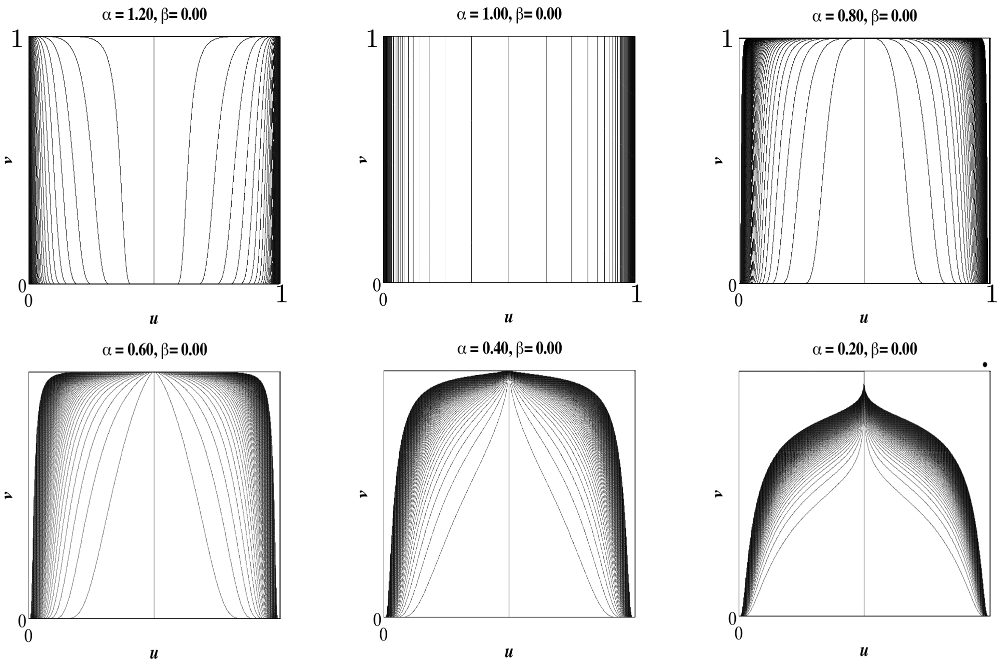


Figure 4.7: Isolines of the map $X = F_{\alpha}(u,v)$ with $u, v \in (0,1)$ for decreasing values of α. The regions with increasingly divergent gradient (upper corners) is not shown beyond $|x| > 600$. Note that the orientation of the isolines flips over with decreasing values of α at exactly $\alpha = 1$.

rearranging

$$\frac{\mathrm{d}v(u)}{\mathrm{d}u} = -\left(\frac{\partial F(u,v)}{\partial v}\right)^{-1} \frac{\partial F(u,v)}{\partial u}. \tag{4.16}$$

With an appropriate initial condition this differential equation defines the isoline $v(u)$ in the coordinate square spanned by u, v. The alternative representation of u as a function of v is equally appropriate from the mathematical point of view, but is less convenient in this case for symmetry reasons. We skip additional considerations on singularities and limiting behaviour. For $\alpha = 2$ and $\beta = 0$ Eq. (4.13) reduces to $X = F_2(U,V) = 2\sqrt{-\log U}\sin(\pi(V - 1/2))$, which is the Box-Muller method for Gaussian deviates with standard deviation $\sigma = \sqrt{2}$. The corresponding map is shown in the upper left of Fig. 4.5. The value x in the condition $X > x$ determines the initial condition for Eq. (4.16) that determines the isoline, i.e. x in the condition $X > x$, and for $\alpha = 2$ it can be chosen on the boundary of the square $U, V \in (0,1)$. Two other analytic limit cases for $\beta = 0$, where $L_\alpha(x)$ can be written in terms of elementary functions, are the Cauchy distribution, with $\alpha = 1$ and $X = F_1(U) = \tan(\pi(U - 1/2))$, and the Lévy distribution, with $\alpha = 1/2$ and $X = F_{1/2}(U,V) = -\tan(\pi(V - 1/2))/(2\log U\cos(\pi(V - 1/2)))$. Note that for values of $\alpha \neq 2$ the map F is singular in the points $(0,1)$ and $(1,1)$. In such cases the initial condition cannot be chosen on the boundary, which considerably complicates the situation numerically.

Starting from the simplest case, insertion of $F_2(u,v)$ into Eq. (4.16) yields

$$\frac{\mathrm{d}v(u)}{\mathrm{d}u} = \frac{\cot(\pi v(u))}{2\pi u\,\log(u)}. \tag{4.17}$$

Insertion of $F_\alpha(u,v)$ into Eq. (4.16) yields

$$\frac{\mathrm{d}v(u)}{\mathrm{d}u} = (\alpha - 1)\left\{\frac{1}{\pi u\log(u)}\left[\tan\left(\pi\left(v(u) - \frac{1}{2}\right)(1-\alpha)\right)(\alpha-1)^2 \right.\right. \\ \left.\left. + \cot(\pi v(u)) - \alpha^2\cot\left(\pi\left(v(u) - \frac{1}{2}\right)\alpha\right)\right]\right\}^{-1}. \tag{4.18}$$

One way to sample directly and uniformly from the area under $v(u)$ would be an area-preserving map of a square domain spanned by two uniform random numbers, e.g. $U, V \in (0,1)$, or any other suitable two dimensional domain onto this area. To our knowledge this solution is not available yet. Alternatively, the function $v(u)$ can be obtained numerically via integration or by appropriate algorithms for the generation of isolines. Once data points for $v(u)$ are obtained, any method that samples uniformly the region $X < x$ or $X > x$ is suitable in principle. With this, the generation of a tail variable constitutes in itself a standard non-uniform variate generation task. It is the initial scenario of sampling uniformly under a curve, but with the great simplification of a finite support. However, this is not the route we propose for three reasons. First, the numerical solution of Eq. (4.18) is cumbersome. Second, the initial condition has to be found *within* the u-v square due to the above mentioned singularities. The subsequent integration in two directions

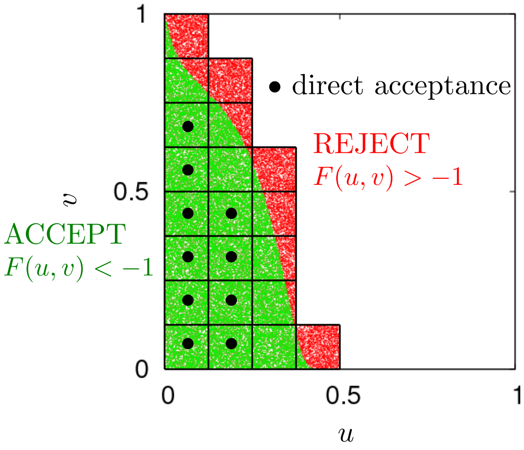


Figure 4.8: Intuitive, coarsely tiled, example of the tiling in the u-v square for sampling symmetric Lévy α-stable random variates with the condition $X = F(u,v) < -1$ and $\alpha = 1.8$, $\beta = 0$. The tiled area can be sampled efficiently while only points in the red shaded region are rejected. Tiles with direct acceptance do not require the acceptance comparison $X = F(u,v) < -1$.

must be guaranteed to work unattended and automatically as a black box with α and β as the only parameters. Third, the outcome is not *exact* in the sense that the sampled random tail variates are distributed with respect to an approximated probability density function based on the data point representation of the isoline. As it will turn out a numerical or analytic representation of the isoline is not a required piece of information and its calculation can be avoided. It can also be shown that the isolines are monotonic in u in the regions $F_{\alpha\beta}(u,v) < 0$ and $F_{\alpha\beta}(u,v) > 0$ which is a useful property in Sec. 4.2.3. Although the approximation of density functions is commonly accepted as a reasonable compromise in several applications we introduce in the next section a simple graphical method without this disadvantage.

4.2.3 Sampling method and example application

We introduce the method using simple intuitive examples. The production algorithm relies on the rejection method whose invention dates back to von Neumann [176] and which we do not rehearse here. Fig. 4.8 demonstrates a computationally efficient concept for uniform sampling in a certain two-dimensional region. In the first example we aim at producing Lévy α-stable random variates with parameters $\alpha = 1.8$, $\beta = 0$ and the condition $X < -1$. The map $F_{1.8}(u,v)$ for this choice of parameters is also shown in Fig. 4.5. It corresponds to a relatively large region in the left part of the square. We performed a straightforward and simple tiling of this region using square tiles that can be refined, for example, iteratively maintaining complete coverage while minimizing the excess area of the tiles that

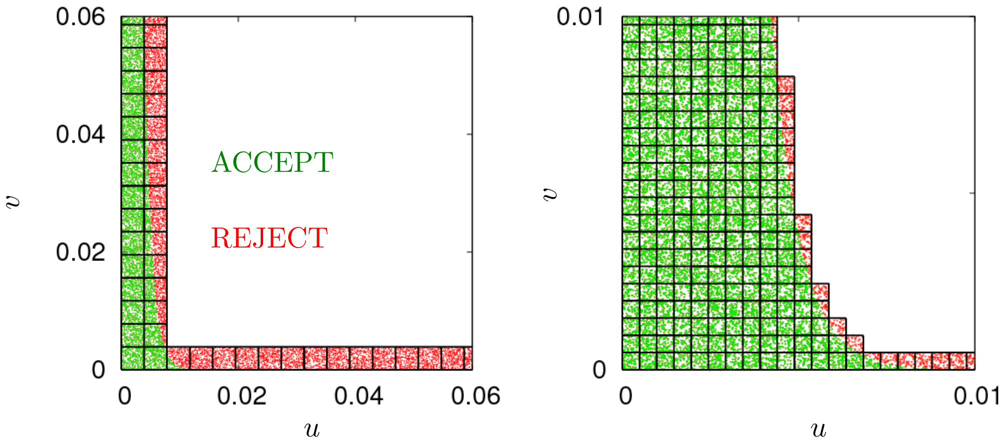


Figure 4.9: Two different tiling refinements of the region corresponding to the condition $X < -12$ which is a narrow strip along the left and bottom of the unit square. Only the lower left corner is shown on a scale that magnifies the tiling to a visible size. The row of tiles on the bottom samples a narrow strip below the isoline. In the right panel the rejection rate is significantly lower. It may help intuition that the colored dots are the uniformly distributed pairs (u, v).

stick out of the region defined by $F_{1.8}(u, v) < x$. Uniform sampling of the tiled area accepting all $X = F_{1.8}(u, v) < x$ and rejection of all other samples achieves the desired tail sampling. The size of the tiles can be chosen to achieve an arbitrarily low rejection rate. In the example shown in Fig. 4.8 the tiling is refined only moderately to convey the situation. For tiles that lie completely underneath the isoline the test $F_{1.8}(u, v) < x$ must not be executed. With dense tiling this comparison is therefore hardly needed and indeed must be avoided to yield a speed-up with respect to the transformation method. The production loop of random variates follows the steps:

```
Algorithm Input: $x \in \mathbb{R}$.
0 Setup:
    Tiling of the region $F(u,v) < x$ using a method of choice.
    Label tiles with an integer index.
    Label tiles that are intersected by the isoline $F(u,v) = x$.
1 Draw a random integer tile index with uniform probability.
2 Draw a random coordinate $(u,v)$ with uniform probability within this tile.
3 Test if the tile is intersected by the isoline (table look-up).
    If yes, go to 4. If no, accept $X = F(u,v)$ and go to 1 (direct acceptance).
4 Test if $(u,v)$ satisfies $X = F(u,v) < x$.
    If yes, accept $X$, otherwise reject and go to 1.
```

Note that for monotonic isolines the position of a tile with respect to an isoline, i.e. whether underneath, above or intersected, can be determined by evaluating the map for at most two corners. Step 4 is unlikely to be carried out if the coverage is dense, giving nearly a zero rejection rate. Overall, this procedure is efficient in setup and production for a sufficiently dense tiling. Furthermore, with small modifications of the above acceptance and rejection conditions in the pseudo code, the tiling and production of random numbers on a finite interval $X \in [x_1, x_2]$ is geometrically and algorithmically equivalent to generating numbers from the tail. This requires the tiling of a region in the *u*-*v* square between two isolines with the condition $x_1 < X < x_2$.

Fig. 4.9 shows the map for the left tail regions of the *u*-*v* square with $X < -12$, which is more realistic for the purpose of tail sampling. This condition corresponds to sampling a narrow strip at the bottom and left sides of the unit square. The figure only shows the corner at the origin. The bottom strip of tiles samples an extremely narrow strip than is not visible on this scale. The iterative tile refinement in the setup stage is acceptably fast, below a second in our non-optimized code, down to the level on the right panel of Fig. 4.9 to achieve a rejection rate below 1%. Different values of $\alpha > 0.1$ as well as not too extreme values of x have no significant influence on the setup performance achieving a rejection rate of 1%. Note that the speed of random number production is independent of the number of tiles. In our case it amounts to 2.3 million tail variates per second on a PC with a 2.4 GHz Intel Pentium 4 processor using the GNU C++ compiler version 3.2.2 and optimization level -O3. As the uniform random number generator we used the XOR shift SHR3 by Marsaglia [117].

The colouring of the acceptance and rejection regions in Figs. 4.8 and 4.9 are produced by green and red coloured dots representing the random uniform coordinates (u,v).

We would like to stress that the method of tiling as well as the form of the tiles is in principle arbitrary. Equal size and shape is computationally advantageous, but this issue is not the focus of the present work. Of course any tiling technique that produces a similar result is suitable, using either square or rectangular tiles. However, the choice of square equal tiles is algorithmically very simple and likely to outrun an adaptive scheme with more complex shapes in setup and also production. The iterative tiling refinement, as performed in the above examples, is robust and fast also for large values of $|x|$. The rejection scheme is in principle similar to the Ziggurat implementation of Ref. [117]. It also needs the setup of a data structure that covers a region by equal area rectangles. The details of the tiling method that is more general and applicable to random number production directly via the probability density are described in the previous section [61].

4.2.4 The Mittag-Leffler probability distribution

Our second example density is less know in scientific applications, even less so its transform. The Mittag-Leffler probability distribution appears in the analytic solution of the time-fractional Fokker-Planck equation [73]. The generalised Mittag-Leffler function is defined as [67, 74]

$$E_{\alpha\beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)}, \quad z \in \mathbb{C}. \tag{4.19}$$

For our purposes it is sufficient to restrict the example to the one-parameter Mittag-Leffler function which plays an important role in the stochastic solution of the time-fractional diffusion equation, see Eq. (2.37), leading to a pointwise representation on a finite interval. Fig. 4.10, top left plot, shows examples with two selected values of β. More details are given in Sec. 2.6. In many applications Mittag-Leffler random numbers where produced by rejection with a look-up table and interpolation of a pointwise representation obtained via Eq. (2.37) in a very cumbersome manner leading to a truncated distribution. Fig. 4.10 shows the map $M_\alpha(U, V)$ of the transform representation Eq. (2.47) as borders between intervals corresponding to $t = 0, \pm 0.5, \pm 1, \pm 1.5, \ldots$ The exponential case with $\alpha = 1$ depends on only one random variable in the u-v square which is expressed by perfectly horizontal isolines. For $\alpha < 1$ the left and right edges develop singularities. It is not recommended to use Eq. (2.37) and summation of many terms for the computation of $E_\alpha(-t^\alpha)$. A more elegant and accurate method is presented in Ref. [67, 74, 145]. For the generation of random numbers we use the implementation in Ref. [64].

4.3 Fast generation of rotationally invariant random matrices

As an extra topping we present a method that produces rotationally invariant matrices in one single step without the summation of rotated matrices as in Eq. (3.36). For the definition of such matrices, their properties and use within the theory of free probability refer to Chap. 3. Random rotationally invariant matrices are characterized by the following alternative pieces of information:

1) The distribution of the matrix entries
2) The distribution of the eigenvalues

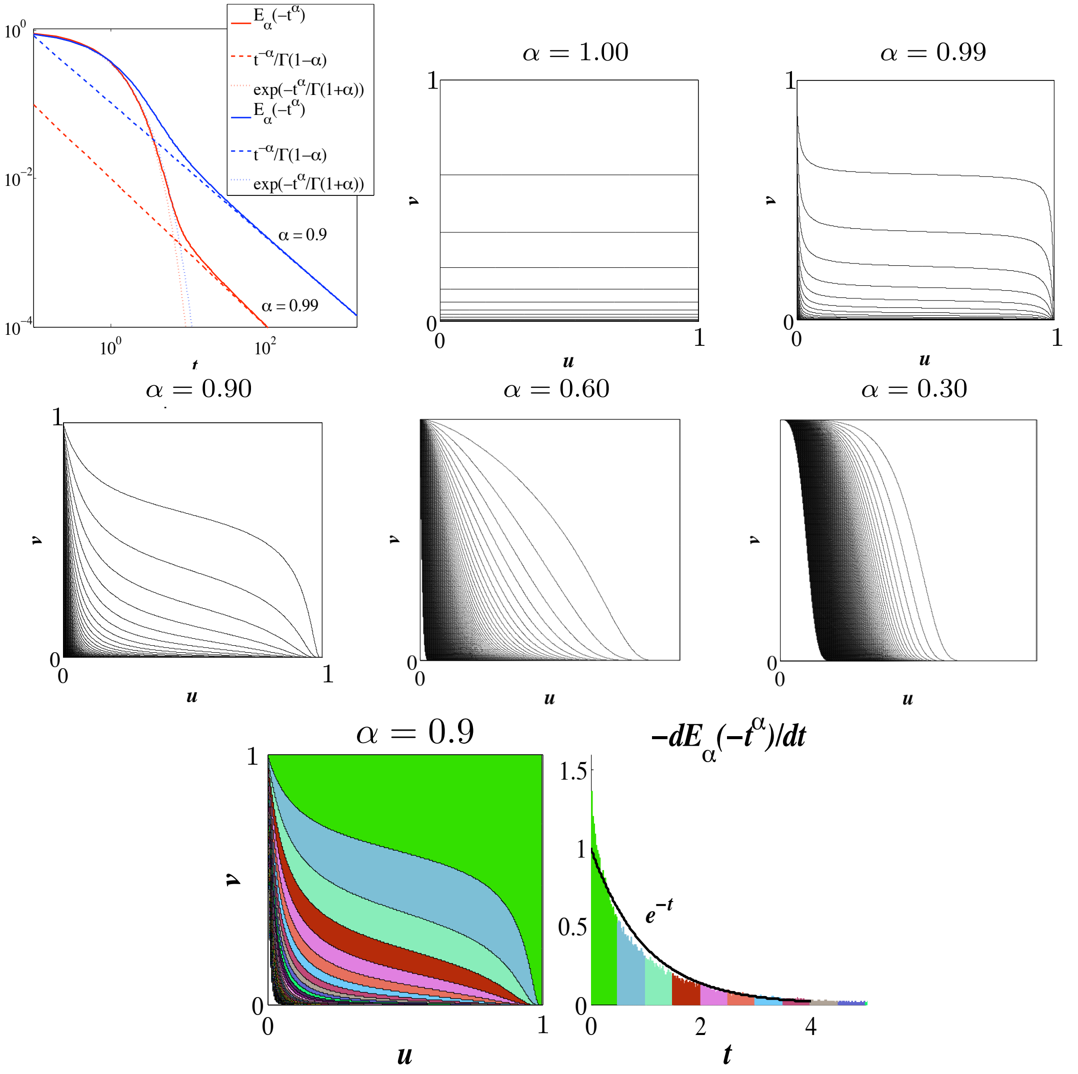


Figure 4.10: The Mittag-Leffler function $E_\alpha(-t^\alpha)$ in a log-log plot (top left) and the transformation map $X = M_\alpha(U, V)$, Eq. (2.47), in terms of isolines for five values of α. The case $\alpha = 1$ corresponds to the standard exponential function. The regions on the left side of the maps is not shown beyond $t > 600$ due to an increasingly divergent gradient. The two plots at the bottom repeat the case with $\alpha = 0.9$ using colors and showing the corresponding histogram.

So, everything is characterised by distributions only. If we choose one set of eigenvalues according to the desired distribution, the corresponding eigenvectors must obey a certain distribution, and vice versa. There is nothing that determines single numbers. Of course everything has to be seen in the limit of large matrices. We can write

$$D = XAX^{-1} \tag{4.20}$$

with $D = \mathrm{diag}(d_1, \ldots, d_n)$ the diagonal matrix of eigenvalues and X the eigenvector matrix. Then rearrange by multiplication with X from the right side and choose D according to the desired distribution and pug into

$$DX = B = XA \quad \text{where} \quad DX = [d_1 x_1, d_2 x_2, \ldots, d_n x_n]. \tag{4.21}$$

The eigenvectors columns in the matrix $X = [x_1, \ldots, x_n]$ can easily be chosen isotropic according to Porter-Thomas Eq. (8.13) or any other suitable isotropic distribution, which are all Gaussian in the end. Resolve $B = XA$ for A.

4.4 Discussion and conclusion

In this chapter we presented a fast method for automatic generation of random variates with arbitrary probability density functions independent of symmetry, number of modes, and discontinuities. The only prerequisites are pointwise computability and finite support. We also explained that the most general thinkable or universal method will require no less but also no more than these two requirements. In the introductory overview on some representative methods it is shown that many less powerful methods exist that truncate the infinite support for analytic density functions with only one mode. The accuracy of our method is exact up to the computation of the probability density function and meets any numerical demand which includes density functions with poles or cusps without additional attention.

The generation of one non-uniform random variate requires only one random integer, one random uniform real, two additions, one multiplication and one table look-up (no float comparison) most of the time. This is close to the minimum of principally required operations, so that additional speed can only come from hardware exploitation or specialized methods. Even for complicated density functions the memory requirements are suitable for any contemporary desktop computer.

These properties are not available in other methods of this kind. We can extend the wish list from Sec. 4.1.2 to a random number generator by additional items:

8. No need for a priori knowledge about the location of *any number* of modes or discontinuities within the density function.

9. Only pointwise computability and representability of the density function is necessary.

10. Fast setup time and fast generation of random variates.

11. The discretisation and therefore sampling efficiency is *asymptotically exact* and can be pre-imposed.

An extension to the multivariate case is simple in principle. It means to substitute a two-dimensional tile with a cube or hypercube. The required storage for data however increases with a power of the dimension.

We have also demonstrated some properties of the Chambers-Mallows-Stuck and Kozubowski-Rachev transform maps exemplifying the production of random numbers with the former. The interpretation as a two-dimensional map from the unit square to the real numbers allows to associate arbitrary intervals on the support of the density with well defined finite regions of the map domain. The uniform sampling of such regions produces directly random numbers exactly within the respective intervals. We have also introduced an efficient concept for the automatic setup of a random number generator that makes use of this property. The resulting generator can in principle produce random numbers in intervals which can be disconnected and any combination of the kind $(-\infty, x_1] \cup [x_2, x_3] \cup \ldots \cup [x_n, +\infty)$, $x_i \in \mathbb{R}$. Most importantly, the sampling of tails as shown here can be used as a tail handling method in fast implementations of random number generators for which a candidate is the Ziggurat method implementation that was proven to greatly outrun simple inversion methods. The present work opens the route for the speedup of many known random number generators that rely on transform representations.

While the application of rotationally invariant Lévy random matrices and in particular of realisations is still very seldom we are convinced that the realisation via summation of rotated matrices will soon be replaced by direct generation. Since the generation takes up significant CPU-time and freeness is approximated in the large size limit the compromises between matrix size and computer time will be less difficult to find with this new direct method.

4.5 Outlook

The connoisseur of random number generation will have realised by now that the method introduced in the last section is just the right one to perform fast sampling in the ratio of uniforms method for densities that are not a simple teardrop shape. The principle of this method is to use the ratio of two uniform variates and sample within a certain region of the unit square [82], see Fig. 4.11. Suppose p is a (not necessarily normalised) density and U, V are uniform on $0 \leq u \ \leq \sqrt{p(v/u)}$ then $X = V/U$ has the density (proportional to) p. The procedure to set up a generator is:

a) Construct in the unit square (u, v) a plane bounded by $0 \leq u \ \leq \sqrt{p(v/u)}$.
b) Draw uniform (U, V) in this bounded plane.
c) Return V/U as the deviate which then will obey p.

Depending on the desired density and concavity properties the tiling of more complicated shapes is a field on its own. For concave or piecewise concave densities it is trivial and the sampling reaches the full infinite support. Radially concave

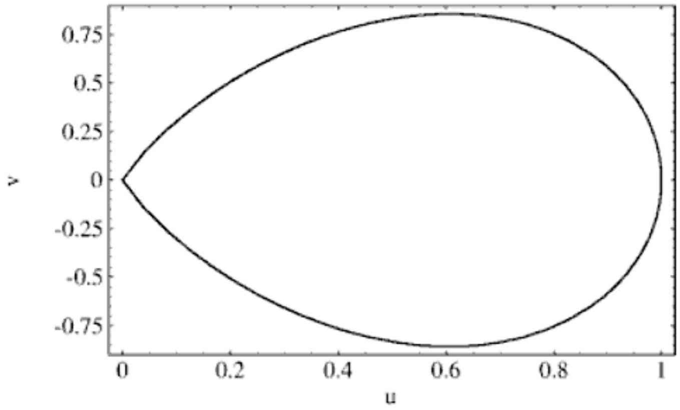


Figure 4.11: Acceptance region in the ratio of uniforms method for normal random variables. The picture is taken from Ref. [149]

shapes, i.e. regions that can be defined via a functional relation $p(\theta, r)$ from an origin, are also simple to tile and encompass a large class of density functions.

Chapter 5

Comparison of the Fourier and Pearson correlation estimators

> *If you want to inspire confidence, give plenty of statistics. It does not matter that they should be accurate, or even intelligible, as long as there is enough of them.*
> *L.C.*

The work in this chapter was supported by the British Council Researcher Exchange Programme 2008 for stays of D.F. at the Department of Economics at the City University London. This part mainly compiles results and arguments but also states a few conclusions that came up in the attempt to settle certain issues that caused debates as well as persistent disagreements between the researchers involved in this topic. The focus of research are the Pearson and Fourier correlation estimators, the latter of which is very new and hardly known. The benchmark using artificial continuous-time random walks creates a scenario very close to high-frequency financial data. So far, the influence of a power-tailed distribution of waiting-times seems to be negligible in the context of correlation matrix analysis with meaningful time series, i.e. not too extreme distributions. It turns out that the Fourier estimator cannot cope better with waiting times, at least in reconstructing the Marčenko-Pastur spectrum for the matrix of correlation coefficients. The influence of power-tailed jumps, however, is clearly visible in the spectrum as well as in the distribution of the largest eigenvalue. In particular the latter is of importance since confidence tests are based on the Tracy-Widom theory of the distribution of the largest eigenvalue in the uncorrelated case. This theory associates a probability for the largest eigenvalue in de-trended correlated real world data. The Tracy-Widom theory, however, is grounded entirely in the Gaussian world of finite jump moments and synchronous (or equi-spaced and equivalently exponential) updates.

5.1 Introduction

The notion of correlation is mathematically well defined with respect to its estimator. For time series or other pairs of real valued data that are synchronous the measurements can be interpreted as realisations of random variables. The Pearson correlation estimator for two real random variables $X, Y \in \mathbb{R}$ with zero mean, unit variance and realisations $X_i, Y_i, i = 1, \ldots, N$ is given by

$$\frac{1}{N-1}\sum_{i=1}^{N} X_i Y_i \tag{5.1}$$

and is the most common estimator for the calculation of correlation. Usually N is exceedingly large and the subtraction of 1 is omitted. The resulting value is mathematically well defined while the value obtained from real world data may only be considered as an approximation of correlation. The estimation is asymptotically unbiased only for those discrete processes that can be produced by a thinning of a continuous Wiener process. Asynchronously updated continuous-time random walks (CTRWs) are discrete processes that take new values at arbitrary continuous times and the term correlation becomes ambiguous. Refer to Sec. 2.2 for the definition of CTRWs. In data analysis the procedures to extract, at least, a similar information via the Pearson correlation estimator must consider the application and model assumptions about the system. Typical procedures are several types of interpolation to obtain synchronous values between the data pairs. The most common is previous tick interpolation usually carried out on an equi-distant grid of sampling times: The sampled increments $\Delta x_{ij} = x_{ij} - x_{i,j-1} \equiv x_i(t_j) - x_i(t_{j-1})$ are defined as the difference taken between two consecutive sampling times t_j and t_{j-1}. Assuming that the averages and means of the increments are zero, the Pearson estimator for the covariance of two CTRWs i and k with unit variance is

$$c_{ik} = \frac{1}{T}\sum_{j=1}^{T} \Delta x_{ij}\Delta x_{kj}. \tag{5.2}$$

In the hydrodynamic limit of large time scales and finite moments the result may approach the Wiener process situation. However, this limit is often not a reasonable assumption as, for example, in high-frequency financial data analysis. It is clear that the mechanism and system state between the updates of stock prices is not expected to be sufficiently characterised by any kind of interpolation of the values. In fact, interpolation even makes assumptions on the system state where no information is available, and where the state refers to a model system. Correlation estimators that rely on interpolated values are expected to give different results. Whether this can be called a bias is unclear before the true correlation is mathematically well defined.

An alternative to methods which explicitly interpolate is the Fourier estimator by Malliavin and Mancino [11, 109, 151] that does not require equi-spaced pairs of data. If a process $x_i(t),\ t \in \mathbb{R}$ is given at non-evenly spaced times $t_1, \ldots, t_N$ the correlation coefficient can be calculated via the Fourier coefficients that are obtained with the following formulas:

$$a_{ki} = \frac{x_i(2\pi) - x_i(0)}{\pi} + \frac{1}{\pi}\sum_{n=1}^{N} x_i(t_{n-1})(\cos(kt_{n-1}) - \cos(kt_n)), \tag{5.3}$$

$$b_{ki} = \frac{1}{\pi}\sum_{n=1}^{N} x_i(t_n)(\sin(kt_{n-1}) - \sin(kt_n)). \tag{5.4}$$

The data is assumed to be rescaled to 2π. The Fourier coefficient of the pointwise covariance matrix is given by

$$a_{ij} = \frac{\pi\tau}{T}\sum_{k=1}^{T/2\tau}(a_{ki}a_{kj} + b_{ki}b_{kj}). \tag{5.5}$$

Since we potentially have infinitely short waiting-times the coefficient above is only exact in the limit $\tau \to 0$. In reality we have a finite τ and $T/2\tau$ is the highest meaningful wave harmonic. The desired correlation matrix elements for the entire data

$$c_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}} \tag{5.6}$$

contain the integrated value of the pointwise covariance matrix

$$\sigma_{ij}^2 = 2\pi a_{ij}. \tag{5.7}$$

The Fourier method is hoped to give under certain circumstances a less noisy or less biased estimation of the same coefficient value as compared to the Person estimator. While the latter is meaningful only on the basis of the (correlation) model assumption, i.e. the specific mechanism of price interaction, it is necessary to study the properties of the Pearson and Fourier estimators quantitatively also under realistic circumstances since correlation analysis is a common tool in many types of data analysis. Such analysis was presented in [146, 147] for pairs of stock market data and we are going to study the spectral properties of, and in particular, the differences between correlation matrices build with the Pearson and Fourier estimators. If the Fourier estimator gives systematically different results in the analysis of certain types of data, then in such a case the choice between the two will have to be made on grounds of purpose, on the model assumption and on the degree to which the relevant features are captured best. Note that the Fourier estimator also interpolates, i.e. it makes assumptions on the behaviour of a postulated price correlation system in the regions between the data points.

5.2 The context of correlation matrices

In the context of correlation matrix analysis it is often the spectrum of the uncorrelated situation that is considered the null-hypothesis benchmark. The distribution of jumps and waiting-times of the benchmark time series have finite moments.

Whether this assumption is realistic is a different issue. In any case, the Pearson estimator is used also for high-frequency financial data where power-tails can be observed in the log-price changes. For the above finite moment null-hypothesis assumption and the Pearson correlation estimator the Marčenko-Pastur theory [118] delivers an analytic eigenvalue density of the correlation matrix in the large size limit. In a comparison of the Pearson and Fourier estimators two questions arise: a) How does the Fourier estimator reproduce the Marčenko-Pastur law? b) How are correlations captured in terms of large eigenvalues and respective eigenvectors? These two questions can be considered first in a more general and clean context of CTRWs with different distributions of jumps and waiting-times. Our choice of Lévy α-stable distributed jumps and Mittag-Leffler distributed waiting-times is motivated by empirical studies of high-frequency financial data [150] and by the possibly convenient connection to the Monte Carlo solution of the space-time fractional diffusion equation [62, 73]. The Wiener process and the standard diffusion equation are recovered as a limiting case. Furthermore, it is necessary to analyse how true correlations are captured in terms of eigenvalues and eigenvectors of the correlation matrix. Both estimators treated here are not well defined and are at least biased for processes other than thinnings of a Wiener process. The justification for comparing the estimators on processes that are more general is simply the contemporary common practice [137] due to the lack of alternatives. Secondly, it is only the outcome in a particular application, e.g. portfolio optimisation, that determines the choice of an algorithm for the respective data analysis and prediction models. The latter is a different issue as well.

Overall, we are confronted with a situation that requires at least a hundred time series to provide a histogram of eigenvalues with sufficient accuracy to constitute a spectrum. A meaningful correlation matrix analysis requires data of a minimum length and information content. We have explained above at length, the number of data pairs and information content are not synonymous in the context of CTRWs. A correlation matrix that does not have full rank because of spurious correlation due to missing values caused by large waiting-times is not obvious to interpret and not in the realm of the Marčenko-Pastur theory either. Some suspect the summation of many zeros in the estimator Eq. (5.2) for the correlation coefficients as a source of differences or bias. These zeros are produced by previous tick interpolation which gives a zero increment if no update occurred within $\Delta t = t_j - t_{j-1}$. In statistics, this situation is known as one of the many so called "missing values" scenarios in covariance matrix estimation [104, 168]. As long as these values are missing in a random manner, the estimator for the covariance (thus, the correlation) is unbiased and we should expect the same result, possibly with some increased statistical error. This assumption, however, is only true for the null-hypothesis. We could therefore expect that the uncorrelated test case with completely decoupled jumps and waiting-times is less likely to show a true difference between the Pearson and Fourier spectrum. In the previous section it was explained that the distribution of coefficients and spectra of correlation matrices are nearly independent of the jump density if it has finite moments. It is only the density's power-tail that is relevant; also see Ref. [16]. In other words, any meaningful distribution of jumps and waiting-times that has finite moments yields the same result in terms of the spectrum of the respective correlation matrix. Therefore we expect the Pearson and Fourier correlation matrices to be less sensitive to missing values if the data has finite moments as compared to data with power-tails. This statement is even stronger for increasingly larger waiting-times which only provoke more missing values. On the other hand, large jumps in data are the typical nuisance for a Fourier transformation which requires at least a linear de-trending between the end points of a time series. In some cases this measure severely distorts the data. Indeed, we observe differences between the Pearson and Fourier spectra with the introduction of power-tailed jumps. Refer to formula Eq. (5.3) in which the first term contains an assumption on the behaviour of the data between the first and the last data point.

For artificially correlated benchmark cases we could expect stronger differences between the Pearson and Fourier estimators possibly also with the introduction of (power-tailed) long waiting-times. An open question is how to introduce test-correlations since it is unknown how prices are correlated due to a lack of a reasonable model. For other generic real world data there will be no model anyway. However, this question is obsolete if one cannot distinguish a real and artificial situation on the basis of *quantified* stylised facts of the time series. This quantification is often rather open in the respective community and we can introduce simple correlation that mimic stock prices for example. The latter case is particularly simple because the correlation is typically very weak and here we only consider equal time correlation. Stylised facts like volatility clustering require more sophisticated faking of data.

Since there is no generic correlated or uncorrelated case, several scenarios can be considered in the attempt to provide information on the behaviour of the Pearson and Fourier estimators in realistic applications. The most associative scenario is possibly, again, high-frequency stock market data:

- Uncorrelated CTRWs with equal distribution of jumps and waiting-times.
- Uncorrelated CTRWs with inhomogeneous distribution of jumps and waiting-times and also scaling factors.
- One equally correlated set of CTRWs in a sea of uncorrelated CTRWs. The correlated set is updated synchronously but not equi-spaced.
- As above but the correlated set is updated asynchronously. In this case the correlation must be introduced in a reasonable way which is not necessarily mathematically well defined.
- The last two items can be repeated with inhomogeneous distribution of jumps, waiting-times and scaling factors.

The above items are carried out in one way or another in different chapters of this document, not necessarily for comparing estimators. It can be questioned whether the introduction of inhomogeneous distribution of jumps and waiting-times and also scaling factors just smoothes out differences between the two estimators or if this is actually the key issue that provokes the differences. In finite sets of data it is not necessarily obvious to recover a posteriori the scale factor and the power-tail parameter from the data points for each individual time series. The correlation matrix spectrum would be the most insensitive tool to test for a mixture of scales and power-tails. We also have to set restrictions on the distribution of waiting-times. It turns out that a too fat tailed density too often produces waiting-times that are longer than the entire length of the CTRW, see Fig. 5.7. This situation would be of limited meaning.

A realistic inhomogeneity of scales and power-tails as found in financial data is achieved by applying specific distributions of the parameters $\alpha, \beta, \gamma_x, \gamma_t$ which were measured [150, 155] before.

5.3 Preliminary numerical results

The data model presented in Sec. 2.2 gives four degrees of freedom to vary the characteristics of the time series. It is important to study the influence of each parameter separately. The initial situation from which parameters are varied is the unscaled Normal Compound Poisson Process (NCPP) that is given with the parameters $\alpha = 2,\ \beta = 1,\ \gamma_t = \gamma_x = 1$.

Alpha variations with uncorrelated CTRWs

The most promising parameter to start with that might provoke a difference between the two estimators is α which governs the power-tail of the jump density. Fig. 5.1 shows the averaged realised spectrum of the correlation matrix of uncorrelated CTRWs produced with $\alpha = 1.4$. We can observe a developing discrepancy at the left end and around the middle of the spectrum. A more pronounced situation with $\alpha = 1.0$ in Fig. 5.2 clarifies the trend: The tails are essentially unchanged while the peak at $\lambda = 1$ is more pronounced for the Pearson estimator. Since the cumulation of eigenvalues at $\lambda = 1$ is the limiting situation for long uncorrelated time series one could conclude that the Pearson estimator reflects the uncorrelated situation better. This conclusion, however, is shortsighted because it is the left and, most importantly, the right tails of the spectrum that are usually indicators of correlation, whether random or not. In this example we are confronted with a more subtle effect that leaves the right tail seemingly unchanged and is most expressed in the bulk.

Since it is of more importance in correlation analysis we investigate the right tail more thoroughly for the test case with $\alpha = 1.4$. Figs. 5.3 and 5.4 show the histogram of the largest eigenvalue unscaled and in a log-log plot respectively. The difference we observe here must be considered in the specific context and are not necessarily significant. The log-log plot shows that both estimators reproduce a power-tail in the density of the largest eigenvalue but with slightly different tail index.

Beta variations with uncorrelated CTRWs

In previous studies we already investigated the spectrum of correlation matrices of uncorrelated CTRWs with power-tailed waiting-times. We observed a very small deviation almost entirely at the right side of the Marčenko-Pastur law in form of a small decaying tail instead of a hard edge. For the same set of CTRW

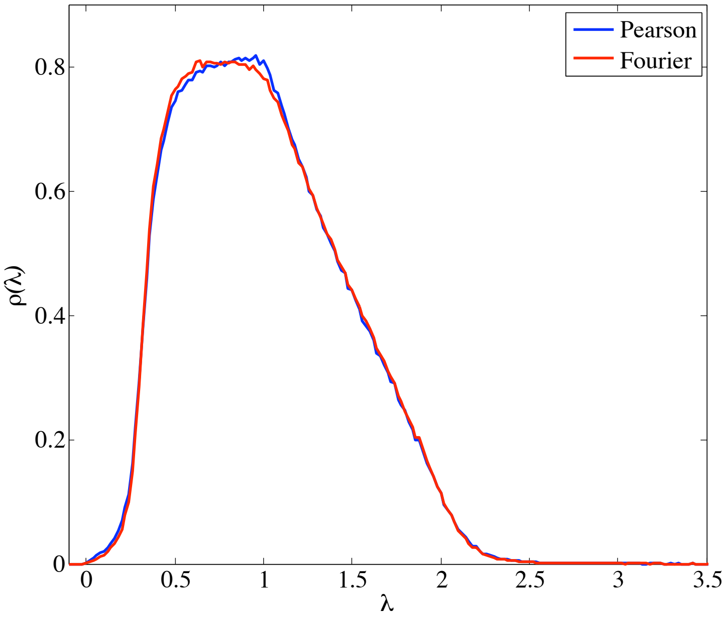


Figure 5.1: Averaged spectra of 9400 correlation matrices of size $N = 100$, $T = 500$, $\alpha = 1.4$, $\beta = 1.0$.

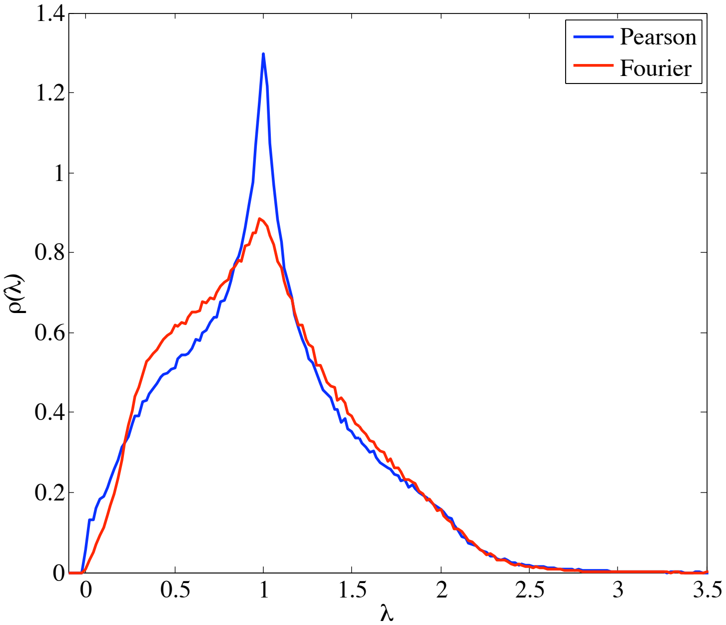


Figure 5.2: Averaged spectra of 2600 correlation matrices of size $N = 100$, $T = 500$, $\alpha = 1.0$, $\beta = 1.0$.

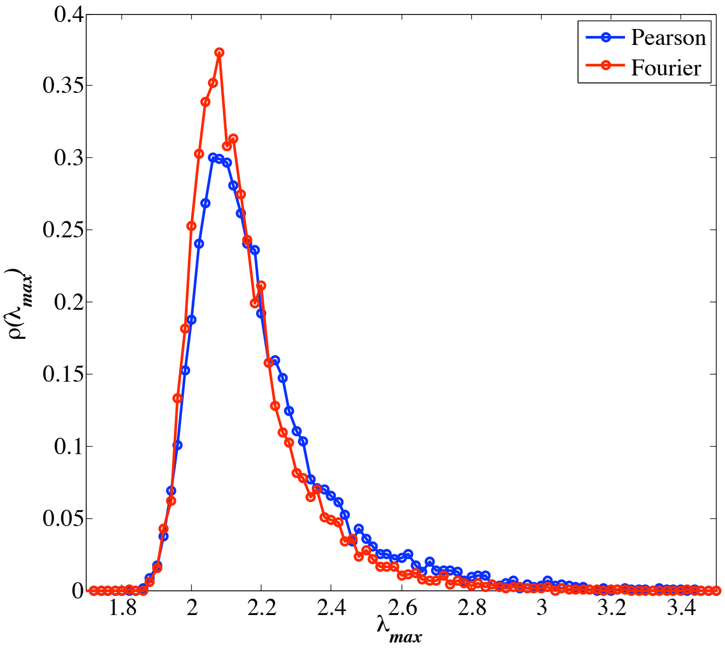


Figure 5.3: Histogram of the largest eigenvalues of 19410 correlation matrices of size $N = 100$, $T = 500$, $\alpha = 1.4$, $\beta = 1.0$.

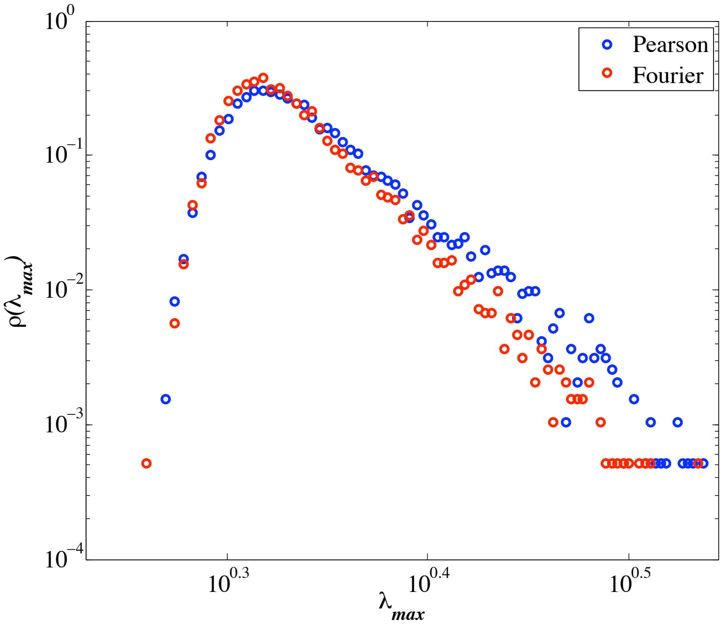


Figure 5.4: Histogram of the largest eigenvalues of 19410 correlation matrices of size $N = 100$, $T = 500$, $\alpha = 1.4$, $\beta = 1.0$. The logarithmic scale indicates that both estimators reproduce a power law distribution for the tails but with a different index.

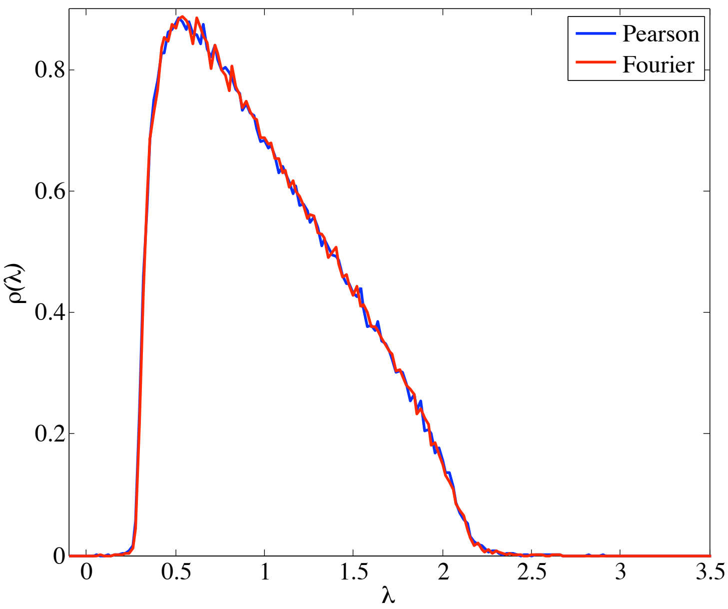


Figure 5.5: Averaged spectra of 900 correlation matrices of size $N = 100$, $T = 500$, $\alpha = 2.0$, $\beta = 0.9$.

parameters the Fourier estimator gives the same result as the Pearson estimator, see Fig. 5.5. The respective CTRWs are shown in Fig. 5.6 to give an impression of the situation. The choice of $\beta = 0.9$ is also close to the lower limit of a meaningful distribution of waiting-times. Lower values are likely to produce waiting-times that are longer than the entire CTRW and in some cases no update occurs. A frequent a posteriori removal of such cases would bias the statistics. It remains to be seen if an inhomogeneous distribution of power-tail indices as well as scale factors provokes a stronger difference than just one choice of β (and α).

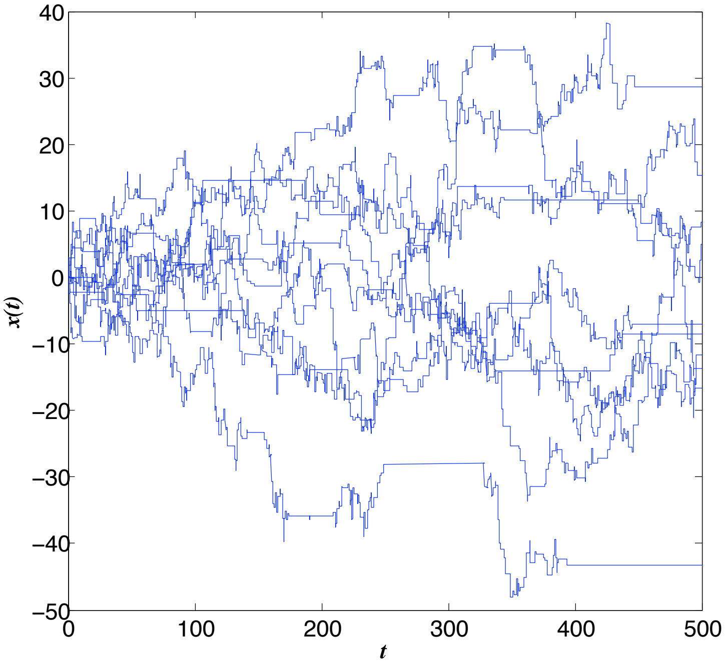


Figure 5.6: Samples of CTRWs of length $T = 500$, with parameters $\alpha = 2.0$ and $\beta = 0.9$.

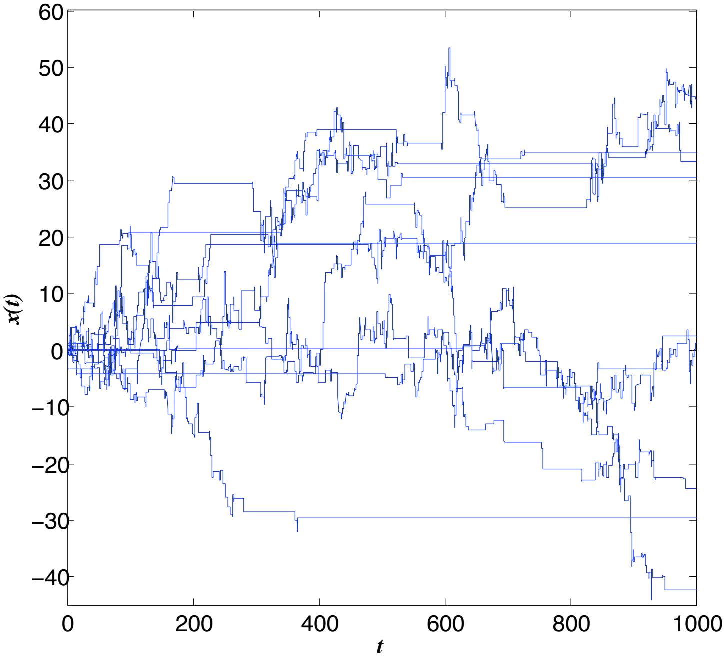


Figure 5.7: Samples of CTRWs of length $T = 1000$, with parameters $\alpha = 2.0$ and $\beta = 0.8$. It can be seen that the probability for updates is so low that in many cases no update falls within 1000 average waiting-times of an NCPP.

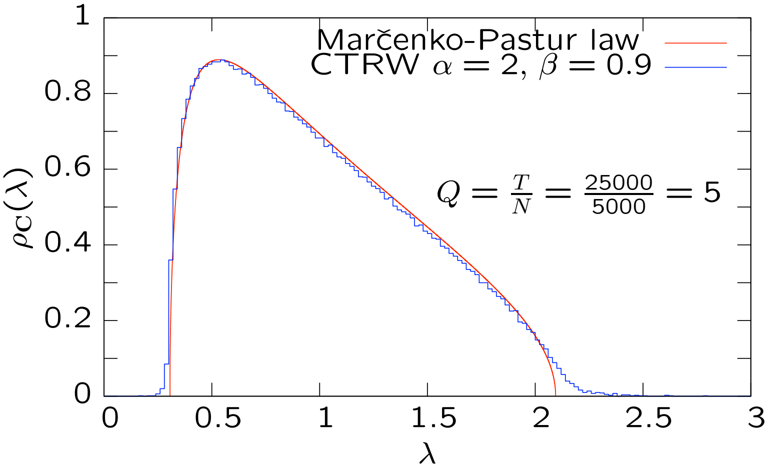


Figure 5.8: Spectrum of a correlation matrix from uncorrelated CTRWs with Gaussian distribution of jumps and Mittag-Leffler distribution of waiting-times ($\beta = 0.9$). The sampling rate for the Pearson correlation estimator is 1 in units of the average waiting-time for the exponential function defined by $\beta = 1$. The number of CTRWs is chosen large to exclude any finite size effect or coincidental large waiting-time within only few time series. Furthermore, the spectrum is averaged over 10 realisations to ensure statistical accuracy.

5.4 Matlab code for the Fourier correlation estimator

The following function pre-calculates along with the CTRWs all Fourier coefficients. With these the subsequent pair-wise correlation speeds up dramatically by avoiding redundant Fourier transformations.

```
function [M,XI,TAU,X,TIME,aa,bb] = makectrws(T,N,delta_t,alpha,beta,gamma_x,gamma_t,den)

M=inf(N,floor(T/delta_t)); % PTI increment matrix
XI=[]; X=[]; TAU=[]; TIME=[]; aa=[]; bb=[];

k_max=floor(2*pi/den);
im_unit=sqrt(-1);

for n=1:N

  x=[];  xi=[];  t=[]; tau=[];
  %Must avoid too few increments leading to singular matrices:
  while (length(tau)<4)
     [x,xi,t,tau]=ctrw(T,beta,gamma_t,alpha,gamma_x); %return continuous-time random walk
  end

  XI(n).xi=xi;
  X(n).x=x;
  TAU(n).tau=tau;
  TIME(n).t=t;
  [tpti,xipti,xpti] = pti(t,x,delta_t,T); %previous tick interpolation for re-use in main
  M(n,:)= xipti;
```

```
  times=2*pi*TIME(n).t/T;
  co=cos(times'*(1:k_max)/2);
  si=sin(times'*(1:k_max)/2);
  a = (XI(n).xi*co)/pi;
  b = (XI(n).xi*si)/pi;
  a=[ flipdim(a,1) 0 a]; %positive and negative frequencies
  b=[-flipdim(b,1) 0 b];
  aa(n).a=a;
  bb(n).b=b;

end
%-- end of function makectrws

------------ main Program:

....

   den=gamma_t*2*pi/T; % wave length limit
   [M,XI,TAU,X,TIME,aa,bb] =  makectrws(T,N,delta_t,alpha,beta,gamma_x,gamma_t,den);

   disp('C Fourier..')
   for i=1:N
       for j=i:N
           corre=aa(i).a*aa(j).a'+bb(i).b*bb(j).b';
           c=2*pi*pi*corre/(2*k_max+1);
           CFourier(i,j) = c;
           CFourier(j,i) = CFourier(i,j);
       end
   end
   for i=1:N
       dia=sqrt(CFourier(i,i));
       CFourier(i,:) = CFourier(i,:)/dia;
       CFourier(:,i) = CFourier(:,i)/dia;
   end

....
```

Thanks to Mauro without whom this code would have taken longer to produce.

5.5 Conclusion

For the scenario of spectral correlation matrix analysis with uncorrelated date the Pearson and Fourier correlation estimators are equivalent for data with finite moment jump distributions and fat-tailed asynchronous waiting-times. With jumps of power-tailed distribution differences appear but mostly in the bulk rather than in the tails of the spectrum. It is the tails that are considered as sensitive for signatures of departure from the random uncorrelated case. This makes comparisons in real world situations with only one realisation or data set difficult since only one realisation of the empirical spectrum is available.

Chapter 6

Correlation matrices of artificial continuous time-random walks and empirical data

The principle of correlation measurement with time series data exhibiting stylised facts as introduced in Sec. 2.2 has been described in Chap. 5, in particular the previous tick interpolation. Here we show how CTRWs appear to be a candidate for high-frequency stock market data at least with respect to the reproduction of the theoretical Marčenko-Pastur spectrum for uncorrelated noise (jumps) with finite moments. Along with this uncorrelated test-case an artificial correlation experiment provides some insight into the behaviour of the eigenvectors which seems not to have been recognised yet.

6.1 Eigenvectors of correlation matrices

Fig. 6.1 compares the spectrum of the correlation matrix built from CTRWs containing exponential waiting-times to the Marčenko-Pastur theory for uncorrelated noise with finite moments. The noise (increments) is Gaussian. Exponential waiting-times are expected to give the equivalent result as equi-spaced random walks on a grid of size 1 in units of the mean waiting time. The mean waiting-time in the exponential case is indeed 1. Finite size effects are negligible in this example with 500 CTRWs. The arbitrary choice of length T is measured units of the sampling grid which in turn is chosen equivalent to the expectation value of the mean waiting-time distribution, $\mathbb{E}[\xi \sim \exp(-t)] = 1$. It will be elaborated to a great detail in Chap. 7 that the waiting-time distribution plays no significant role with this choice of scales and sizes. In other words, in this section and generally in the context of correlation matrix analysis we are in the hydrodynamic limit. The results do not change with increasing size parameters N and T.

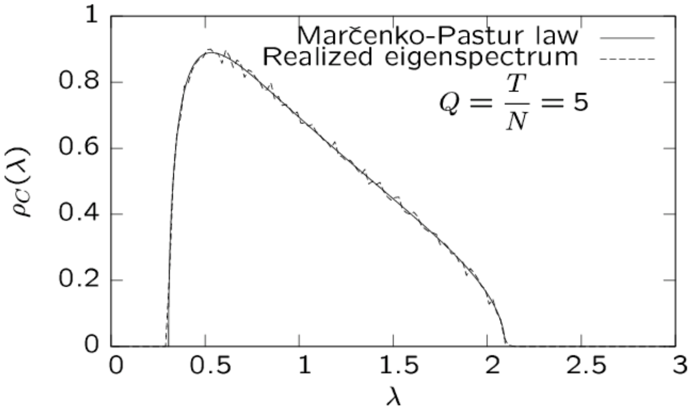


Figure 6.1: The Marčenko-Pastur spectrum for uncorrelated noise with finite moments is reproduced perfectly with CTRWs containing exponential waiting-times. The empirical spectrum is averaged over 60 matrices of size 500×500.

Of particular interest in our setting of time series analysis are the eigenvectors of the correlation matrix which code information about the contained "modes". This term is not uniquely defined and its meaning depends on the measurement instrument or "sensor" with which the coefficients in the coefficient matrix are calculated. With the Pearson correlation coefficient a mode can often be recognized by eye as a co-movement of the time series. We will use this for a simple toy example:

Toy data model with controlled artificial correlations

- $N = 500$ CTRWs with exponential waiting-times with average 1
- One artificial correlated homogeneous group of 50 CTRWs
- $x^*(t)$ is a prefixed CTRW where $\max(t) = T$.
- Type of correlation:

$$x_n(t) = x_n(t)(1-c) + x^*(t)c, \quad n = 1, \ldots, 50, \quad t \in \mathbb{R} \tag{6.1}$$

  and $c \in [0, 1]$ is a correlation coefficient.
- $c = 0.4$

This scenario uses an arbitrary method to correlate a group of CTRWs. Anything that somehow simulates an attractive force among CTRWs is completely sufficient for this type of study. It is not necessary to associate the parameter c with the mathematical correlation coefficient. In this case it is the integrated values of the random variables (jumps) that is correlated. This leads to a group of CTRWs staying together indefinitely. In a later chapter toy examples are shown where the jumps are correlated. Both methods are equivalent with respect to the respective effects and objects of interest.

Fig. 6.2 shows at a glance in one figure the main features of the result. Panel A shows the correlated group of CTRWs in red within a background of green uncorrelated CTRWs. The group of 50 correlated CTRWs are correctly reflected in a small heap of large correlation coefficients around 0.4. The correlation strength is chosen such that the group produces an eigenvalue of the order of a typical market mode eigenvalue. Panels B to F mark single selected eigenvalues whose eigenvector is shown graphically below.

Panel B starts with the (largest) eigenvalue provoked by the correlated group. As expected, the first 50 elements of the eigenvector are equally large while the other elements are essentially zero. Panel C: We elaborate in greater detail in Chap. 8 that the smallest eigenvalue's vector also contains nonrandom structure. The first 50 elements clearly stick out by high variance but are seemingly still Gaussian as all other elements in the indices of uncorrelated CTRWs. Panel C: Moving to the right, just inside the Marčenko-Pastur spectrum no structure can be observed. D & E: Closer to the right edge we recover again non-random structure, i.e. low variance in the indices of the correlated CTRWs. This finding is particularly interesting in C because the common sense in literature does not mention this but states that everything inside the Marčenko-Pastur part does not carry information. This is of relevance in publications like Ref. [171] where the Marčenko-Pastur part is used to filter out the noise part from a data set of stock price time series to extract the relevant dynamics. Here we show that the subtraction of the entire noise-part can be harmful. The procedure of mode reconstruction is rehearsed in detailed in Chap. 8 and promises a way out.

This issue of the meaning of eigenvectors and filtering with their help is also criticised in the following analysis of real-world data. Fig. 6.3 shows the result using one month of 1000 stocks from the New York Stock exchange. The largest eigenvalue and shape reproduces a typical spectrum for this type of analysis [88]. The sampling is $\Delta t = 100$ seconds such that possible power-tails do not show up due to the inherent truncation of price jumps in the trading mechanism. The market mode coded in the largest eigenvalue can be recognised by eye in Fig. 6.4. The overall downtrend is very faint but since most stocks take part the resulting eigenvalue is large. The market mode is commonly understood to be some empirically observed co-movement of the entire market. At least it is thought so. Fig. 6.5 shows in addition to the eigenvector of the largest eigenvalue also the eigenvector of the second largest eigenvalue. It is obvious that there is a "mode" that is more common to all stocks than the market mode. The fact that the signs are opposed indicate that the modes are anti-correlated. This result proves that the subtraction of the apparently informationless market mode by removing the mode coded in the largest eigenvalue's eigenvector is possibly incorrect as well as the assumption that the market mode is the most common co-movement.

It is important to note that the size of the largest eigenvalue is determined by the number of participating time series and the strength of the correlation. It is therefore not clear what the market mode is. Chap. 8 suggests a measure to improve the situation. In any case, the stylised fact: "There is a co-movement of stocks that does not carry information." is vague.

6.2 Toy example: artificial stock market data

The realistic situation in financial data is composed of time series with different volatility and liquidity. Both of these properties are modeled within the CTRW model with two parameters each which determine the scale (multiplicative factor) and the power-tail of the respective distribution: α, β, γ_x, γ_t; refer to the respective section. This is a rather primitive approach but produces reasonable fake data that can hardly be distinguished from real data. To recreate realism each stock will have his own combination of such parameters. Studies have produced empirical results on the distribution of the above scale factors and power-tails. We draw 1000 independent combinations of the 4-tuple $(\alpha, \beta, \gamma_x, \gamma_t)$. These parameters are all taken from truncated Gaussians, which are parametrised according to a a reasonable and realistic choice. Fig. 6.6 shows the cloud of 4-tuples (at least in a 3 dimensional sub-space) and the resulting spectrum of the correlation matrix. The sampled length in terms of data pairs is 28800 reflecting the seconds of cumulated trading in one months. The distribution of γ_t is indeed chosen such that $t = 1, 2, 3, \ldots$ is empirically meaningful to denote seconds. At a glance:

Parametric null-hypothesis high frequency stock market model

- $N = 1000$ individual CTRWs with parameters $(\alpha, \beta, \gamma_x, \gamma_t)$
- $T = 28800$ reflecting tick-by-tick data of one month
- Sampling grid: $\Delta x = 1$ corresponding to one second or tick
- Parameters $(\alpha, \beta, \gamma_x, \gamma_t)$ are drawn independently according to a Gaussian distribution $N(\mu, \sigma; m, M)$ where m, M are cut-offs. These numbers are chosen such that the unit of time is one second.
- The Gaussians are parametrised according to Table 6.1.

	μ	σ	m	M
α	1.6	0.20	1.2	2.0
β	0.95	0.05	0.9	1.0
γ_x	0.0003	0.00005	0.0002	0.0004
γ_t	10.0	5	1.0	20.0

Table 6.1: Parameters of the Gaussian distributions $N(\mu, \sigma; m, M)$ truncated at m, M used to generate the parameters α, β, γ_x, γ_t of the artificial market.

In Fig. 6.6 the eigenvalues that leak out both ways of the spectrum indicate spurious correlation evoked mostly by power-tailed jumps and to a small extent by power-tailed waiting-times. The mechanism with the latter is totally different, however, and this issue is picked up in Chap. 5. In an application one would have to compare a spectrum taken from one set of data to the respective null-hypothesis. Only confidence statements on single eigenvalues are possible for the purpose of which the null-hypothesis is also required to provide empirical information on the dynamics of the outer eigenvalues. The latter requires statistics on single eigenvalues, i.e. repeated simulations of the artificial market. This task is of the order of several hours on contemporary desktop computers.

What was condonely ignored so far is the normalisation in the Pearson correlation estimator. It needs the variances of the sampled increments whose expectation values do not exist in the case of power-tailed jumps, i.e. $\alpha < 2$. However, in this example we average over realisations of fixed length. The realised averaged distribution of variances does converge and so the spectrum of the correlation matrix. It should be mentioned that the normalisation is a choice that depends on the theory one likes to recover. For example, the theory by Burda et al. [24, 25, 26, 30, 33] requires an α-dependent normalisation and the covariance estimator Eq. (3.35). Consequentially the resulting spectrum is different to the one in Fig. 6.7. It shows example spectra for five values of α calculated from CTRWs with Gaussian waiting-times and Lévy distributed jumps. Each spectrum corresponds to one α and all CTRWs for one spectrum are parametrised with the same value. We can see that the realised spectrum Fig. 6.6 is quite similar (up to statistical uncertainties) to the spectrum with $\alpha = 1.7$. This indicates that a mixture of αs produces a situation that corresponds roughly to the mean/median of αs. This can be seen as an amendment to the previous chapter on the comparison of the Pearson and Fourier correlation estimators.

Chap. 3 presented a theory that is not capable to deal with a distribution of αs. The realistic null-hypothesis will therefore still be a numerical Monte Carlo calculation.

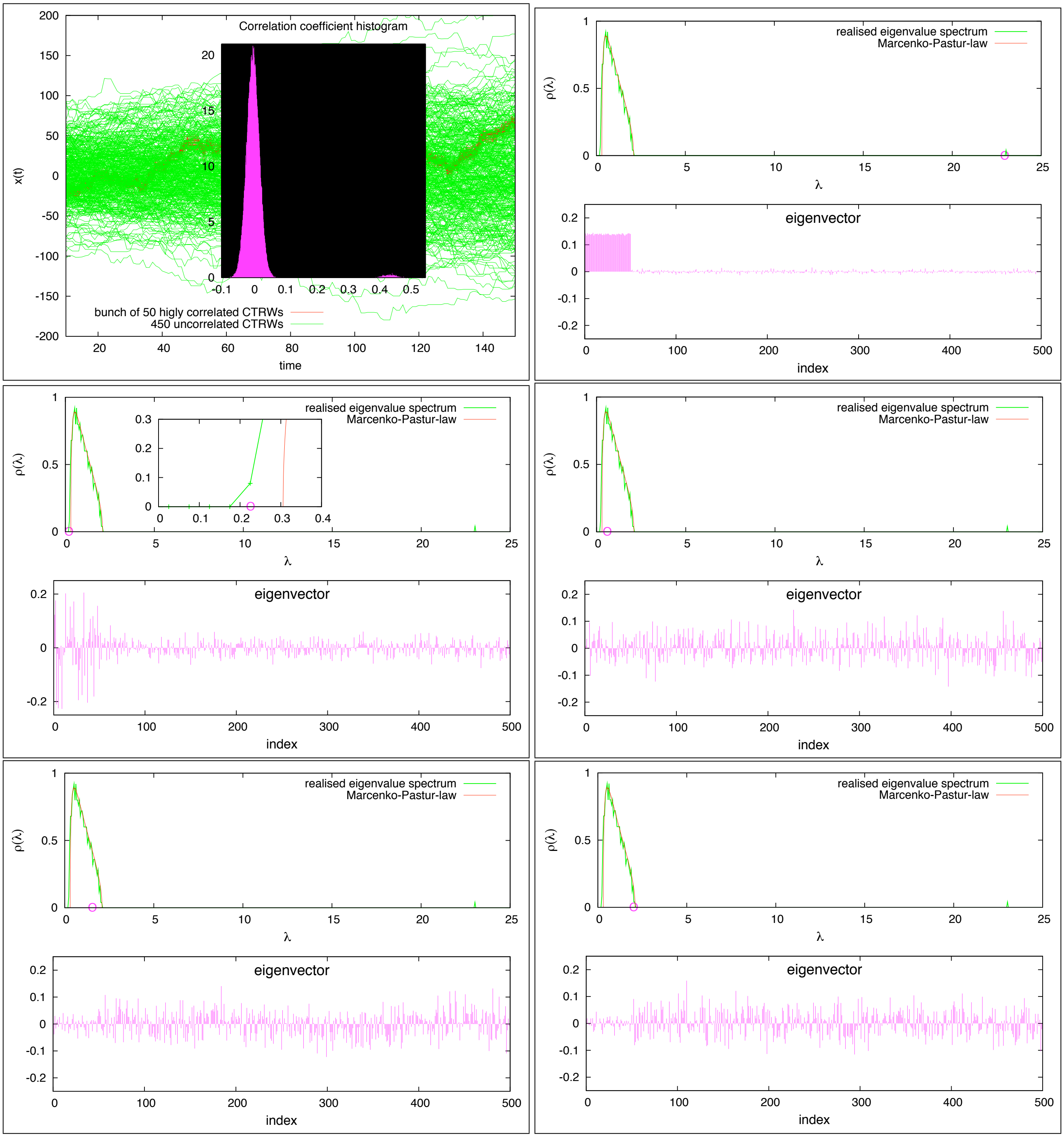


Figure 6.2: Eigenvector study on a correlation matrix of CTRWs with a correlated group. A: 50 red correlated CTRWs embedded in the 450 uncorrelated CTRWs. The histogram reveals the correlated group as a pile of high coefficients at 0.4. B: Eigenvector of the largest eigenvalue. C: Eigenvector of the smallest eigenvalue. D: Eigenvector of an eigenvalue just inside the Marčenko-Pastur spectrum. E: Eigenvector of an eigenvalue close to the right edge of the Marčenko-Pastur spectrum. F: Eigenvector of an eigenvalue at the right edge of the Marčenko-Pastur spectrum.

A	B
C	D
E	F

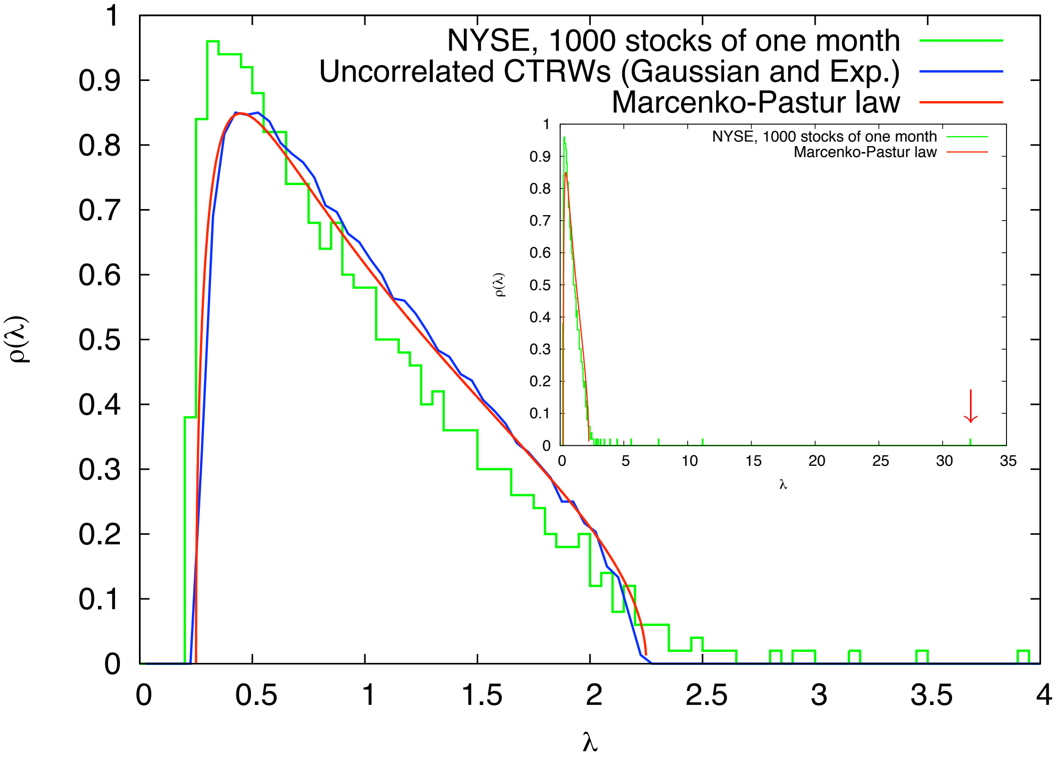


Figure 6.3: Eigenvalue spectrum of the correlation matrix produced of one month of 1000 stocks from the New York Stock exchange. The sampling for the previous tick interpolation is 100 seconds. The red arrow marks the largest eigenvalue.

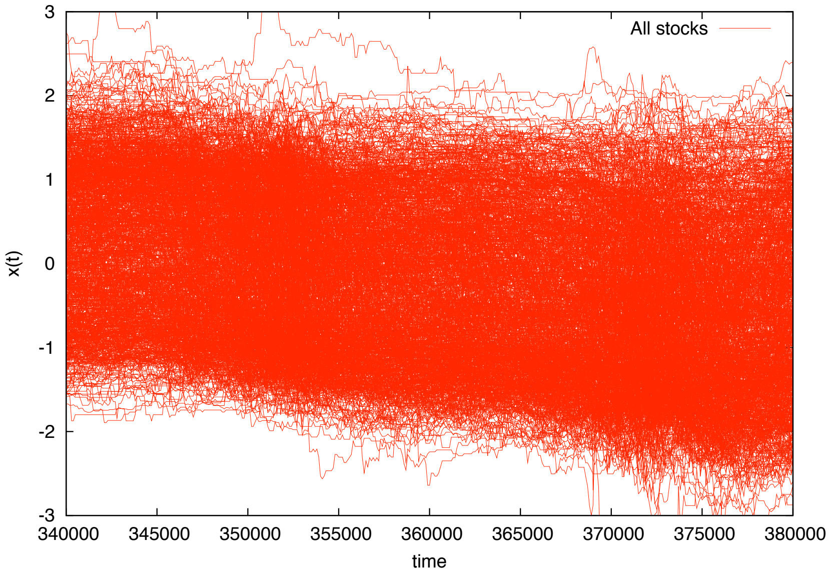


Figure 6.4: A small section within one month of 1000 log-prices from the New York Stock exchange. The time series are normalised to unit variance and zero mean. The downwards trend is reflected in one large eigenvalue.

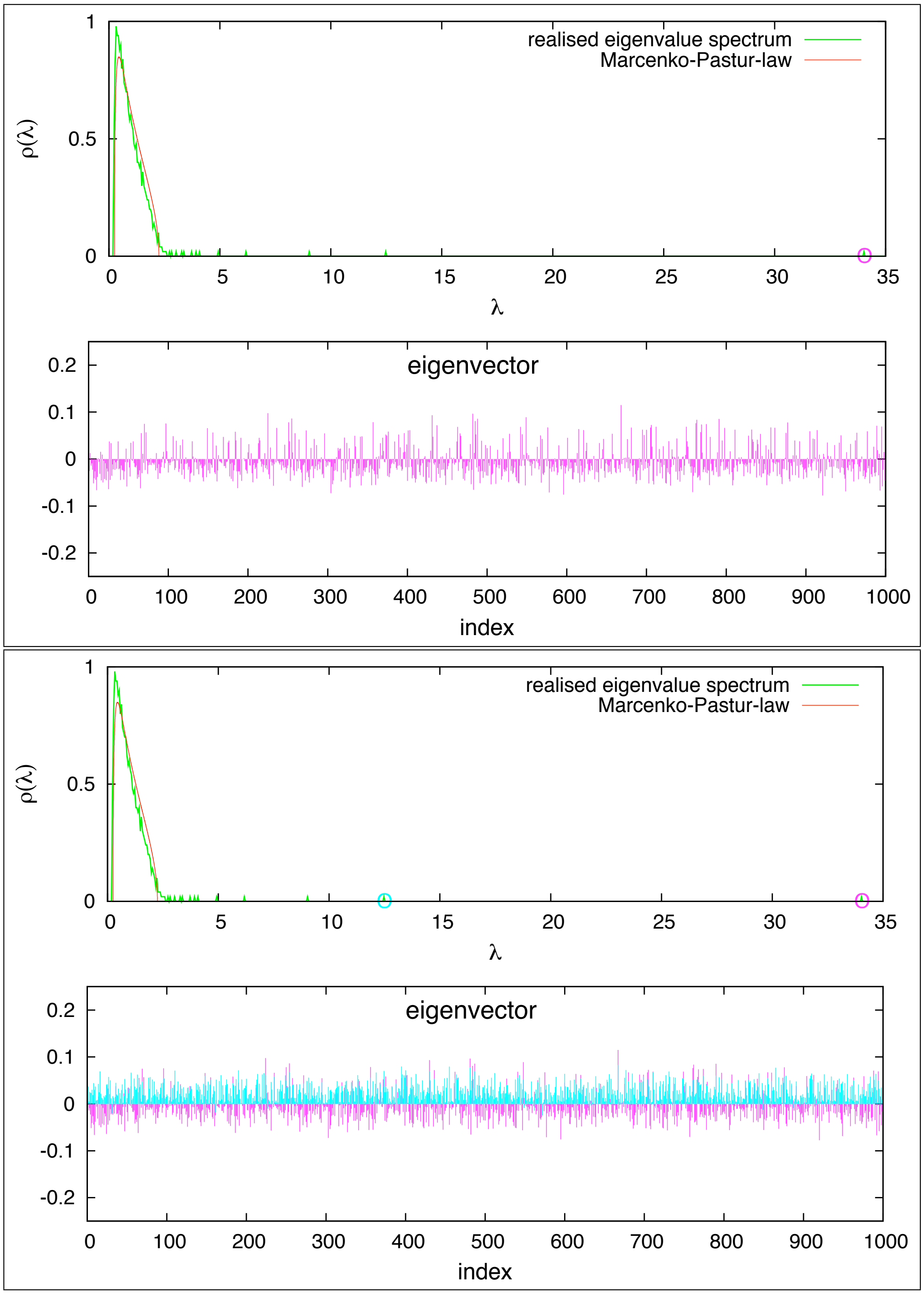


Figure 6.5: Eigenvalue spectrum and selected eigenvectors from the correlation matrix produced of one month of 1000 stocks from the New York Stock exchange. The top panel shows the eigenvector of the largest eigenvalue. This plot is repeated below with the addition of the eigenvector of the second largest eigenvalue plotted on top (the circle colours correspond). The elements of the second largest eigenvalue's vector (turquoise) are almost all positive.

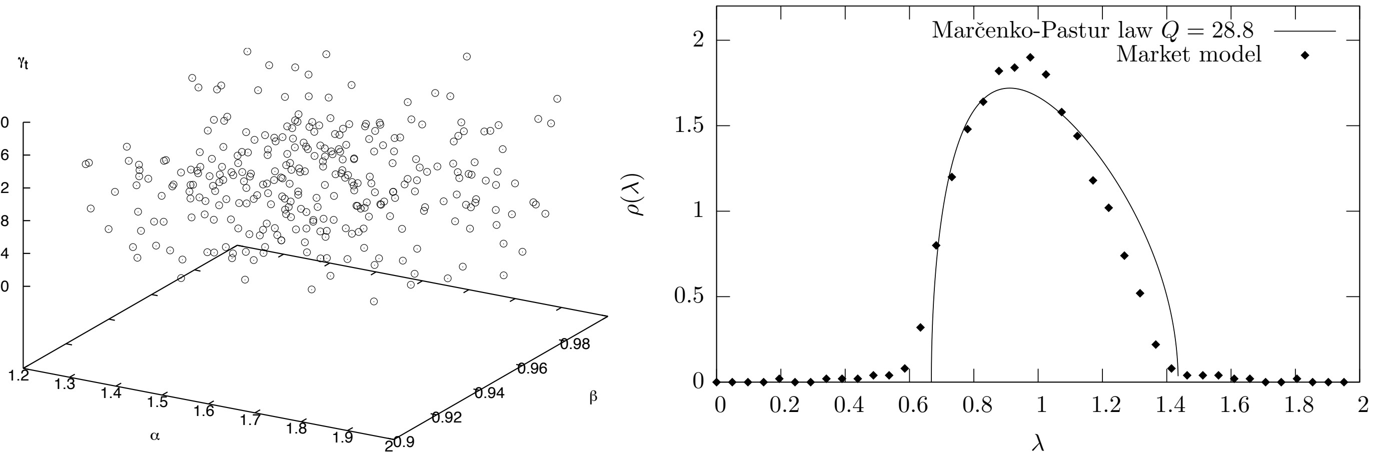


Figure 6.6: Parametrisation and spectrum of an artificial null-hypothesis test market with a realistic content of scale factors and more or less power-tailed jumps and waiting-times.

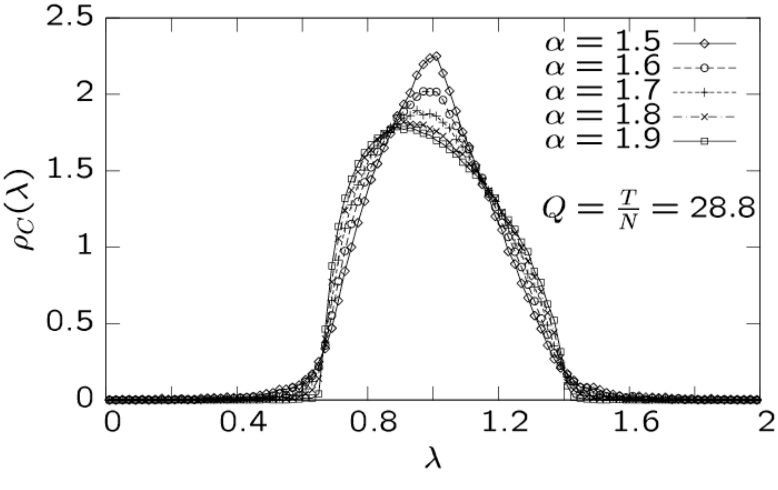


Figure 6.7: Spectra of correlation matrices of CTRWs produced with the Pearson correlation estimator. The increments of the CTRWs contain power-tailed jumps according to the Lévy distribution with the above list of parameters α. Despite the diverging variance of increments or sampled increments ($\Delta t = 1, \gamma_t = 1$) the spectra converge. Each spectrum is averaged over 50 realisations.

Chapter 7

On the relevance of the sampling of continuous-time random walks in correlation matrix analysis

> *'What is the use of a book', thought Alice, 'without pictures or conversations?'*
>
> *L.C.*

7.1 Introduction

For the Pearson correlation estimator a time series is most meaningfully sampled at equi-distant times with previous tick interpolation to achieve a uniform weighting of time. This assumption requires homogeneity and stationarity of the data. Moreover, it requires a justification of equal weighting of time, which is usually part of the model assumption and implied by the questions asked. This cannot be stressed too often. Here we deal with time series of the kind continuous-time random walk as a model for financial time series. Considering the completely uncoupled case for all random variables the arising questions are:

1. How are the correlation coefficients distributed and do the distributions of waiting-times and increments influence the correlation coefficient's distribution?
2. Does the spectrum of a correlation matrix change with the distribution of waiting-times and increments? Or equivalently, does the joint probability change?

While the answer to 1 might be negative, it is the joint probability density that drives the spectrum. Likewise, two different distributions of matrix elements can give the same spectrum. The following reasoning is situated in the Gaussian regime, i.e. with distributions of finite moments, and starts with a theoretical consideration of the distribution of increments as a function of the sampling frequency, or better: as a function of the ratio of sampling frequency and mean exponential waiting-time. A glimpse ahead: The result is that the distribution is rather insensitive to variations of the sampling frequency. Then we argue that the distribution of correlation coefficients is bound to converge very quickly towards a Gaussian. The answer to question 1 is that all distributions of jumps and waiting-times with finite moments give indistinguishable results for situations that are meaningful in the context of correlation matrix analysis and random matrices in the realm of the Marčenko-Pastur theory where the number of samples (or better sampled increments) is of the order or greater than the number of time series. The answer to question 2 is that in the uncorrelated case equally uncorrelated waiting-times do not truly change the joint probability distribution of covariances. Any changes in the spectrum due to the latter are statistical and similar to the finite size effect.

7.2 The distribution of sampled increments Δx

The curtosis of the distribution of increments sampled at equi-distant times $\Delta x(t) = x(t) - x(t - \Delta t)$ is a function of the sampling interval:

$$\beta_2 = 3\frac{\tau_0}{\Delta t}, \tag{7.1}$$

where τ_0 is mean waiting-time in $x(t)$ and Δt is the sampling interval. Traditionally the curtosis is denoted by the symbol β_2 with the classification: $\beta_2 > 3$: leptocurtic, $\beta_2 = 3$: mesocurtic, $\beta_2 < 3$: platycurtic. Since β_2 is a function of Δt one might also expect a dependence of the distribution of correlation coefficients on Δt. For a CTRW as introduced in the previous chapter and position definition according to Eq. (2.27) the solution of the master equation (the probability to be in position x at time t) is given by [154]

$$p(x,t) = \sum_{n=0}^{\infty} P(n,t)\lambda^{*n}(x), \tag{7.2}$$

where

$$\lambda^{*n}(x) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \mathrm{d}\xi_{n-1} \dots \mathrm{d}\xi_1 \, \lambda(x - \xi_{n-1}) \dots \lambda(\xi_1) \tag{7.3}$$

is the n-fold convolution of the distribution of jumps $\lambda(\xi)$ and $P(n,t)$ is the cumulative density of n jumps up to time t

$$P(n,t)=\int_0^t \psi^{*n}(t-\tau)\Psi(\tau)\mathrm{d}\tau, \tag{7.4}$$

with Ψ the cumulative waiting-time density. And finally

$$\psi^{*n}(t)=\int_0^t \dots \int_0^{\tau_1} \mathrm{d}\tau_{n-1}\dots\mathrm{d}\tau_1 \ \psi(t-\tau_{n-1})\dots\psi(\tau_1) \tag{7.5}$$

is the n-fold convolution of the waiting-time density. For analytically not nicely behaved functions the objects above are unlikely to have an analytic solution. For a normal compound Poisson process (NCPP) with exponential waiting-times $\psi(\tau)=\exp(-\tau/\tau_0)/\tau_0$ and mean τ_0 and Gaussian jumps with variance σ_ξ^2 we can get at least a closed form for the cumulative density:

$$P(n,t)=\frac{(\Delta t/\tau_0)^{(\Delta t/\tau_0)}}{(\Delta t/\tau_0)!}, \tag{7.6}$$

and the multiple convolution of a Gaussian is again a Gaussian with variance $n\sigma_\xi^2$:

$$\lambda^{*n}(x)=N\left(0,\sqrt{n}\sigma_\xi\right)=\frac{1}{\sqrt{2\pi n\sigma_\xi^2}}\exp\left(-\frac{x^2}{2n\sigma_\xi^2}\right). \tag{7.7}$$

Because of the stationarity and homogeneity of the process the distribution density of sampled increments at intervals Δt is then given by [154]:

$$p(\Delta x)=P(x,\Delta t)=e^{-\Delta t/\tau_0}\sum_{n=0}^{\infty}\frac{(\Delta t/\tau_0)^n}{n!}N\left(0,\sqrt{n}\sigma_\xi\right), \tag{7.8}$$

which is clearly not Gaussian, but a sum of many different Gaussians. However, Eq. (7.8) is approximated well with

$$p(\Delta x)\approx Ke^{-\Delta t/\tau_0}\left(\frac{(\Delta t/\tau_0)^{(\Delta t/\tau_0)}}{(\Delta t/\tau_0)!}\right)N\left(0,\sqrt{\frac{\Delta t}{\tau_0}}\sigma_\xi\right), \tag{7.9}$$

because the factor

$$\frac{a^n}{n!} \tag{7.10}$$

has one sharp maximum at $n=a$, see Fig. 7.1. The added constant K accounts for neighbouring Gaussian's probability mass that is not included in the approximation. The sum in Eq. (7.8) is dominated by few terms in a narrow interval where $n\approx\Delta t/\tau_0$, in other words, a sum of a few but similar Gaussians. This sum essentially gives a Gaussian if the sampling frequency is sufficiently low compared

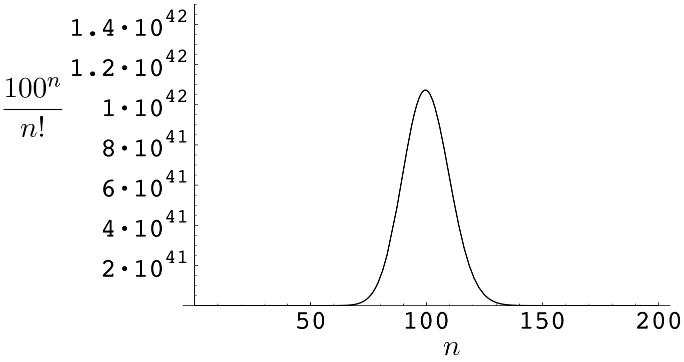


Figure 7.1: The plot of $100^n/n!$ shows a shape with one sharp maximum at $n = 100$.

to the mean exponential waiting-time so that the curtosis of the distribution of increments is negligible.

The above reasoning is valid for all jump densities with finite moments. Due to the central limit theorem the sum of sufficiently many jumps becomes normally distributed very quickly. For uniform jumps the summation of only 12 already gives a good approximation. The latter is well known as the method of the 12s for the production of normal random numbers from the early days of computing. If the sampling frequency allows for sufficiently many summations of jumps, we get essentially indistinguishable results for moderate variations of the sampling frequency.

7.3 The distribution of correlation coefficients with missing data

The next issue in answering the questions on the influence of the sampling is the distribution of correlation coefficients. If random numbers X and Y are independent and follow a Gaussian distribution then this means that they are also jointly Gaussian distributed. Consequentially, the pair (X, Y) follows a bivariate Gaussian distribution:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_X^2} - \frac{y^2}{\sigma_Y^2} - \frac{2rxy}{\sigma_X\sigma_Y}\right)\right), \quad (7.11)$$

where r is the correlation between X and Y. For N realisations (or samples) of the *uncorrelated* pair (X, Y) the distribution of correlation coefficients is given by

$$p(r) = \frac{(N-2)\Gamma\left(\frac{N-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{N}{2}\right)}(1-r^2)^{(N-4)/2} = \frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}(1-r^2)^{(\nu-2)/2}, \quad (7.12)$$

where $\nu \equiv N - 2$ are the degrees of freedom. Note that with $N = 2$ samples there are zero degrees of freedom which indeed makes little sense if talking about a matrix of correlations. Eq. (7.12) is derived from the Student-t distribution and if N grows large it quickly approaches the normal distribution, see Fig. 7.3. It is a variant of the definition of the Student-t distribution also with respect to the definition of the degrees of freedom. In the case of correlation coefficients the (sensible) minimum number of measurements is $N = 4$ whereas in general it is $N = 2$ and the degrees of freedom are defined as $\nu \equiv N - 1$.

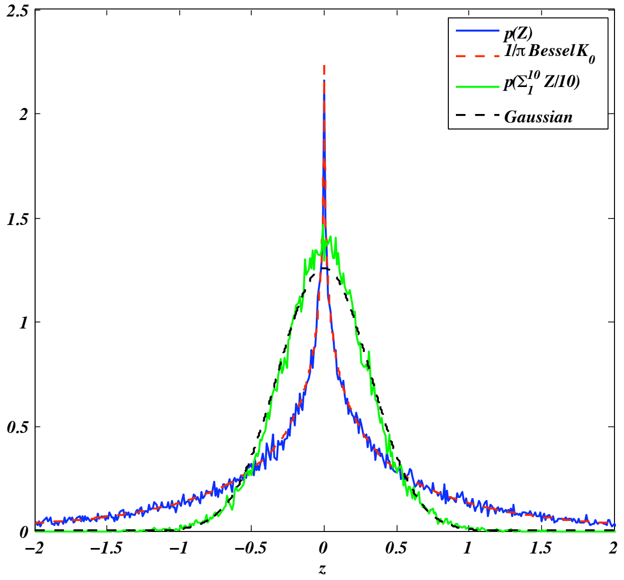


Figure 7.2: The product of two Gaussian random variables $Z = XY$ is distributed according to $K_0(z)$, the modified Bessel function of the second kind with index zero. The sum of only 10 random Z is already close to a Gaussian. The continuous lines are histograms.

The above reasoning can be regarded as stemming from the large size limit. This also works from the perspective of summing single random variables. The limiting distribution of the sum of products $\sum XY$, where X and Y are Gaussian random variables in reference to the distribution of increments from above can be calculated analytically via the folding of the two respective densities according to a special case of the Mellin integral [66]:

$$p_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z/x) \frac{1}{|x|} \mathrm{d}x. \tag{7.13}$$

In this example the joint probability density $f_{X,Y}$ factorises and the product $Z = XY$ is distributed according to the density function

$$p_Z(z) = \frac{1}{\pi \sigma_X \sigma_Y} K_0 \left(\frac{|z|}{\sigma_X \sigma_Y} \right), \tag{7.14}$$

where $K_0()$ is the modified Bessel function of the second kind with index zero and

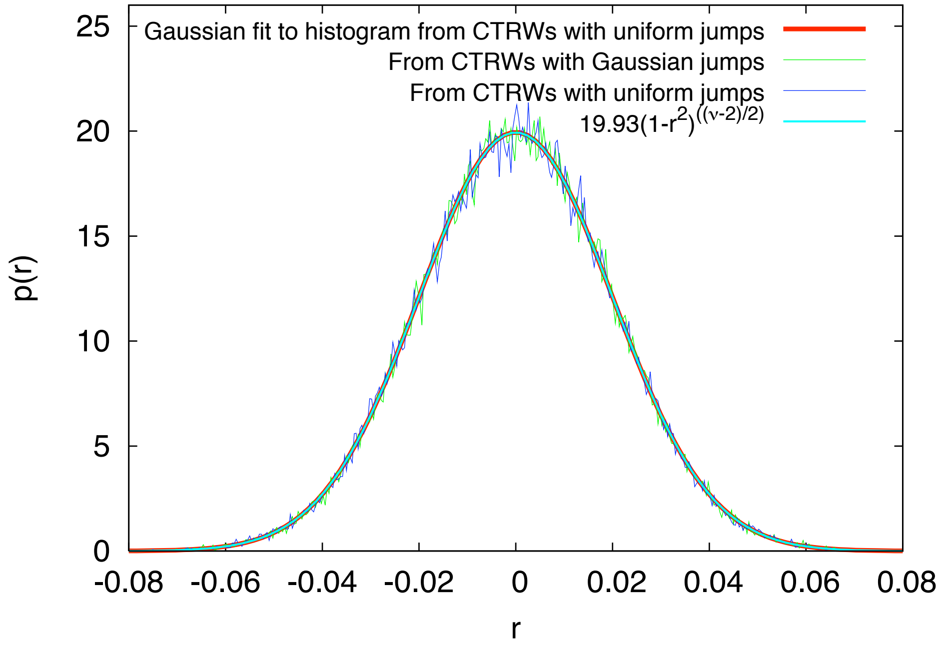


Figure 7.3: Empirical distribution of correlation coefficients and comparison to theory. The plot shows two empirical histograms from correlation coefficients (blue and green) calculated between CTRWs with Gaussian jumps (NCPPs) and CTRWs with uniform distribution of increments $\xi \in [-1/2, 1/2]$. The waiting-times in both are exponential and the sampling rate is $\Delta t = 10\tau_0$ in units of the mean waiting-time. The number of CTRWs is $N = 500$ with a length of $2500\Delta t$. The Gaussian fit to the latter ($\sigma = 0.0200388 \pm 1.418 \times 10^{-5}$) shows that both histograms are indistinguishable from a Gaussian. Imposed onto the fit is the Student-t density function with $\nu = N - 2 = 2498$.

is given by

$$K_0(z) = \int_0^\infty \frac{\cos(zt)}{\sqrt{t^2+1}} \mathrm{d}t. \tag{7.15}$$

This function, interpreted as a density, has finite moments. The asymptotic behaviour is [182]

$$K_n(z) = \sqrt{\frac{\pi}{2z}}\, e^{-z} \left(1 + O\left(\frac{1}{z}\right)\right), \tag{7.16}$$

which exhibits exponential decay. With this we find ourselves in the realm of the Marčenko-Pastur theory and the Pearson correlation coefficient converges to a Gaussian. Fig. 7.2 shows in terms of histograms that the (normalised) sum of only 10 scaled Bessel distributed variables produces a Gaussian distribution. The message of Figs. 7.2 and 7.3 is to show that the distribution of jumps has negligible influence on the distribution of correlation coefficients as long as its moments are finite. A sampling grid of $\Delta t = \tau_0$ in the data for Fig. 7.3 gives exactly the same picture which supports the conjecture that the random introduction of missing values, e.g. by randomly picking out jumps, has no or little influence on the distribution of increments, especially so if the moments of the jumps are finite. This also demonstrates that the curtosis of increments is generally of negligible influence in meaningful examples of correlation matrix analysis with CTRWs. We can even repeat the same scenario with the distribution of jumps $\lambda(x) = 1/2(\delta(x-\sigma) + \delta(x+\sigma))$, which is the most un-Gaussian density whose moments are all σ.

As an answer to question 2 we can adopt the reasoning from literature on how to deal with missing variables of which there is not many, however [104, 168]:

> For covariance matrix estimation with randomly and independently missing values the pair-wise estimation of covariances is **un-biased** if the data is stationary.

We expect from this:

> The deletion of values from the covariance estimator via independent and uncorrelated waiting-times across all time series does not truly influence the joint probability density of correlation matrices of uncorrelated CTRWs.

The above "truly" allows for differences which are of a more subtle nature but are entirely caused by statistical uncertainty that is introduced by random deletion of values. This type of influence we regard as similar to the finite size effect in small correlation matrices that does contradict the Marčenko-Pastur theory.

See Fig. 5.8 for an example with uncorrelated CTRWs using Gaussian jumps and power-tailed waiting-times. It is noteworthy that by keeping $Q = T/N$ constant the shape of the empirical spectrum does not change for different values of T. This indicates a universal behaviour introduced by systematic removal of information from the Pearson covariance estimator. In the following Section 5 we demonstrate that the Fourier estimator gives statistically the same result up to negligible fluctuations.

When performing the analysis above on real data and observing a significant deviation from the blue curve presented in Fig. 5.8 we can suspect non-stationarity as one of the reasons of the data or that some parts, even single data points, carry important information. This is a particularly important implication.

7.4 Conclusion

The main statements of this chapter are given in boxes where they come up naturally during the course of the reasoning. The "real world" question we sought to answer was on the role of waiting times in the context of correlation matrix analysis in scenarios similar to financial stock market data. It seems that equal-time correlation analysis is hardly influenced if the stocks have some kind of “minimum liquidity” to provide at least a little information for the correlation estimator. In the Chap. 8 we will see that equal time correlation is not the end of the story and that the disposal of time, as the Pearson estimator does, is sometimes not appropriate. Moreover, it also seems that the correlation matrix spectrum is a rather insensitive tool to the information that is contained in time series data.

Chapter 8

Spectral properties of correlation matrices – towards enhanced spectral clustering

> *As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality.*
>
> *A. Einstein (This time not by L.C. because it fits very nicely.)*

This chapter compiles some properties of eigenvalues and eigenvectors of correlation and other matrices constructed from uncorrelated as well as systematically correlated Gaussian noise. The situations depicted in this setting are found in time series analysis as one extreme variant and in gene profile analysis with microarrays as the other extreme variant of the possible scenarios for correlation analysis and clustering where random matrix theory might contribute. The main difference between both is the number of variables versus the number of observations. To what extent results can be transferred remains to be seen. The origin for this material was a project on microarray data and clustering at the University of Alessandria with the goal to cluster data of differential expressions via random matrix theory. While random matrix theory as such makes statements about the statistical properties of eigenvalues and eigenvectors, the expectation is that these statements, if used in a proper way, will improve the clustering of genes for the detection of functional groups. In the course of the scenarios the relation and interchangeability between the concepts of time, experiment and realisations of random variables will play an important role. The mapping between a classical random matrix ensemble and the microarray scenario is not yet obvious. It was also necessary to analyse the spectral properties of correlation matrices from the bottom up starting from the opposite end of the spectrum of scenarios.

8.1 Introduction

We review spectral properties of correlation-like matrices from a general point of view. One question to answer: What meaning do the eigenvalues and eigenvectors *exactly* have, what information can be extracted that can be used to improve clustering, for example? The established association of *the* large eigenvalue and respective eigenvector with some kind of “dominant” mode [89] in the underlying data seems to be just half of the story.

From a mathematically abstract point of view the situation and task is, or likes, to be interpreted as the following: We are presented with stochastic variables $\xi_n, n = 1, \ldots, N$. The values taken by these variables are indexed by $t = 1, \ldots, T$ by writing $\xi_n(t)$ to allow the association with time series while this labeling may refer to the experiment number or any other label that expresses meaningfully that variables $\xi_1(t), \ldots, \xi_N(t)$ belong to “one measurement”. This pedantry is necessary because the interpretation and the choice of methods crucially depend on the mappability between the mathematical object and the real world. In the former there is no concept of time and its introduction must be well defined and justified. The entire data can be arranged in a matrix $\mathbf{M}$ of dimension $N \times T$. Assuming that the average is zero, the Pearson estimator for the covariance matrix (C_{ij}) is given by Eq. (3.2). The covariance or correlation matrix $\mathbf{C}$ is often associated with the Wishart matrix for which Marčenko and Pastur derived an analytic spectrum in the large size limit if the variables $\xi_n(t)$ are independent and identically distributed with the condition of finite moments [118]. In previous chapters the equivalence between the correlation matrix and the Wishart matrix was taken for granted.

A typical task is to extract sets of variables that form correlated groups, or rather groups that have something in common. The above mentioned correlation coefficient is just one of many possible “linkages” between (real valued) random variables or even other random objects. It is to view clustering as a special case of spectral reconstruction (approximation) of matrices or related networks [40]. The notion of correlation can be extended to any coefficient that measures a link between two random objects in terms of a real number for which a suitable pair-wise distance can be defined. The definition of a correlated group is therefore somewhat arbitrary, likewise is the resulting clustering of different methods more or less different. In real world data there will usually be no correlation in the mathematical sense but possibly something very similar and interpretable as correlation. Many methods act on matrix $\mathbf{C}$ to extract information. Specialised methods make model assumptions on the type of correlation (or link value) and are thus *empirically optimised* to cluster the random variables that work best for the given source of the random variables.

Additionally, the mathematically abstract context of realisations of random variables at equal times to which correlation measurement is often mapped to is not justified in some cases. It does not hold, for example, if not all random variables provide a realised value for each time index. This is the case in high-frequency fi-

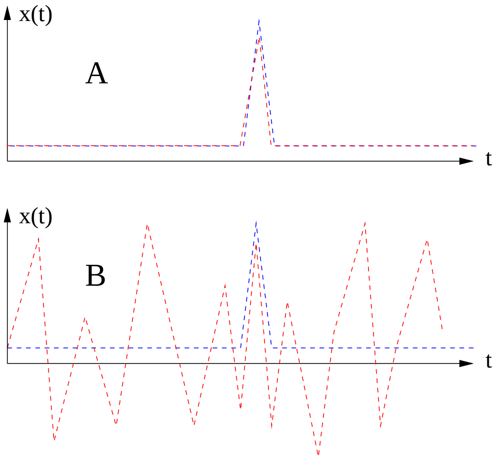


Figure 8.1: Schematic time series. In panel A both show a peak at the same time. In B the peaks are still identical while Red seemingly fluctuates at random. The correlation coefficients calculated from both situations are identical if the increments lie on the same grid. The coefficients are likely to be very similar even with continuous-time random walks and meaningful interpolation schemes.

nancial data where waiting-times produce zero returns between samplings. Neither are the (differential) expression values extracted from microarrays easily justified to be interpreted as random variables since for equal experimental settings equal values must be expected. Consider the situation shown schematically in Fig. 8.1. The time series change only once and at the same time in panel A. In panel B time series Red fluctuates a lot and has, by coincidence or not, an identical spike together with Blue. A correlation measurement combined with some typical interpolation technique would produce identical or at least very similar correlation coefficients for both situations. It depends strongly on the system, application and questions asked if it is appropriate to call either situation correlated or not. It is most likely A that depicts a significant connection between the two time series. Note that this likelihood increases with the total duration or number of data points! We must therefore realise that the regions in time having zero increments do contain information, in particular if their time scale is of the similar order of magnitude as other time scales in the respective situation, for example the total duration of the measurement. Ergo:

Any post-processing that only considers the correlation coefficients disregards time and produces in such a case a joint probability density in the (dis-)similarity matrix that does not correctly reflect the connection between the time series.

It will be demonstrated later on that with time series with behaviour as described in Fig. 8.1, or respective stylised facts, the data cannot be disregarded but should be used in the reconstruction of "modes" and separation of correlated clusters that are otherwise not separable. *Note, that the term "mode" is not mathematical and mostly intuitive if used in real world data measurement.*

Clustering with matrices and their eigenvalues and vectors is well established in graph theory for a long time under the term "spectral clustering". A good tutorial is Ref. [175]. In short, it is based on a certain dissimilarity measure matrix $\mathbf{L}$, while definitions sometimes disagree, and it uses the eigenvectors of the smallest k eigenvalues. This value has the same meaning as in k-means clustering, i.e. the a priori estimate on the number of clusters. $\mathbf{L}$ is symmetric and called Laplacian and one frequent definition is

$$\mathbf{L} = \mathrm{diag}(d_i) - \mathbf{C}, \tag{8.1}$$

where $\mathbf{C}$ is the unweighted (positive) adjacency matrix and $\mathrm{diag}(d_i)$ is the diagonal matrix of vertex degree:

$$d_i = \sum_{j=1}^{N} C_{ij}. \tag{8.2}$$

One can consider L to contain a dissimilarity measure via the negated $\mathbf{C}$ which in turn is analogous to the absolute value $|\mathbf{C}|$ sometimes used in a distance measure. For example, the comparison of clustering methods in Ref. [125] uses a dissimilarity measure that is close to the one used later. In the end it is unlikely that mathematical reasoning leads to the best choice of distance measure, as explained above. In the following examples and figures it is demonstrated that with some (dis-)similarity measure it makes sense to consider also other eigenvectors than the large eigenvalues's eigenvectors. In spectral clustering one is free to choose a suitable clustering method, for example k-means, which then performs the clustering using these eigenvectors. This also means that the number of clusters must be guessed beforehand. The motivation to consider here also a dissimilarity measure is to keep track of what type of matrix other non-spectral methods as k-means or PAM use.

The idea for an improved spectral clustering uses the correlation matrix as a similarity measure since it contains no less information than any dissimilarity matrix. Moreover, a theory exists on its random case spectrum. It seems that the use of the Laplacian matrix $\mathbf{L}$ in standard spectral clustering is mostly to achieve plausibility since it matches with a mathematical construction in graph theory, i.e. just for plausibility. Furthermore, for a "mode" carrying a correlation information we also have small eigenvalues leaking out of the Marčenko-Pastur law of uncorrelated data. This is true for the similarity (correlation) matrix and, in an analogous way, also for the dissimilarity matrix used here and defined later. Since we assume that we are faced with a "noisy" situation we must use all information we can extract. Since these small eigenvalues and associated eigenvectors are likely to contain redundant information about the correlated cluster it is appealing not to ignore this information. The naive mapping of a real world situation to simultaneous realisations of random variables is often not easy to be justified and is mostly argued for because of reasonable results. An example are liquid together with illiquid stocks. There the data itself can be used in the reconstruction of the correlations.

The following sections construct artificial situations that are "extreme" for didactic purposes in the sense that they are not realistic but allow to recognise features in the eigenvalue and eigenvector spectra that could be used in the better exploitation of the information content given in a more noisy and more realistic data set.

8.2 Scenario 1 – Correlated noise with many variables and many measurements per variable

8.2.1 One correlated cluster

The scenario demonstrated here mimics synchronous financial data analysis, i.e. at least as many measurements as variables:

- $T = 200$ number of realisations per random variable
- $N = 200$ number of random variables
- $N_c = 1$ number of independently correlated groups of variables
- $N_1 = 20$ number of correlated variables in group 1 (only one here)
- $\xi_n(t)$ Gaussian noise data set n where $t = 1, \ldots, T$.
- Type of correlation within group i:

  $$\xi_n(t) = \xi_n(1-c) + \Xi(t)c \tag{8.3}$$

  Ξ is a prefixed "parent" noise vector specific for the correlated group. $c \in [0, 1]$ is a correlation coefficient.
- $c = 0.93$ (very high correlation)

The choice of $Q = T/N = 1$ is to avoid any factor Q if it appears in some normalisation. The mathematical/numerical construction of the artificial correlation is not so relevant since the realistic case will not provide a mathematical correlation coefficient either. Fig. 8.2 shows the correlation matrix created from the series ξ_1 to ξ_N. For identification the first 20 are correlated. Also shown is the more realistic disordered situation if the correlated data sets are unknown, i.e. shuffled. Fig. 8.3 shows the respective random walks (RWs) $x_i(t)$, $t \in \mathbb{N}_+$, that are created from the realisation of the random variables in the same manner as in the previous sections with the exception that these are equal-time random walks on a regular grid. For the purposes of this chapter this is entirely sufficient and the data points $x(t_1), x(t_2), \ldots$ are connected with straight lines. In Fig. 8.3 the correlated group is drawn red.

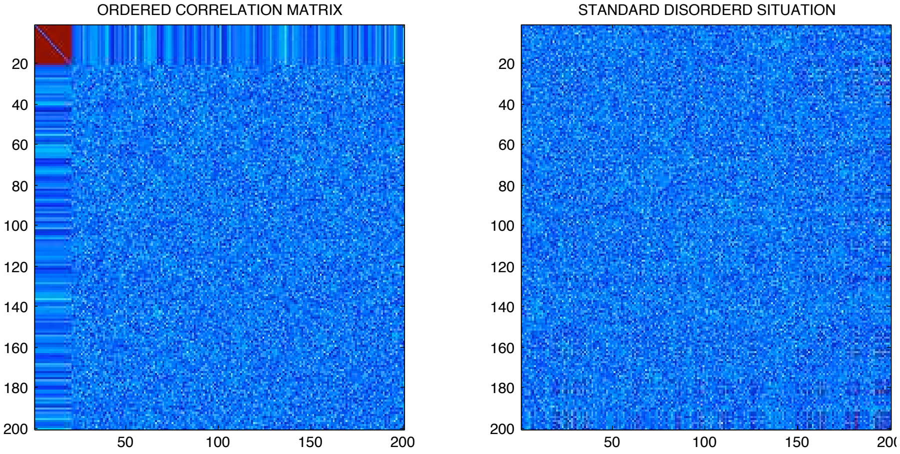


Figure 8.2: Correlation matrix of uncorrelated noise with one cluster of 20 artificially correlated variables. The right panel is reshuffled to imitate the standard disordered situation in reality, where the red dots of high coefficients are randomly distributed.

In addition to the correlation matrix **C** the following figures also show the results using a dissimilarity matrix **D**. The definition of a dissimilarity is a bit arbitrary. The measure used here is

$$\mathbf{D} = \mathbf{1} - |\mathbf{C}|. \tag{8.4}$$

Fig. 8.4 shows the main part of the eigenvalue spectra of both matrices. The numbering of eigenvalues is by size, i.e.:

$$\lambda_1 < \ldots < \lambda_N. \tag{8.5}$$

Some features are outside the plot range. Note that λ_-^C denotes the lower bound of the Marčenko-Pastur domain which is zero in this case with $Q = 1$. In the finite size situation with low correlation parameter the classification of eigenvalues as belonging to the correlated group is not unique due to the overlap of the distributions with the informationless bulk of the spectrum. Likewise, the expected size of eigenvalues fluctuates. We therefore use the order notation with $\mathcal{O}(\cdot)$ to indicate that an eigenvalue is expected to have the value $\mathcal{O}(x)$ or the number of eigenvalues in a distinct group is expected to be $\mathcal{O}(N)$. This is not to be confused with the usual meaning of order notation. For larger values of Q, N and correla-

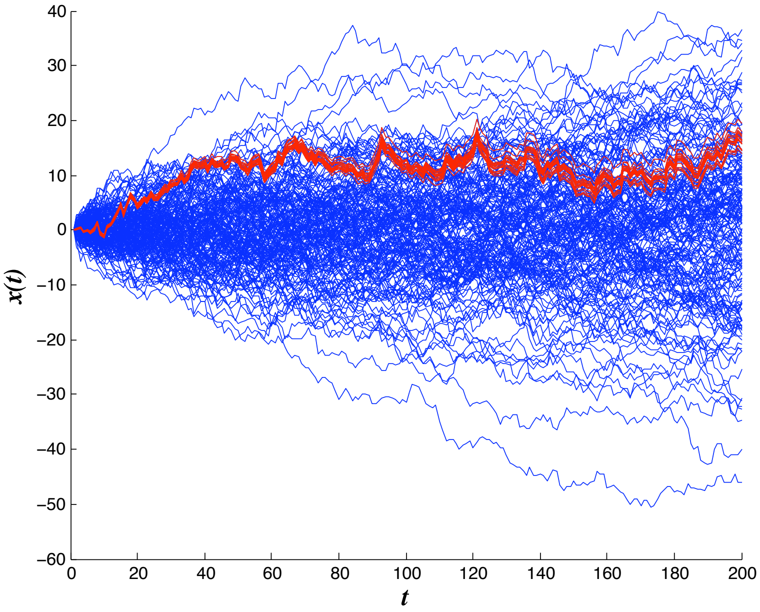


Figure 8.3: Random walks created from the noise in vectors ξ_1 to ξ_N. The correlated group is in red.

tion coefficient the groups of, e.g., zero eigenvalues, can be distinguished very well, see Fig. 8.5. The histograms in the eigenvalue figures are again deliberately not normalised to convey the absolute counts. The Laplacian's **L** smallest eigenvalue is always $\lambda_1^L = 0$ by construction [40].

Fig. 8.6 shows the eigenvector matrices of the correlation matrix as well as of the dissimilarity matrix **D**. The realistic (shuffled) situation is also shown. Columns are eigenvectors with column number corresponding to eigenvalue index. The ordering is entirely arbitrary as long as the pairs of eigenvalue and eigenvector are maintained well.

In the situation created here with one correlated cluster we find the following (partly empirical) properties.

A) Properties of the eigenvalues and vectors of similarity (correlation) matrix **C**:

1. The spectrum is strictly positive definite with a lower Marčenko-Pastur bound λ_-^C.

2. Conservation law in the limit of high correlation:

$$\lambda_N^C - N_1 = 0. \tag{8.6}$$

   This not only holds if $1 \ll N_1 \ll N$ but as long as $N \ll T$ and $c \approx 1$.

3. One large eigenvalue $\lambda_N^C = \mathcal{O}(N_1)$.

4. The eigenvector V^C belonging to the only large eigenvalue λ_N^C contains N_1

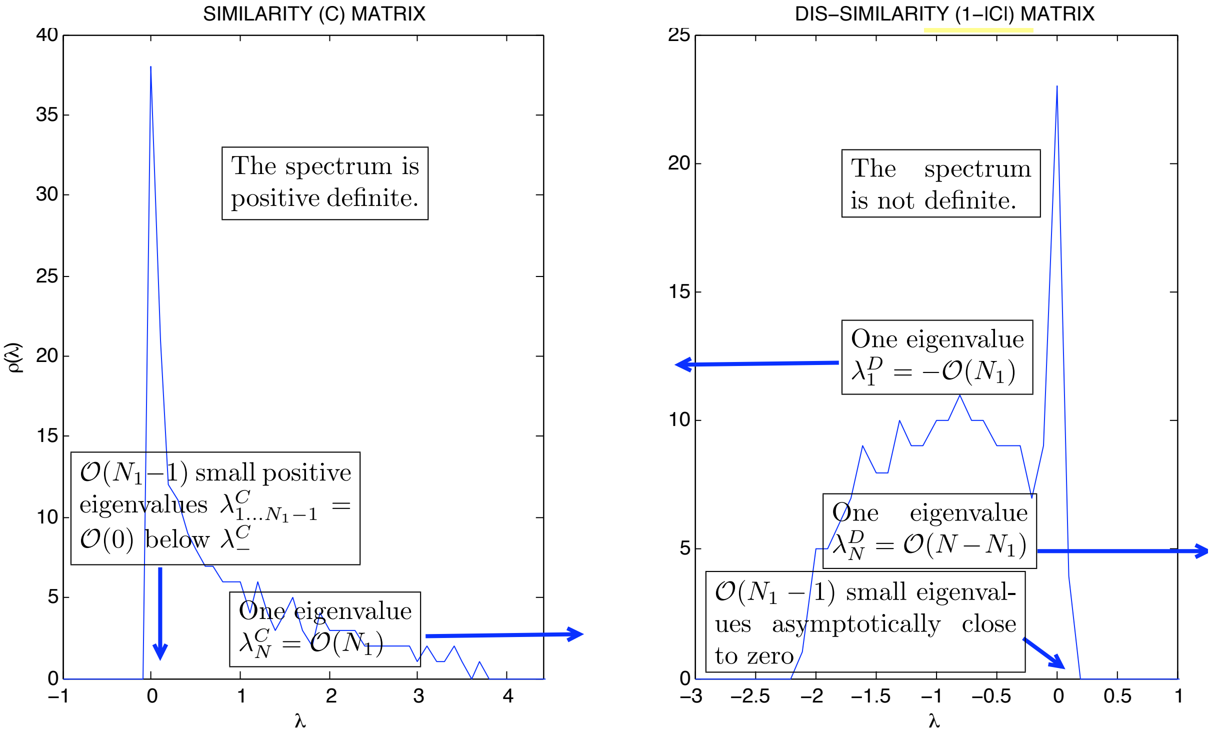


Figure 8.4: The features of the spectra of the similarity matrix and dissimilarity matrix are similar in their content of information on the number of correlated data sets. Due to “conservation of weight” we have the above (approximate) relations for the extreme values and number of eigenvalues that are close to zero. In the limit of high correlation and $T \gg N$ these become equalities. In this example we have $N_1 = 20$ correlated among $N = 200$ random variables. For the dissimilarity matrix we find $\lambda^D_N = 186.15$ and $\lambda^D_1 = -17.57$ only approximately correspond to 200 and 19 respectively. $\lambda^D_N + \lambda^D_1 - N \approx 0$ holds well, however. The histograms are not normalised on purpose to convey the absolute counts.

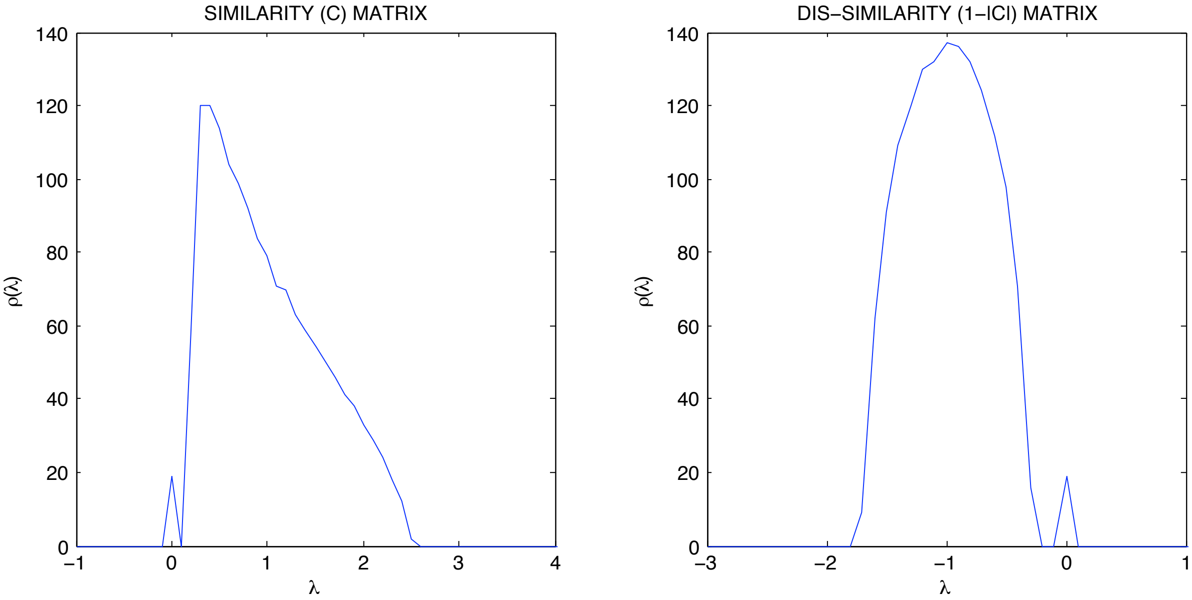


Figure 8.5: Spectrum of similarity and dissimilarity matrix doing the same experiment as in Fig. 8.4 but with larger numbers $N = 1500$ and $T = 4500$. Here the group of $N_1 - 1 = 19$ near zero eigenvalues can be identified easily.

relevant large elements in the indices of the correlated cluster proportional to their contribution to the respective "mode".

5. A group of $\mathcal{O}(N_1 - 1)$ small positive eigenvalues λ_1^C to $\lambda_{N_1-1}^C$ below λ_-^C. This group is well distinguishable due to a clear gap to the bulk above λ_-^C. See Fig. 8.5.

6. The eigenvectors $v_1^C, \ldots, v_{N_1-1}^C$ belonging to this group of small eigenvalues are well distinguishable by N_1 elements of higher variance in the element indices that belong to the correlated cluster. See Figs. 8.7. With increasing cluster correlation and matrix size the other elements of these vectors get asymptotically close to zero.

7. Due to conservation of weight and with increasing cluster correlation and matrix size all other eigenvectors $v_{N_1}^C, \ldots, v_{N-1}^C$ are asymptotically close to zero in the indices that *do not* belong to the correlated cluster. These small elements fluctuate below the variance of the random majority of the eigenvector matrix. Again see Fig. 8.7.

8. The random bulk of the eigenvector matrix may indeed fluctuate beyond the magnitude of the elements in V^C. This can be observed in Fig. 8.7.

B) Properties of the eigenvalues and vectors of diss-similarity matrix $\mathbf{D}$:

1. The uncorrelated spectrum is not definite due to one large positive eigenvalue. However, all other eigenvalues have a strictly negative upper bound

$\lambda_+^D < 0$.

2. Empirical conservation law in the limit of high correlation:

$$\lambda_N + \lambda_1 - N = 0. \tag{8.7}$$

   This not only holds if $1 \ll N_1 \ll N$ but as long as $N \ll T$ and $c \approx 1$.

3. One left large eigenvalue $\lambda_1^D = -\mathcal{O}(N_1)$.

4. The eigenvector V_{left}^D belonging to the left large eigenvalue $\lambda_1^D = -\mathcal{O}(N_1)$ contains N_1 large elements in the indices of the correlated variables.

5. A group of $\mathcal{O}(N_1 - 1)$ small eigenvalues λ_1^D to $\lambda_{N_1-1}^D$ asymptotically close to zero with increasing cluster correlation. This group is well distinguishable in the large size limit due to a clear gap to the bulk below λ_-^D. See Fig. 8.5.

6. The eigenvectors $v_1^D, \dots, v_{N_1-1}^D$ belonging to this group of small eigenvalues are well distinguishable by elements of high variance in the indices that belong to the correlated cluster and low variance below the random bulk variance. With increasing cluster correlation and matrix size the other elements are asymptotically close to zero. See Fig. 8.7.

7. One right large eigenvalue $\lambda_N^D = \mathcal{O}(N - N_1)$

8. The eigenvector V_{right}^D belonging to the right large eigenvalue $\lambda_N^D = \mathcal{O}(N - N_1)$ is $(1, \dots, 1)/\sqrt{N}$ plus some partly systematic fluctuation due to finite size.

9. As opposed to the similarity matrix, the eigenvectors with large fluctuations in the group indices are not situated at the far left end opposite to the vector V_{right}^D but a few columns closer. This property has no explanation yet.

10. The gap between the informationless bulk of the eigenvalues and the group of small eigenvalues around zero is wider than in the spectrum of the similarity matrix. See Fig. 8.5. It remains to be seen if this property is an advantage.

11. The random bulk of the eigenvector matrix may indeed fluctuate beyond the magnitude of the elements in V^C. This can be observed in Fig. 8.7.

The list of properties is not finished with the above items. Of particular interest is the linear combination of RWs created from the correlated noise and eigenvector information since the resulting “modes” allow recognition of similarities by eye that otherwise remain hidden in the series of increments. The artificially correlated cluster is produced via a fixed set of increments, called parent noise, that is reused for the production of all members of the cluster by adding more or less noise depending on the correlation parameter c, see Eq. (8.3). Fig. 8.8 shows in black the RW obtained by this “parent” noise that lies within the correlated cluster. The

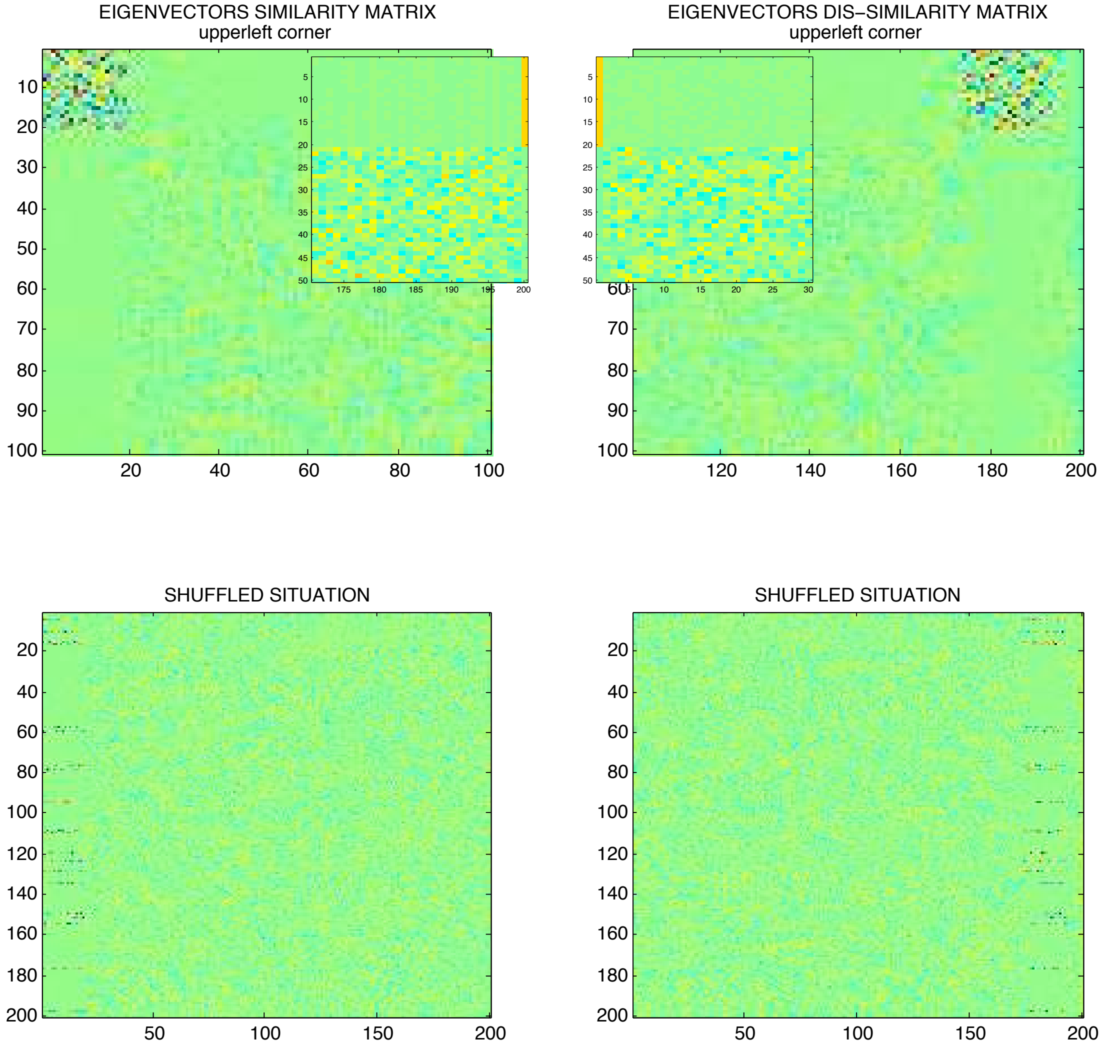


Figure 8.6: Colour coded eigenvector column matrix of the correlation matrix $\mathbf{C}$ and of the dissimilarity matrix $\mathbf{D}$. Shown are the magnified left and right corners. The $N_1 - 1$ eigenvectors belonging to the small group of eigenvalues evoked by the correlated cluster can be identified as a square of $N_1 \times N_1 - 1$ strongly fluctuating elements as compared to the informationless part of the eigenvectors. Due to the normalisation to length 1 the remaining elements in these vectors are close to zero (column of uniform green area). Also note that all other eigenvectors are essentially zero in the first $N_1 - 1$ rows as well. The shuffled situation is also shown below.

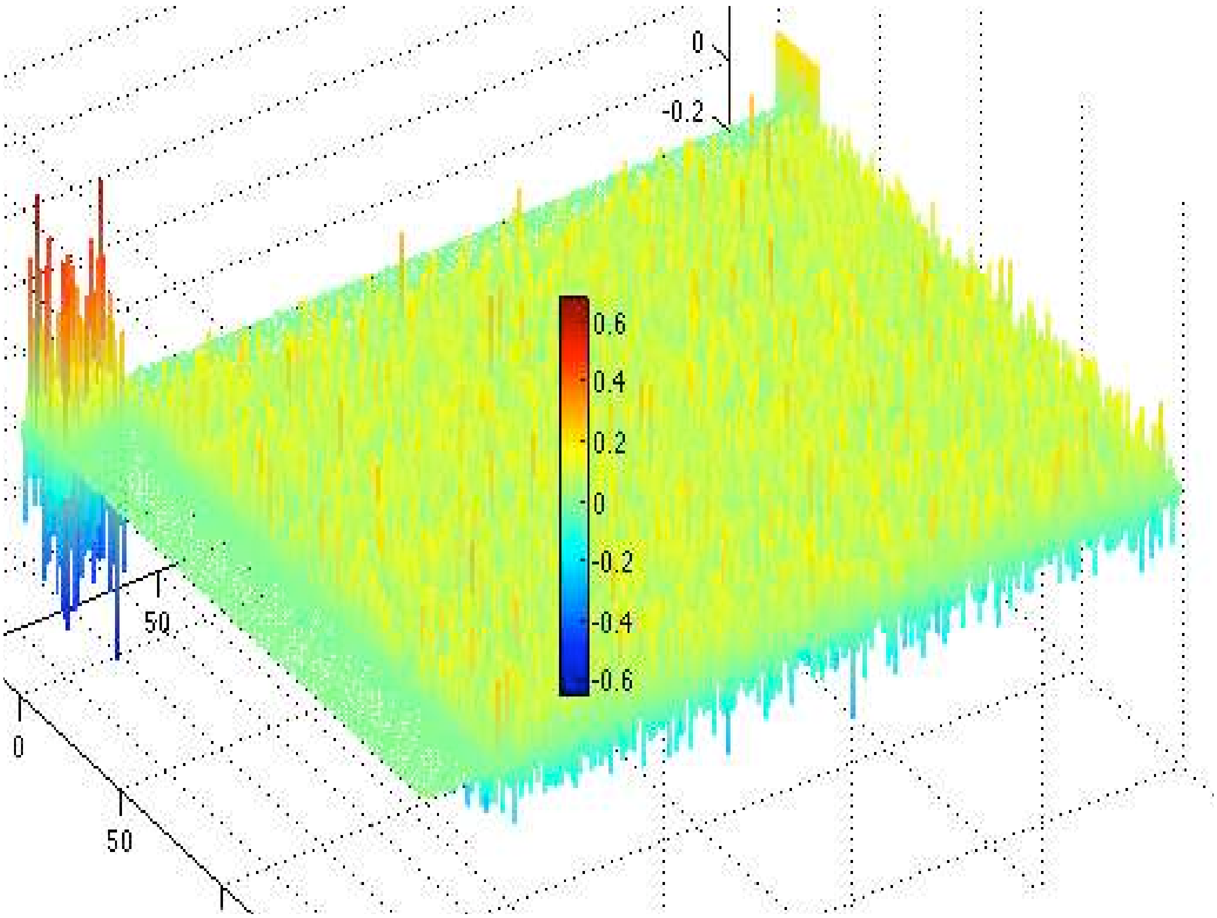


Figure 8.7: Impressionistic view of the eigenvector matrix of the similarity matrix **C** from Fig. 8.6. This perspective gives an impression of the overall structure and magnitude of the eigenvectors. Vector V^C can be recognised by the ridge of equally large positive elements at the right upper corner.

following notation, that is maintained later, assumes for simplicity's sake vectors if the entire range $t = 1, \dots, T$ is referenced, i.e. $x = (x(1), \dots, x(T))$. We start with the similarity matrix $\mathbf{C}$ and its two main objects of interest:

- The linear combination of RWs using the (in this case first) N_1 relevant large elements of eigenvector V^C as coefficients:

  $$x_{V^C} = \sum_{n=1}^{N_1} V^C_{(n)} x_n, \tag{8.8}$$

  where $V^C_{(n)}$ denotes vector element n which is also the index of the random variable. x_n is the RW-vector constructed from the respective increments:

  $$x_n(t) = \sum_{\tau=1}^{t} \xi_n(\tau). \tag{8.9}$$

  Even though the sum runs over the first N_1 vector elements of V^C the correlated mode is reproduced well. Fig. 8.8 shows the curve obtained via Eq. (8.8) as blue circles.

- The linear combination of RWs using the N_1 relevant elements of the $N_1 - 1$ eigenvectors v^C_n belonging to the small group of eigenvalues:

  $$x_{v^C_n} = \sum_{i=1}^{N_1} v^C_{n(i)} x_i, \quad n = 1, \dots, N_1 - 1 \tag{8.10}$$

  This gives $N_1 - 1$ different "modes". Fig. 8.8 shows these as several blue lines. Even though the sum uses only the first N_1 vector elements and RWs, the modes are reproduced well.

- Of possibly greatest interest could be the sum of all modes according to Eq. (8.8) and Eq. (8.10):

  $$\sum_{n=1}^{N_1-1} x_{v^C_n} \quad \text{blue squares in Fig. 8.8} \tag{8.11}$$

  and also the sum of *all* modes according to Eq. (8.10) plus Eq. (8.8)

  $$\sum_{n=1}^{N_1-1} x_{v^C_n} + x_{V^C} \quad \text{blue squares in Fig. 8.9.} \tag{8.12}$$

It is not clear yet which of the last two curves and with which factor or sign comes to sit exactly on the parent mode, or if there is any predictability at all. The difference between these two examples in Figs. 8.8 and 8.9 is only different seeds in the random numbers, i.e. they are different realisations. There are four possible outcomes:

a) Example realisation Fig. 8.8 requires Eq. (8.11).
b) Example realisation Fig. 8.9 requires Eq. (8.12). The match in the latter is better in this example, actually so good that it appears very unlikely that any other choice could be correct and that this is merely a coincidence.
c) In some other realisations it is a sign change that makes either Eq. (8.11) or Eq. (8.12) sit on the parent mode.
d) Other examples do not make it obvious which sign or factor is missing.

In any case, there do not seem to be intermediate cases in between. Overall, the modes are still the same up to a factor. Apart from that, it seems that the mode obtained by the elements of V^C does not play any special role within the modes obtained via each v_n^C, $n = 1, \dots, N_1 - 1$.

- Finally, it appears that one of the modes v_n^C, $n = 1, \dots, N_1$ contains the "zero-mode". Empirical examples indicate that up to some fluctuation that is decreasing with increasing T one possibly special v_n^C is essentially zero on the entire axis.

The entire exercise on linear combination of modes can be repeated with the eigenvectors of the dissimilarity matrix, but this has to be skipped for now since a detailed comparison requires a good amount of resources and has to be left to future work.

8.2.2 Two correlated clusters

Figs. 8.11 and 8.12 demonstrate how the situation of two correlated clusters shown in Fig. 8.10 is coded in the eigenvectors of the respective correlation matrix. With two clusters an additional feature appears: *both* eigenvectors belonging to the two large eigenvalues code the mode within their elements. The clusters are chosen such that the first (1–20) and the last (180–200) noise vectors are correlated to simplify identification by eye. First observe the following characteristics of this particular realisation of the respective RWs in Fig. 8.12:

1. We have two eigenvectors that code the two modes. As expected they are located at the right of the eigenvector matrix because they belong to two large eigenvalues of essentially equal value up to some noise due to the finite size of the situation.

2. Both eigenvectors are non-zero in the indices of *both* of the correlated clusters.

3. The sign of these non-zero elements is either entirely positive or entirely negative. The sign depends on the realisation, yet not all combinations are possible. Two signs in one group, e.g. 1–20, must be equal, the other in 180–200 two are then opposed.

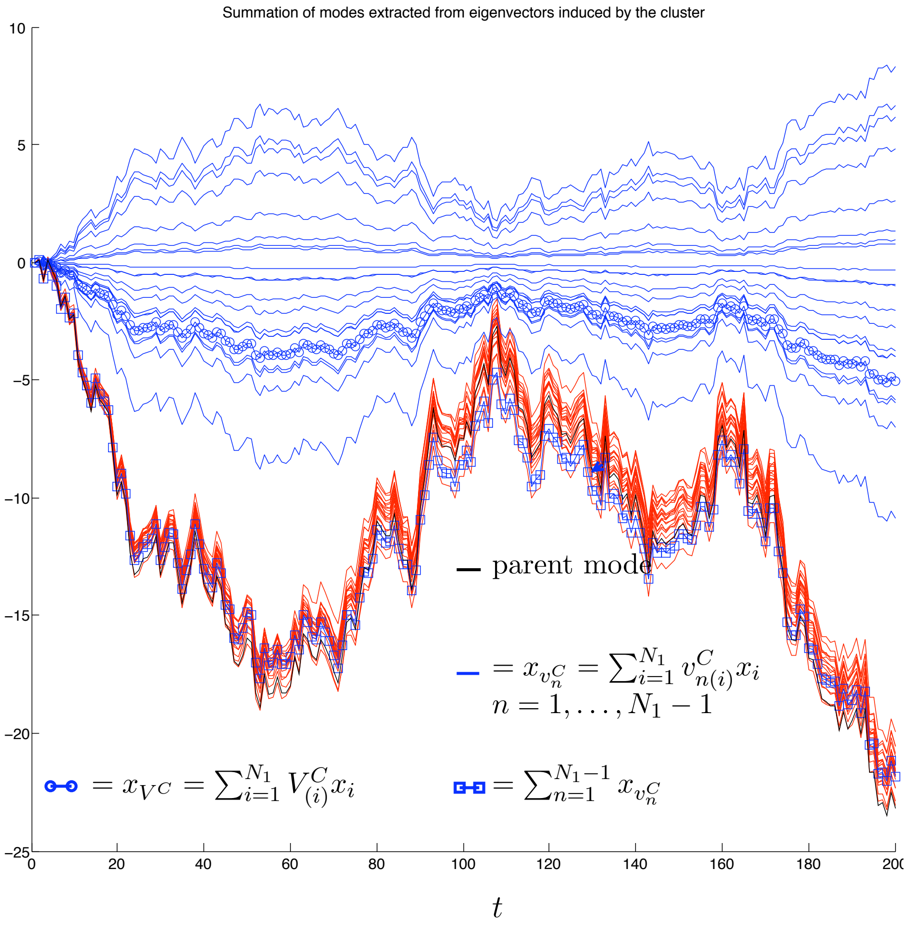


Figure 8.8: From the coefficients of eigenvectors that belong to eigenvalues outside the random bulk of the spectrum several "modes" can be reconstructed that are identical up to some statistical fluctuations. The parent mode is created with one pre-fixed set of noise used for generation of the correlated bunch (red), see Eq. (8.3). The legend gives the respective summation formulas. The notation uses the short cut $x = (x(1), ..., x(T))$ the subscript (i) denotes the ith vector element.

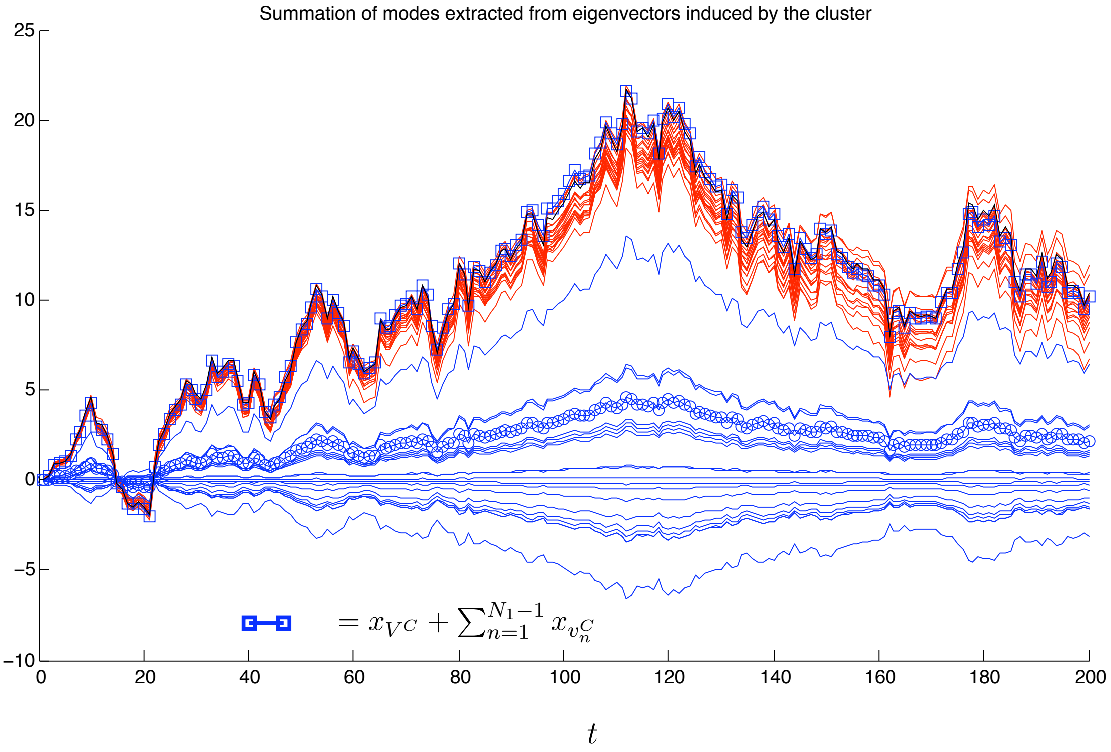


Figure 8.9: In this example the summation giving the blue squares includes the mode x_{V^C} in contrast to Fig. 8.8; the legend is the same. The result lies even closer to the parent mode. They match so well that one should assume that this construction is the right choice. The legend provides its summation formula. The notation uses the short cut $x = (x(1), \dots, x(T))$, the subscript (i) denotes the ith vector element.

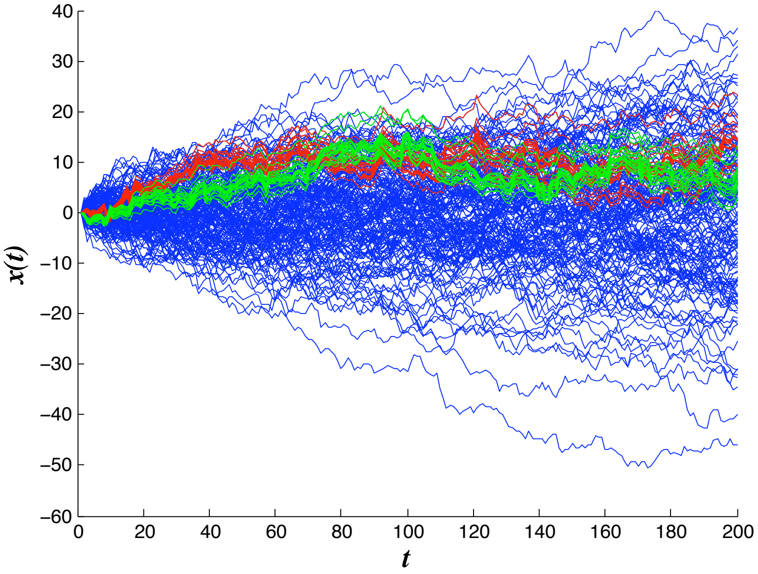


Figure 8.10: RWs with two independent correlated groups of random walks whose increments are correlated with the correlation coefficient $c_1 = c_2 = 0.8$.

Fig. 8.13 demonstrates in detail which eigenvector matrix elements recover the mode. The matrix elements are used for linear combinations of the corresponding RWs according to Eq. (8.11) or Eq. (8.12). The elements are marked in the schematic overview Fig. 8.14 and can be recognised in Fig. 8.12 as columns. It is apparent that the respective recovered modes are cluster-wise numerically identical up to an overall factor. The reconstructions approximate the parent cluster mode Eq. (8.3). With correlation coefficient $c_1 = c_2 = 0.8$, i.e. equal for both clusters, we do not expect a perfect recovery of the parent mode.

8.3 On the distribution of eigenvector elements

In the random uncorrelated case the elements of eigenvectors $u_{(i)}$ taken from the correlation matrix follow the Porter-Thomas law which is essentially a Gaussian with variance $\sigma^2 = 1$:

$$P(u) = \frac{1}{\sqrt{2\pi}} e^{u^2/2}. \tag{8.13}$$

The figures in Sec. 6 provide examples of how eigenvectors that carry information deviate from this law. It has been said that the comparison of the distribution of vector elements with Eq. (8.13) alone does not suffice to reveal informative structure. The cumulated information of a histogram easily hides non-random structure, i.e. some information content.

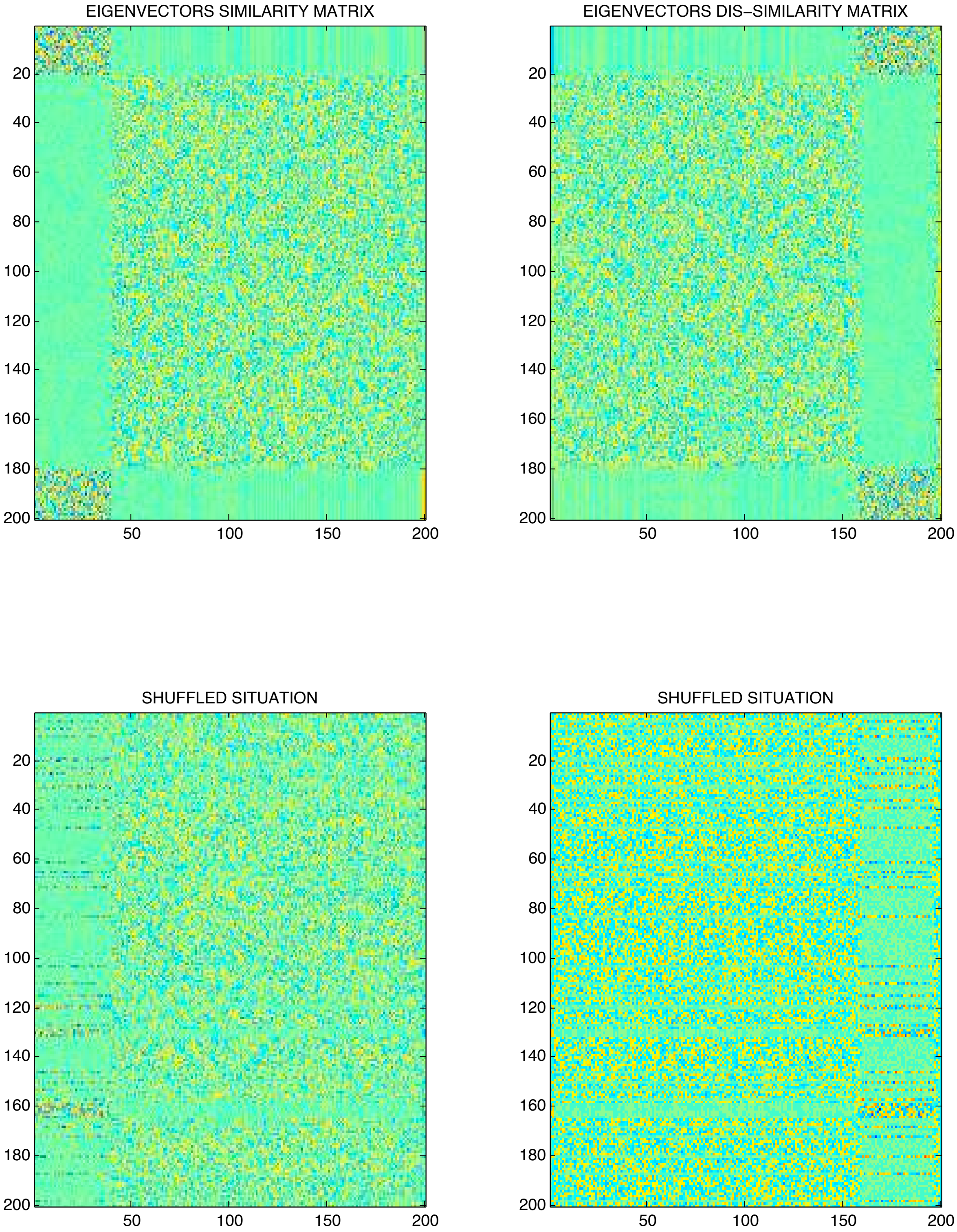


Figure 8.11: Colour coded eigenvector column matrix of the correlation matrix **C** and of the dissimilarity matrix **D**.

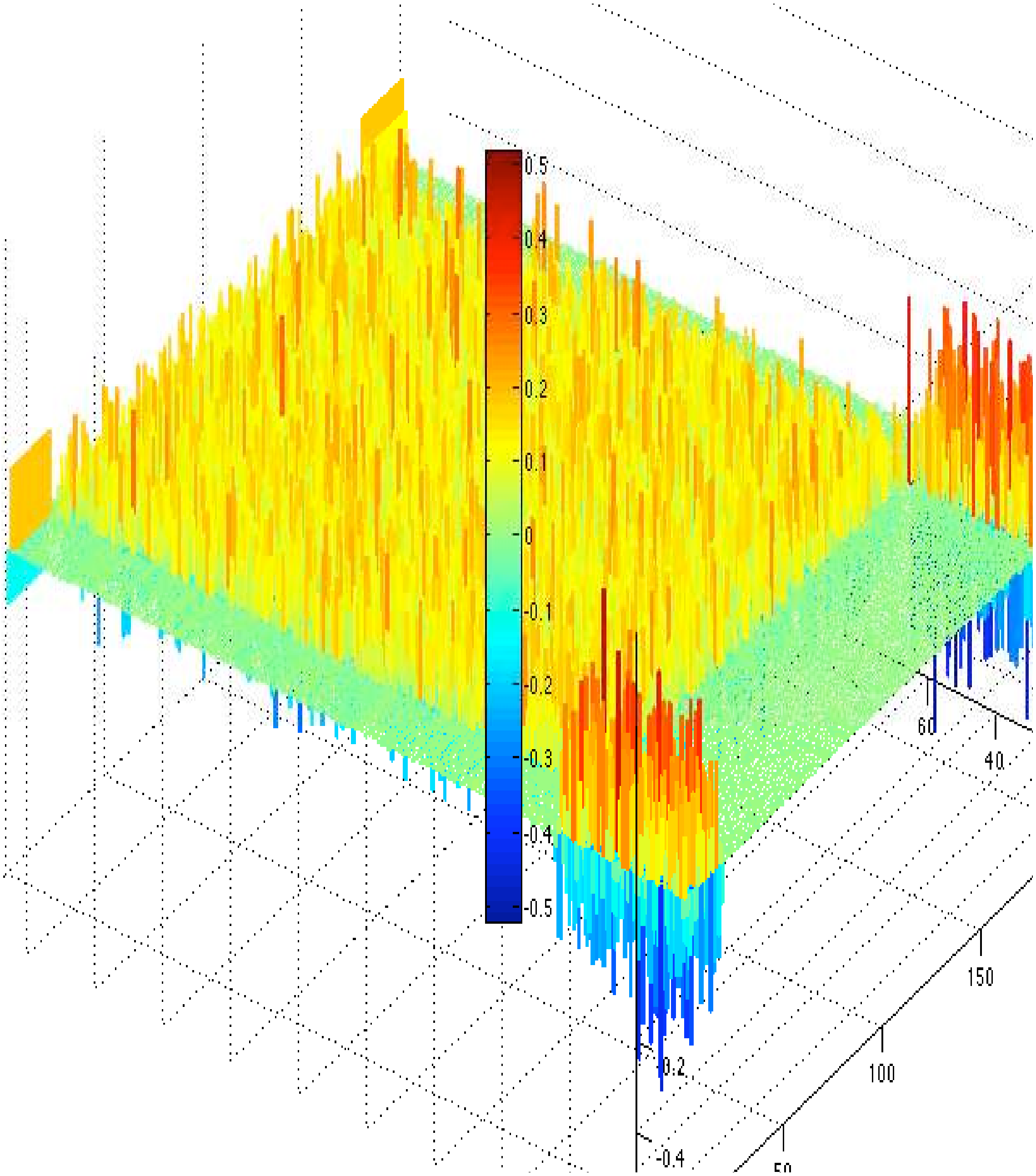


Figure 8.12: Colour coded eigenvector matrix of the correlation matrix **C**. This perspective shows that the two eigenvectors belonging to the two large eigenvalues are non-zero in the indices of both correlated random variables. The regions at the “other end” of the matrix also “code” the respective modes.

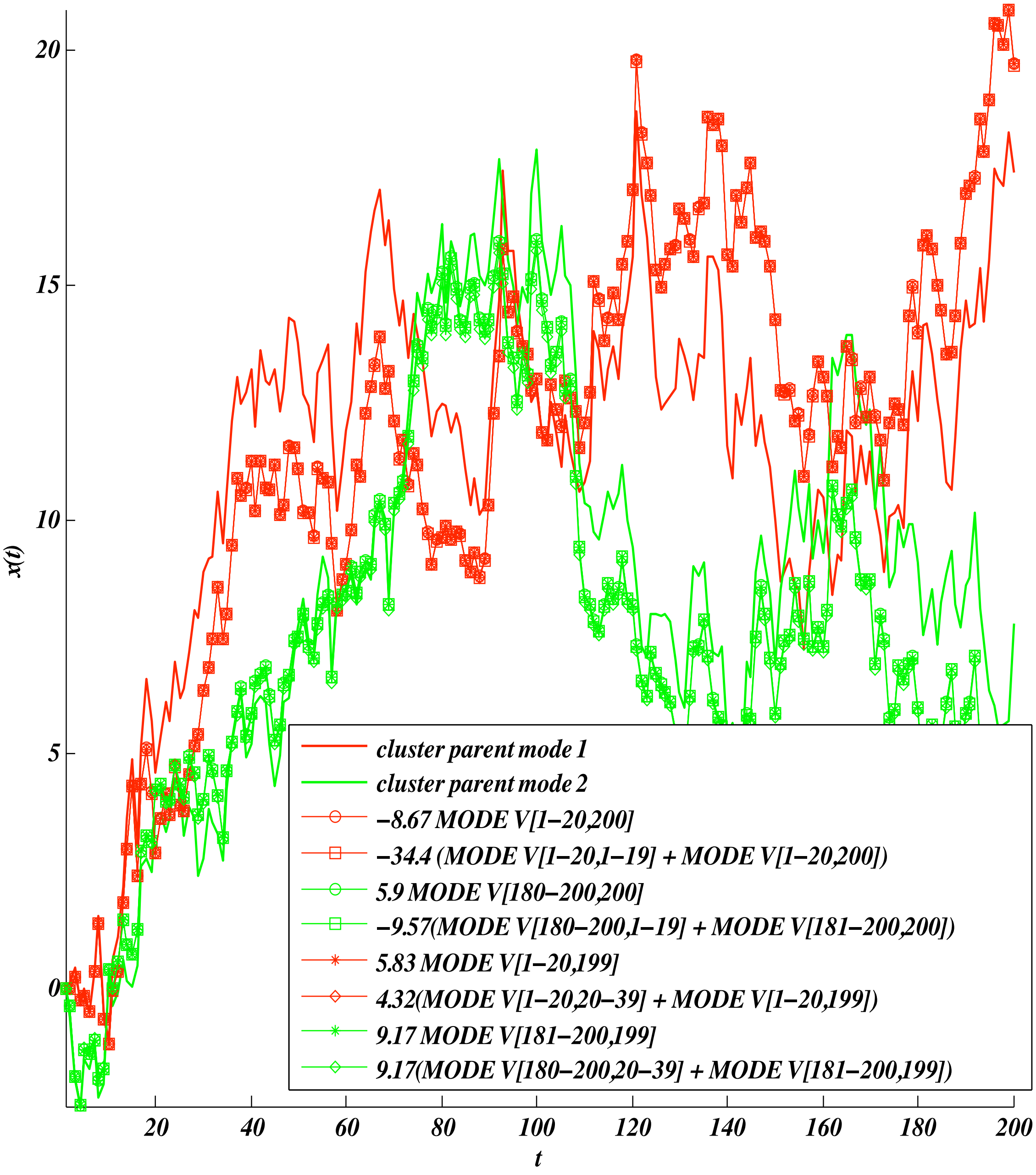


Figure 8.13: The two “parent” modes used for construction of two artificial clusters and the reconstructed modes are shown as red and green continuous lines. The notation MODE V$[i-j, k-l]$ denotes the eigenvector matrix elements used in the mode reconstruction by linear combination of the respective RWs according to Eq. (8.11) or Eq. (8.12). The factors are empirical. In this example we have 20 possibilities to reconstruct exactly the same mode up to a linear factor. These reconstructions are identical but only approximate the parent mode.

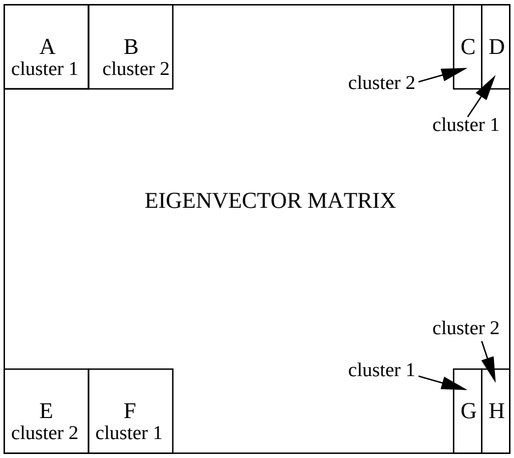


Figure 8.14: In the case of two correlated clusters of random variables in the indices (1–20) and (180–200) the regions in the eigenvector matrix of the correlation matrix whose elements code the respective mode are known beforehand up to an arbitrary ordering of clusters 1 and 2. All eigenvectors that stem from a large eigenvalue contain all modes. In the realistic, or shuffled, situation all rows would be reordered in a random way but the overall order within the columns remains.

8.4 Improved spectral clustering

We make use of the fact that the small eigenvalues belonging to a mode also code the mode up to a factor. In contrast to, for example, k-means the following procedure has the advantage of estimating the number of clusters and it also allows for cluster overlaps.

Algorithm:

1. Decide on the number of significant modes (or clusters) by counting the number k of "large" eigenvalues, number of significant eigenvectors, etc. using a criterion of choice.
2. Take the corresponding k eigenvectors. For all k vectors and do:
3. "Plot" the mode coded in V_k via Eq. (8.8).
4. Decide for V_k which η_k largest elements are significant, e.g. by comparing with the Porter-Thomas law or via contribution to the resulting mode, etc..
5. Find among all vectors the corresponding set $\{v_{ki}\}$ of size $\eta_k - 1$ (region A, B, etc.) There are different ways: E.g. try to recover the mode coded in V_k doing many fits with linear combinations as in Eq. (8.10).
6. Test if a different number than $\eta_k - 1$ gives the best fit by repeating 5 trying an additional v^*_{ki} (consequently also an additional vector element in all v_{ki} as well as v^*_{ki}). Then η_k, the number of significant elements in the large eigenvalue's eigenvector, can be adjusted.
7. Remember these eigenvector element indices as the cluster $\{N_k\}$.
8. Repeat at 2).
9. We are left with k sets of numbers. Some may overlap.

The first, main, improvement is located in steps 5 and 6. We use η_k eigenvectors to determine the number of correlated random variables in cluster k instead of only the one eigenvector V_k belonging to the kth-largest eigenvalue. Furthermore, random matrix theory suggests a threshold and other criteria above which eigenvalues are significant, thus giving an estimate of the contained clusters. Third, the above classification allows overlapping clusters which is meaningful.

Remark on step 1) One can also select different complementary criteria. For example, a measure that tells how significant an eigenvector that belongs to a potentially large significant eigenvalue departs from the Porter-Thomas law. Another criterion checks for any structure in the eigenvector even though the distribution of elements is Porter-Thomas. That this criterion is not redundant has been shown in Sec. 6. Furthermore, we have seen there that *within* the Marčenko-Pastur bound close to the right edge the eigenvectors are not necessarily informationless.

Remark on step 5) There are two pieces of information in the eigenvector matrix that help in this task:
a) We can use the η_k significant elements determined above of each vector v_{ki} (of which there are $\eta_k - 1$). Of great help is the fact that the relevant elements of *all* the to-be-found v_{ki} are located in the *same index*, i.e. row, of the relevant (=large) elements in V_k. This fact can be observed also in reality as ridges of high (color) variance in the reshuffled version of the eigenvector matrices in Fig. 8.6. As shown in examples with more clusters there is no ambiguity. So, in the end, the found v_{ki} arranged next to each other form a rectangle of size $\eta_k \times (\eta_k - 1)$ of high (colour) variance. The remainder is made up of only small elements.
b) The number η_k of relevant (not necessarily large!) elements in the regions A, B, E, F, etc. is equal to the number of significant vectors η_k minus one! The best fit must be in accordance with this, otherwise the algorithm has to switch to step 6 again.

Remark on step 6) In other words: Try to add an additional vector element to the set of relevant vector elements in V_k, then see what happens. This implies that an additional vector v^*_{ki} has to be selected as well as an additional relevant vector element in *all* other v_{ki} so far selected. (This of course includes the additional v^*_{ki}). So we use an increased rectangle A, B, etc.. If the fit of all coded modes does not improve then the best clustering was reached. Furthermore, a “quality” value can be chosen that associates the best clustering with the smallest overlap between the clusters or anything else that is suitable for the application.

With the above procedure the accidental clustering of the three time series in Fig. 8.1 does not happen. The red time series in panel B is not included in the cluster with the other two.

Note that in this field the widely used Kernighan-Lin algorithm [103] is the best example of a purely empirical clustering method that is justified by achieving satisfactory results in practice.

8.5 Scenario 2 – Un-correlated noise with more variables than measurements per variable

The Marčenko-Pastur theory for uncorrelated independent Gaussian noise in the Wishart matrix ensemble was developed for $T > N$ in the large size limit, i.e. more realisations per random variable than variables. It has been recognised only recently by Lehmann [94] that for $T < N$ the theory persists essentially unchanged. In particular, for $T \ll N$ the limiting distribution is the Wigner semi-circle law. The difference is a shift of variables and a delta-contribution of zero eigenvalues.

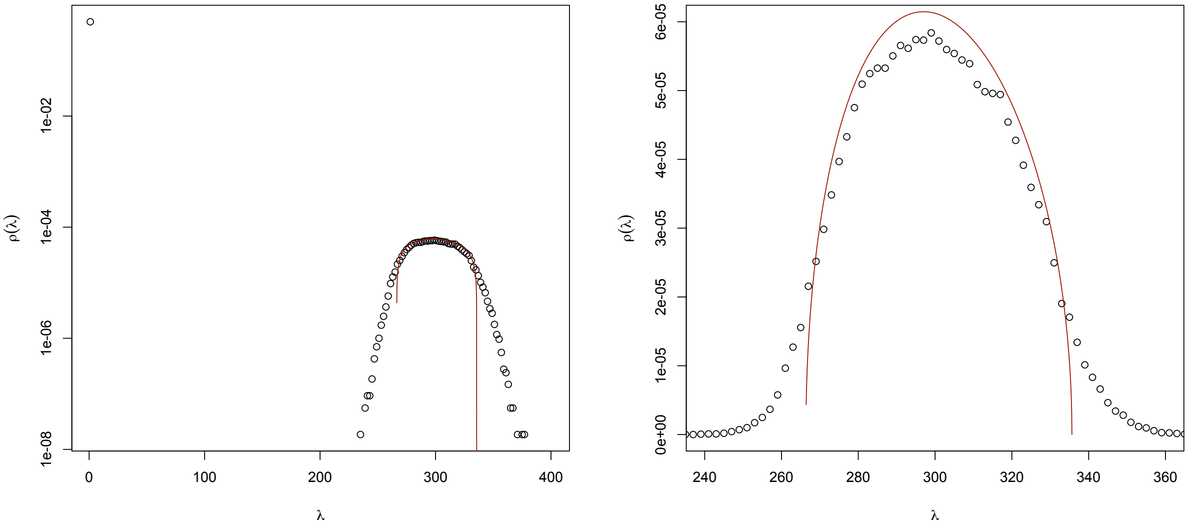


Figure 8.15: Spectrum averaged from 12000 Wishart matrices $\mathbf{A}\mathbf{A}^T$ where $\mathbf{A}$ is $N \times T = 900 \times 3$ with uncorrelated Gaussian noise. In the limit $N \rightarrow \infty$ the histogram approaches the shifted and rescaled GOE spectral density of 3×3 matrices, i.e. the Wigner semi-circle. The single data point at $\lambda = 0$ marks the delta-function in Eq. (8.14) representing many zero eigenvalues.

With $m = N/T$ the eigenvalue density can be expressed as

$$\rho(\lambda) = \frac{1}{2\pi\lambda}\sqrt{4m - (1 + m - m\lambda)^2} + \delta(\lambda)(1 - m), \quad \lambda \in (\lambda_-, \lambda_+), \tag{8.14}$$

$$\lambda_\pm = (1 \pm \sqrt{m})^2/m. \tag{8.15}$$

This law is intuitively understandable since the number of non-zero eigenvalues in the product of two iid random matrices is the rank given by $\min(T, N)$. The non-zero eigenvalues resulting from the matrix products $\mathbf{A}^T\mathbf{A}$ or $\mathbf{A}\mathbf{A}^T$ with rectangular $\mathbf{A}$ made of independent random variables are even numerically identical up to a global normalisation factor. Eq. (8.14) contains a delta-function that represents the zero eigenvalues. Since the eigenvalue density of a GOE random matrix is independent of the matrix size beyond ca. $N, T > 50$, the matrix products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are equivalent in the non-zero part of the spectrum.

Formula (8.14) is independent of N and T and in this section we deal with $T \ll N$. To minimise finite size effects in obtaining the Marčenko-Pastur density in the non-zero part of the spectrum N has to be chosen quite large to allow T to be sufficiently large to achieve $T \ll N$. Eq. (8.14) can be reconstructed in numerical experiments, the spectra are shown in Fig. 8.15 with $T = 3$ and $N = 900$. The logarithmic y-scale allows the Delta-function at $\lambda = 0$ to be observable together with the non-zero part of the spectrum. In the linear plot

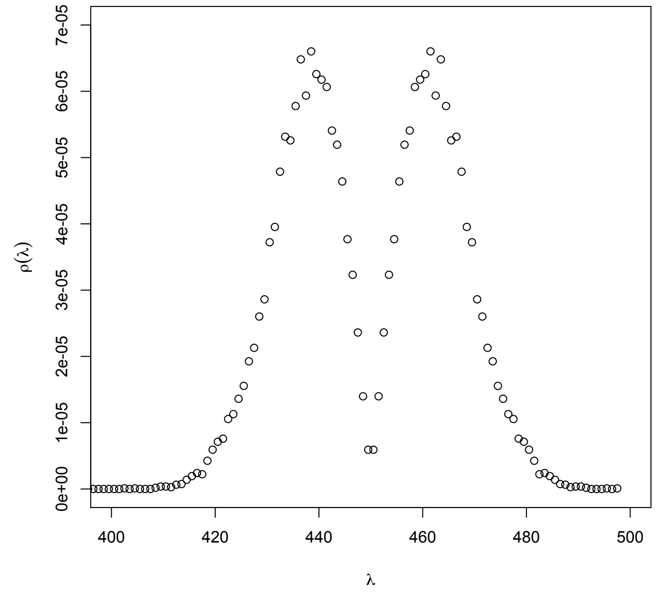


Figure 8.16: Non-zero part of the spectrum averaged from 12000 sample correlation coefficient matrices with uncorrelated noise; $N = 900,\ T = 3$.

the shape of the Marčenko-Pastur law can be recognised but it lies very far out as compared to the standard case with $T \geq N$. In the following most examples are based on the choice $m = N/T = 900/3$ according to the example presented in Lehmann [94]. Larger values for T are better calculated on a parallel machine or with a lot of time because many realisations are necessary to obtain a reasonable accuracy in the histograms.

In respective literature several fundamental results in random matrix theory are presented as candidates for a correspondence with some real world situation or some mathematical/statistical object that cannot easily be calculated. The same is the case with the above result by Lehmann. The point of view is simply carried over from applications in finance where the Wishart matrix ensemble is considered to be sufficiently close to the respective correlation matrix. However:

> The matrix of *sample* correlation coefficients is an inappropriate approximation of the Wishart matrix ensemble if T is small such that the true and realised mean and variance of iid noise differ significantly.

The statement above can be restated as follows: Assume $\xi_1(t), \dots, \xi_N(t)$ to be iid and uncorrelated random variables with zero mean and standard deviation $\sigma = 1$

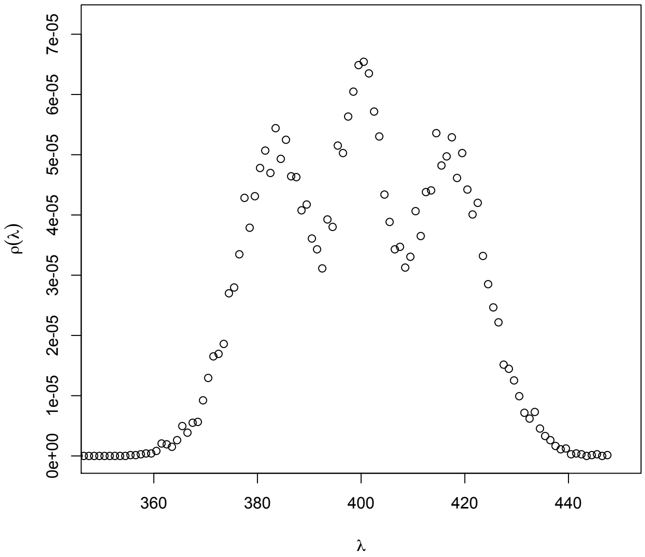


Figure 8.17: Non-zero part of the spectrum averaged from 6000 sample correlation coefficient matrices with uncorrelated noise; $N = 1200,\ T = 4$.

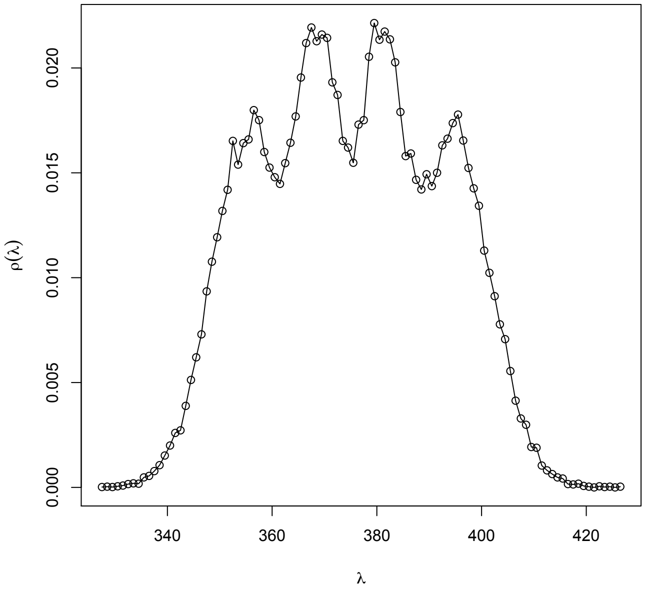


Figure 8.18: Non-zero part of the spectrum averaged from 15000 sample correlation coefficient matrices with uncorrelated noise; $N = 1500,\ T = 5$. The data points are connected with lines to make the series of maxima more apparent.

and realisation index t. Then the matrix of sample correlation coefficients

$$C_{ij} = \frac{\sum_t (\xi_i(t) - \bar{\xi}_i)(\xi_j(t) - \bar{\xi}_j)}{(T-1)\mathrm{sd}(\xi_i)\mathrm{sd}(\xi_j)} \tag{8.16}$$

with sample standard deviation $\mathrm{sd}(\xi)$ and sample mean $\bar{\xi}$ is significantly different from the Wishart ensemble

$$\mathbf{C} = \frac{1}{T}\mathbf{M}\mathbf{M}^{\mathsf{T}} \tag{8.17}$$

according to the definition of $\mathbf{M}$ as in Eq. (3.2). This is simply due to the insufficient number of realisations to obtain a good estimate of mean and variance. The spectrum produced via the sample correlations for the same data as in Fig. 8.15 is shown in Fig. 8.16. The differences are peculiar:
1) The approximate middle spectrum is shifted to a higher value of $\lambda = 300 \rightarrow \lambda = 450$.
2) The spectrum consists of two disconnected parts.
One can show, at least empirically by analysing the eigenvalues directly, that there is no eigenvalue falling into the point region between the two supports of the halves in this example.

How can this be explained? We first observe that for $N \nearrow T$ up to $N = T$ the matrix rank of the sample correlation coefficient matrix is not full but exactly one less than the number of independent rows or columns in the Wishart ensemble. Looking at the list of eigenvalues we discover one numerically zero eigenvalue below the (empirical) Marčenko-Pastur spectrum, which is strictly positive definite. We get in increasing order:

```
from sample corr. matrix    from Wishart ensemble

  5.392423500457690e-17     1.176099194760301e-06
  1.345617510432624e-06     6.217387115086680e-06
  1.261396309228272e-05     2.086802587513426e-05
  2.459788786131458e-05     2.636339685910131e-05
  2.985884814656815e-05     4.599760096704007e-05
  5.245327169211368e-05     8.182996374009664e-05
  1.371004860632502e-04     1.405440978395324e-04
  1.538041928161666e-04     1.640231118492298e-04
      ...                        ...
```

The lists above are the sorted eigenvalue outputs with $N = T = 900$. The strict positivity in the Wishart ensemble is empirically expressed in very small but clearly numerically non-zero eigenvalues down to 10^{-7}. A zero eigenvalue is as unlikely as it is unlikely to find a perfectly uncorrelated series of length T, i.e. with realised covariance equal to zero. With $N = 900$, $T = 901$ the zero eigenvalue in the column for the sample correlation matrix disappears. We can also measure the rank against the number of random variable realisations T. See Table 8.1. The last three rows of the table correspond to Figs. 8.15, 8.17, and 8.18. These three spectra are calculated for fixed $m = N/T = 900/3 = 1200/4 = 1500/5$. The

N	T	rank sample correlation matrix	rank $\mathbf{C} = \frac{1}{T}\mathbf{M}\mathbf{M}^{\mathsf{T}}$
900	902	900	900
900	901	900	900
900	900	899	900
900	899	898	899
900	898	897	898
1500	5	4	5
1200	4	3	4
900	3	2	3

Table 8.1: With the transition $T > N$ to $T \leq N$ the rank of the sample correlation matrix drops earlier and persistently by one than the rank of the Wishart matrix.

rank of the matrix is reflected in the number of independent rows and columns as the QR-decomposition used for these calculations is numerically able to tell. The output is consistent with the observed eigenvalues. With increasing T and constant m the number of maxima increases and we can expect the spectrum to converge against the analytic prediction for the Wishart ensemble because the position of the spectrum wanders closer to the position of the analytic curve and the number of maxima increases.

So far, these are empirical facts that seem to be relevant. The explanation of why exactly the maxima and the shift appear is still elusive. A guess about the shift is that the normalisation by realised means and variances introduces correlation, on the average, in all pairs of variables.

Whether microarray experiments with many more genes than expression values can benefit from the results of random matrix theory has to be seen. Since we have at best a few microarrays, each giving about 10 expression values we get very few non-zero eigenvalues for comparison with the null-hypothesis spectrum above. The nuisances of microarray experiments and data analysis is an epic in itself. What can be claimed already is that no single experiment with two orders of magnitude more genes than experiments can contain the information to separate more than very few functional groups. And even if the data was perfect in the sense of no technical and biological variance and with only 3–4 functional groups that reveal themselves perfectly in significant up (and down?) of the genes, which would be biologically quite a luxury, then the groups could be identified by eye in the data already.

8.6 Scenario 3 – Correlated noise with more variables than measurements per variable

The time series data from Fig. 8.10 can be used as a starting point for making T smaller than N. First we prune the data by exactly one data point, i.e. $T =$

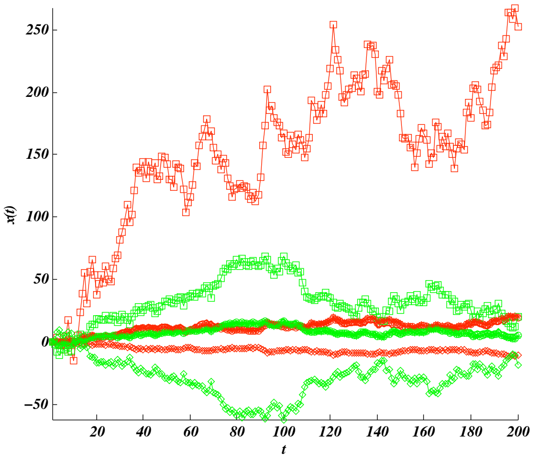


Figure 8.19: Exactly the same data as in Fig. 8.13 except that the last data point is cut off, i.e. $T = 199$. Despite this small change, all factors on the reconstructed mode do not fit anymore. This is caused by reordering of nearby eigenvalues and their eigenvectors.

$200 \searrow 199$. The procedure for the reconstruction of modes as well as the factors from Fig. 8.13 are kept identical. The result can be seen in Fig. 8.19, all factors are not in the right place anymore. Despite the small change in T and despite all eigenvectors being normalised to unit length, all factors on the reconstructed mode do not fit anymore. Indeed, we have no influence on the ordering of the two large eigenvalues or which mode the nth eigenvector happens to capture. This is the usual problem of unique identification of eigenvalues very similar to electronic band structure calculation, where band crossings are not distinguishable from small band gaps [60].

The next demonstration is more extreme with $T = 100$, keeping $N = 200$. We now show the eigenvector matrix of the correlation (similarity) matrix and of the dissimilarity matrix $D = 1 - |C|$, see Fig. 8.20 showing both from the same perspective. First note one eigenvector $(1, \dots, 1)/\sqrt{N}$ in the highest index of the dissimilarity matrix. The explanation is analogous to why the Laplacian also always has $(1, \dots, 1)$ as an eigenvector, see Ref. [133]. At this point it also becomes apparent that for $T < N$ the behaviour of the two matrices departs more significantly and the detailed analysis of the vectors is left open for future work.

8.7 Intermediate discussion

Usually the matrix rank is equal to the number of non-zero eigenvalues and is considered an integer number. For some purposes or applications this view is possibly too restrictive. What happens in the above scenario when two out of many time series are correlated? We get a large and two small eigenvalues growing out of the Marčenko-Pastur spectrum of eigenvalues that represent uncorrelated noise. The

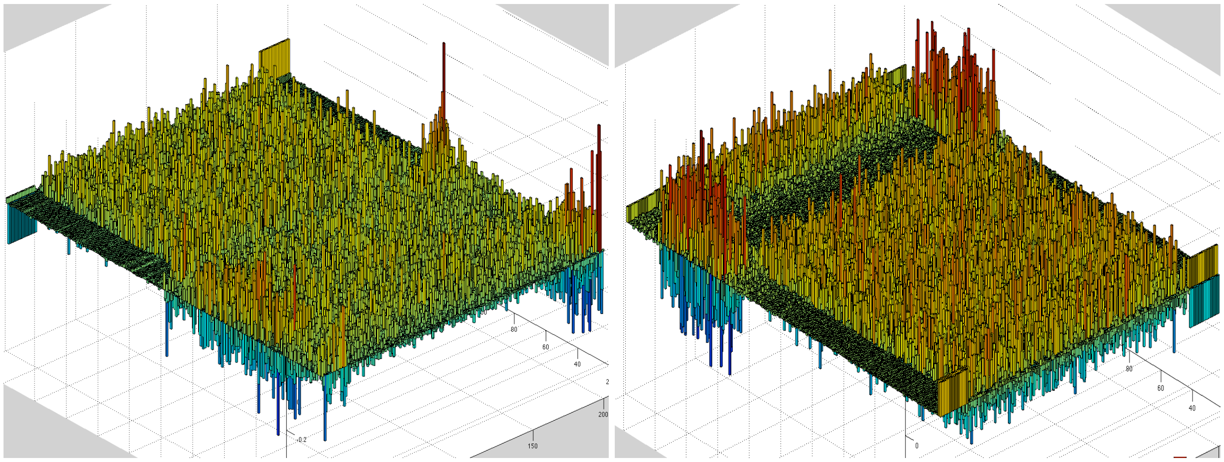


Figure 8.20: Eigenvector matrices of the similarity (correlation) matrix $\mathbf{C}$ (left) and the dissimilarity matrix $1 - |\mathbf{C}|$ (right). With $N = 200$ and $T = 100$ the two correlated clusters cannot be easily recovered in the vectors anymore. The behaviour of both eigenvectors is rather different, too. In the right picture the clusters are not even expressed symmetrically.

two small eigenvalues move towards zero with increasing correlation. With perfect correlation they are zero and the matrix rank suddenly drops by two. On the way however, the continuous transition can be regarded as a continuous reduction of matrix rank of the associated Wishart matrix by correlation or “similarisation”. The details of this process do not not matter as long as it is continuous and we reach equality, i.e. perfect correlation, at the end.

Whether the equality $T = N$ is a special barrier, at least in the Wishart ensemble, depends on the questions asked. We have seen in the previous section that sudden changes occur in certain measures like the matrix rank. The less information with decreasing T, the less eigenvalues and eigenvectors are able to capture the behaviour. With $T \ll N$ the data can hardly be distinguished from randomness. We saw in Fig. 8.20 that the “disturbance” of randomness caused by correlations spreads out to more eigenvectors than originally contained in the artificial clusters. This can also be observed with $T = N$ and more so in the dissimilarity matrix: A closer look at the eigenvector matrices in the regions A, B, E, F as defined in Fig. 8.14 reveals that the fluctuations decay slowly towards increasing column index. This is coincidental misinterpretation by the correlation estimator of a random walk as correlated with a cluster or vice versa. Neither are the edges of the Marčenko-Pastur spectrum hard. The dissimilarity matrix seems less prone to such false-positive and false-negative measurement because it disregards information due to the equal treatment of correlation and anti-correlation.

8.8 Genetic profile scenario of microarray data on differential expressions

In the previous sections the problem of extracting clusters from correlation matrices was discussed. In microarray experiments it is common to deal with values of differential expressions

$$x_i = \log\left(\frac{I_{\text{red}}}{I_{\text{green}}}\right) \tag{8.18}$$

for each gene i. The values I_{colour} are dye intensities on the array wafer that code the number of detected RNA molecules. One of the two usually contains the "control experiment" as a reference base. Any deviation from zero indicates that the gene is differentially expressed with respect to the reference value. This is the extremely idealised situation that is actually far from reality. Such is the point of view of random matrix theory, though. Likewise, we assume that the deviations around zero are linear and symmetric, etc. It is common in practice to work only on a subset of approximately 300–500 genes that are expected to contain one or more functionally related genes that are "provoked" to express differentially, i.e. to change expression levels, under certain possibly changing experimental conditions. The term "experiment" is in this case used on a more general level and can refer to different spots on the wafer or different wavers containing replicas or data from different biological conditions. In both cases the types of systematic errors are different. Commonly, the number of experiments under different experimental conditions is often restricted below 10, mostly due to financial limits. There are three scenarios that can be considered:

0) The true null-situation. All I_{colour}, where colour $\in$ {red, green}, contain the same experimental condition. Any extracted pattern is some systematic error. This test is non-trivial since the different dyes behave differently. The house-keeping genes in particular, which inadvertently could also be regulated, must pass the test.

1) We are given the values $x_i(t)$ where $t = 1, \ldots, T$ indexes the experiments with the *same* experimental condition, i.e. replicas. One may assume that the values fluctuate randomly and independently around the same (*expectation?*) value due to technical or biological variance. This scenario can be considered as a measure to extract *at least some* possible systematic errors that are introduced technically or biologically and influence the measurement within the same experiment index t. This situation should correspond to the null-hypothesis of uncorrelated random variables. A grand un-provable theory of experimental physics states that in a chain of errors the resulting final error is most likely Gaussian. Therefore, we have a good chance that this will also be the case here and the scenario of Bessel distributed random numbers

from Sec. 7 is repeated if $X, Y \sim N(0, \sigma)$ and we deal with one of the above null-situations.

2) We are given the values $x_i(t)$ where $t = 1, \ldots, T$ indexes the experiment with *different* experimental condition, for example increasing cell stress or anything biologically sensible that is hoped to provoke expression changes in certain functional groups. One may assume that the expression values do carry the expected information if they are biologically expected to do so by the experimentalists. The abstractional step from microarrays to realisations of random variables is debatable in this case. We expect the differential expression $x_i(t)$ to have a functional relationship with t. If t indexes the experiment number with increasing cell stress level or any other meaningful condition, then subsequent values $x(t_j)$, $x(t_{j+1})$ are highly dependent, probably monotonic and possibly even nearly linear in t if the cell stress is increased slowly. The latter would implicate that $\Delta x(t) = \text{const}$, leading to scenario 1) after normalisation to $\langle \Delta x(t) \rangle = 0$. This experimental setting of "small" changes is likely to achieve the initial goal of controlled differential expression of the same set of functional groups best. Yet since this ideal situation is not to be expected this second scenario is probably still a distinct case.

In the light of the previous section but also by considering scenario 2 it can be questioned in how far it might be sensible to normalise the variance and the means of the expressions for one gene with the sample values or any other value. In addition, it is debatable whether to take the increase of the differential expression level as the to-be-correlated variable or maybe $x_i(t)$ directly. As we have seen, the realised variance, mean and probably other statistical measures of interest, loose their meaning with extremely small number of realisations. The ideal genetic profiling draws no information from additional experiments (e.g. increased cell stress) if the relation between $x_i(t_j)$ and $x_i(t_{j+1})$ is (ideally) nearly linear. This poses a paradox since the microarray business considers many experiments as beneficial. This line of reasoning however, leads to fundamental debates about the current view on microarray experiments and the information they can contain and well as which statistical prerequisites/algorithms to use. The information contained in the data set $x_i(t_j = 1, \ldots, T)$ is then essentially the slope independent of T. Thus, the information contained in the data set can be coded entirely into one matrix element.

Random matrix theory only accounts for equal-time or "equal-experiment-index" correlation. This statement clearly points to the previous section on the reconstruction of modes while it is yet unclear how to deal with the situation of reduced matrix rank. The following statement may arise from the reasoning above:

Scenario 2 provides highly dependent realisations, possibly even linear, of variable $x_i(t)$ with index $t_j \rightarrow t_{j+1}$. As opposed to time, the difference $t_{j+1} - t_j$, e.g. coding cell stress level, is not meaningless. For such a highly systematic situation the ordering sequence $t_1, \ldots, t_T$ cannot be disregarded. It would be inappropriate for any analysis to ignore this dependency, thus to stick to the random matrix point of view alone.

Nevertheless, for the null-situations depicted above we can still perform some sand box simulations with the luxury of nearly infinite sand in order to be able to calculate densities for the sample correlation matrix ensemble. The spectral density for the null-situation with six experiments is shown in Fig. 8.21 with unknown variances and means, i.e. we are dealing with the sample correlation matrix ensemble. Note that this density is obtained from averaging over many histograms. If variances and means are known (somehow) we get the Wishart ensemble back. Fig. 8.22 shows an artificial experiment to create a possible scenario in microarray data analysis with differential expression of a subset of replicas with high correlation:

- $T = 6$ number of realisations per random variable
- $N = 300$ number of random variables
- $N_c = 1$ number of independently correlated groups of variables
- $N_1 = 20$ number of correlated variables
- $\xi_n(t)$ Gaussian noise data set n where $t = 1, \ldots, T$, $n = 1, \ldots, N$.
- Type of correlation within group the group:

  $$\xi_n(t) = \xi_n(1-c) + \Xi(t)c \tag{8.19}$$

  Ξ is a prefixed “parent” noise vector specific for the correlated group. $c \in [0, 1]$ is a correlation coefficient. Ξ is identical for all realisations!
- $c = 0.9$ (very high correlation)

The correlated cluster induces a bump of eigenvalues on the right of the null-hypothesis spectrum. Despite the extremely high correlation coefficient the additional bump does not lie far outside the uncorrelated null-spectrum. We conclude from this that the correlation measurement, whether via Wishart or sample correlation matrix ensemble, is rather insensitive. In practice we would have only one single eigenvalue that has to be judged by its position with respect to the null-spectrum. And since it turned out to be a non-trivial extension of one correlated cluster Fig. 8.23 shows the following scenario with two independent and equally correlated clusters:

- $T = 6$ number of realisations per random variable
- $N = 300$ number of random variables
- $N_c = 2$ number of independently correlated groups of variables
- $N_1 = 20,\ N_2 = 20$ numbers of correlated variables in each group
- $\xi_n(t)$ Gaussian noise data set n where $t = 1, \ldots, T$.
- Type of correlation within group i:

  $$\xi_n(t) = \xi_n(1 - c_i) + \Xi_i(t)c_i \tag{8.20}$$

  Ξ_i is a prefixed "parent" noise vector specific for group i and $c_i \in [0, 1]$ is a correlation coefficient. Ξ_1 and Ξ_2 are identical for all realisations!
- $c_1 = c_2 = 0.9$ (equal and very high correlation)

In spite of equal number of variables and magnitude of correlation in both groups we get *two* maxima in the distribution of eigenvalues that are pushed out of the null-spectrum. The explanation for this is elusive as well.

Note again that the histograms are produced by averaging over many realisations, thus we cannot compare two curves in the application but can do a confidence test using the previously calculated expected hypothesis density. In reality we only have six eigenvalues that have to be judged by their likelihood of appearance with respect to the informationless bulk of the spectrum. Furthermore, the data for the above toy examples is perfect in the sense that there are

1. no outliers,
2. no (systematic) measurement problems,
3. no technical variance (usually also systematic).

Such errors have to be modeled and obtained from experiments. The null-hypothesis will require the a priori calculation of the null-hypothesis density by averaging over many realisations. The data must contain the above error model in a parametrised fashion.

This is not a sort of nit-picking since the impact of some error or outlier within the series of only six experiments for a gene on the resulting six non-zero eigenvalues is very large. There are articles devoted entirely to error modelling and treatment in microarray experiments and data analysis [55].

The creation of the null-hypothesis spectrum would have to include these problems under the assumption that they are stationary and reproducible *during the experiment* in a parametrisable fashion.

Note that in principle this is the same procedure as in the null-hypothesis toy market from Chap. 6.

8.9 Summary and conclusion

We have shown so far to what extent the information content in the eigenvalues and eigenvectors of a correlation matrix is redundant as long as we have more measurements than variables. Spectral clustering ignores this redundancy which is justified as long as the data is perfect. Moreover, clustering methods that disregard the data and only consider the correlation matrix make errors due to ambiguity. In case of having far less measurements than variables we have demonstrated how the correlation matrix spectrum differs, using either the sample mean and covariance or the expectation values. In the scenario of microarrays the application of correlation matrix analysis for the extraction of functional groups turns out to be questionable. It can serve, however, in the analysis and modelling of errors.

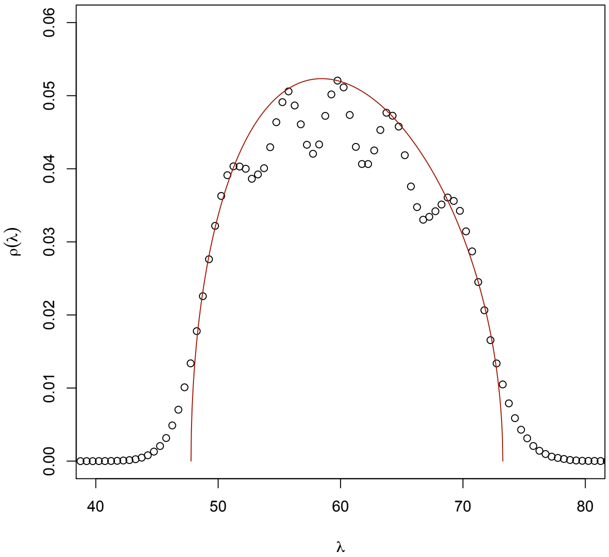


Figure 8.21: Non-zero part of the spectrum averaged from 250000 sample correlation coefficient matrices with uncorrelated noise; $N = 300,\ T = 6$. The normalisation in this pictures is to the total number of non-zero eigenvalues. Also shown is the scaled and shifted analytic prediction for the Wishart ensemble to allow comparison of the shapes.

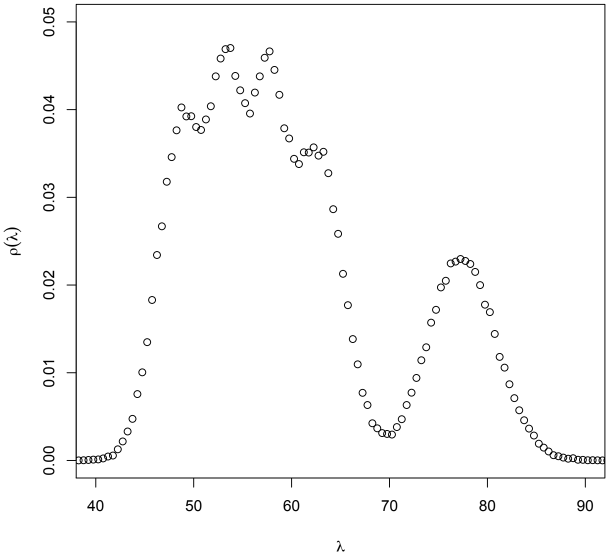


Figure 8.22: Non-zero part of the spectrum averaged from 50000 sample correlation coefficient matrices; $N = 300,\ T = 6$. The first 20 variables are artificially correlated with each other with a coefficient $c = 0.9$ according to Eq. (8.19). The normalisation in this pictures is to the total number of non-zero eigenvalues which is the only significant part of the spectrum.

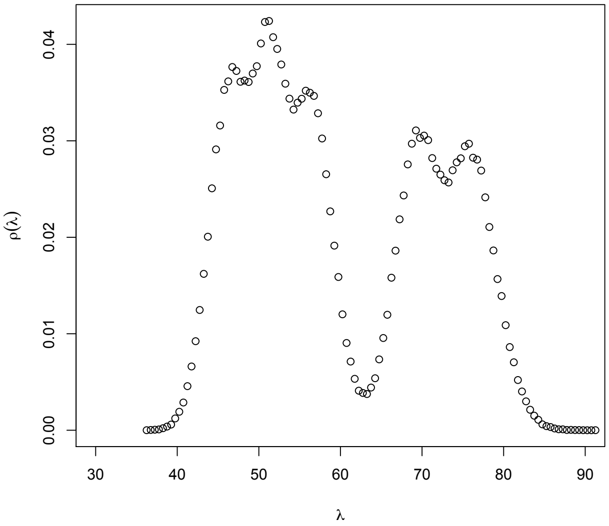


Figure 8.23: Non-zero part of the spectrum averaged from 80000 sample correlation coefficient matrices; $N = 300,\ T = 6$. The first 20 and the second 20 variables are artificially correlated clusters independently of each other with a coefficient $c = 0.9$ according to Eq. (8.19). The normalisation in this picture is to the total number of non-zero eigenvalues which is the only significant part of the spectrum.

Chapter 9

Networks of synchronisation in electroencephalographic activity

> *It's a poor sort of memory that only works backwards.*
>
> *L.C.*

This section presents the very recent and preliminary results of synchronisation analysis of electroencephalographic activity (EEG) measurements that where recorded during the *Siesta* project some years ago. The data analysed here comprises 6 channels giving 36 signals after band pass filtering into the typical frequency bands of brain waves. Data sets for two nights of three people were available. These results were produced in joint work with the group of Prof. Jan Kantelhardt, Department of Physics at the Martin-Luther Universität in Halle and with Dr. Ronny Bartsch at the Department of Physics, Bar-Ilan University in Tel Aviv within the EU-project *DAPHNet* (http://www.daphnet.eu) between June and November 2008. In particular the data pre-processing was carried out during the foregoing years by other workers at different locations after the original measurement campaign. This material is also incorporated into the final *DAPHNet* scientific report D3.2-2009 and represents, among other material, the contribution by the Institute for Scientific Interchange (ISI), Torino (for which D.F. did this work). At the time of writing the ISI's part to bridge the gap between pure fact finding in data and "modelling of everything" is still under construction and will probably not be completed before the end of the project or this millennium.

In brief: We find that the measure of synchronisation as done in this work is surprisingly sensitive to the sleep stages or changes of sleep stage and can possibly detect additional characteristics in comparison to the typical inspection by eye performed for sleep stage classification.

9.1 Definition and measure of synchronisation

Correlation and synchronisation are concepts that provide a measure to describe two types of mutual behaviour of two signals or time series that are sufficiently continuous in time. While correlation is concerned with the amplitudes of the signals, a synchronisation measure is completely independent of the local amplitude and considers the phases only. The numerical execution may be seen as a black box while the meaning of synchronisation is explained best on an intuitive level. The mathematical definition including data pre-processing and interpretation of the result is a field where the bigger errors can be made. For completeness the definitions are introduced as well as one way among several for the calculation of the phases of the signal. Some methods for synchronisation analysis do not possess mathematical justification or correspondence to a mathematical concept but can be equally meaningful in the context of the questions that are asked. Likewise, it might be meaningless to try interpret the "instantaneous" frequency even though the variant via the Hilbert transform explained below also gives numerical values untouched by the fact that a signal may contain several distinct frequencies and might not even be periodic at all. In both cases the notion of "phase" is unclear.

For simplicity we start with the assumption that the signals under consideration have no negative frequencies. This is an issue which is not of practical concern as explained below. A (complex) signal can then be represented as

$$z(t) = \frac{1}{2\pi}\int_0^\infty Z(\omega)e^{i\omega t}\mathrm{d}\omega, \tag{9.1}$$

where $Z(\omega)$ is the complex Fourier transform. Any sinusoidal function $K\cos(\omega t + \phi)$ can be converted into a complex periodic sinusoidal function $K\exp(i(\omega t + \phi))$ with positive frequency by adding to the original signal the same but phase shifted and complex conjugated signal:

$$K\exp(i(\omega t + \phi)) = K\cos(\omega t + \phi) + iK\sin(\omega t + \phi). \tag{9.2}$$

A multiplication by the imaginary unit can be interpreted as a rotation of the complex pointer by $\pi/2$. This is an example the mapping of a sinusoidal function onto a circle on the complex plane has a simple analytic representation and we can easily identify a phase in the resulting complex signal. Real data may contain more than one frequency and the above procedure is not obvious to carry out. One possible method is to use a Hilbert transform filter denoted by $\mathcal{H}[x(t)] = H(t)$ that shifts each component by the above $-\pi/2$ for a positive frequency and $+\pi/2$ for a negative frequency. The new complex signal can then be constructed by using the Hilbert transform as the imaginary part:

$$z(t) = x(t) + iH(t). \tag{9.3}$$

The cancellation of negative frequencies can be demonstrated using a classical example: Be $x(t) = 2\cos(\omega t) = \exp(i\omega t)+\exp(-i\omega t)$ such that the Hilbert transform is $\exp(i\omega t - i\pi/2) + \exp(-i\omega t + i\pi/2) = 2\sin(\omega t)$.

In practice similar considerations are necessary as with the Fourier transformation. Usually we deal with finite non-periodic and discrete data using a method made for infinite and periodic data. Note that the Hilbert transformation is a way to get rid of problems occurring in situations where the meaning of a phase is not obvious due to a broad-banded signal. The transformation *always* gives a number regardless of the signals. In fact, one possibility to calculate the Hilbert transform is via the Fourier transformation which is connected to the Hilbert transformation in the following way:

$$\mathcal{F}[H(x(t))](\omega) = -i \ \mathrm{sgn}(\omega) \ \mathcal{F}[x(t)](\omega). \tag{9.4}$$

It is obvious from this mathematical connection that the true Hilbert transformation is not causal, in particular if t is time. Forget for the moment about the numerical details of the Fourier transformation, data pre-processing and smoothed windowing, etc. to obtain the Hilbert transform. The instantaneous amplitude and phase of the signal can be calculated by

$$A(t) = \sqrt{x(t)^2 + H(t)^2}, \tag{9.5}$$

$$\phi(t) = \tan^{-1}(H(t)/x(t)). \tag{9.6}$$

Some people discuss the meaning of the local frequency $f = \dot{\phi}/2\pi$ which is sometimes a pseudo-discussion since the meaning depends on the application and questions asked as it is always the case. Only pure mathematics does not require any discussion. With two data-sets containing the phase information of two signals $\phi_1(t)$ and $\phi_2(t)$, $t = 1, \ldots, T$ a measure of synchronisation can be defined as the average over all phase differences:

$$s_{ij} = \frac{1}{T}\sum_{t=1}^{T} \exp(i\pi(\phi_1(t) - \phi_2(t))). \tag{9.7}$$

This measure accounts for two signals whose frequencies are potentially integer multiples of each other and should be considered as synchronised. Note that by dealing with narrow banded signals radio technicians implemented the Hilbert transformation long ago using circuits that extract the approximate local frequency and then only shift the signal by 90°. Thus the detection of synchronisation in practice is not only different from paper-work but cannot implement in any case the above mathematical definition of phase precisely since the Hilbert transformation is, most of all, non-causal in the case of temporal variables. A quick reference for this topic is Ref. [162].

Technical remark The pre-processing and filtering into six bands is a totally different issue and described in any textbook on neurological measurements. The amounted CPU-time for the calculation of the phases of the amplitudes of the 36 EGG signals for one night is several hours using an optimised C-code by Fabian Gans from the Martin-Luther Universit"at Halle. This includes the two-fold execution of the described procedure for the calculation of phase on each of the six frequency bands to extract the phase of the *amplitudes* of the original EEG signals. The results presented here provide the phase information at a rate of 40 Hz.

9.2 Non-mathematical comments

There is at least one strong reason for the choice of the amplitudes of EEG activity as the signal of interest with respect to their synchronisation properties:

> The amplitudes of envelopes are meaningful to compare between measurements of very different frequency contents, amplitude dynamics and even different meaning.

This can be put differently: As opposed to correlation, the synchronisation measure is meaningful between signals that have “oscillating” character, i.e. a relatively narrow band in the frequency spectrum as well as with different locations of the bands. The term “different meaning” refers to either physiologically “distant” signals as EEG voltage and blood pressure or to the units of measure as pressure and voltage and with strong individuality. We believe that a network expresses “connections” between entities subject to a meaningful definition of such connections that must be based on some a priori model. As long as there is no satisfactory stylise fact model of the body or “just” the brain yet, and most likely will not be in the intermediate future, we must start with the simpler approach of using one kind of data and one kind of measure of this connection. We are convinced that this does not lead to under-challenge, though.

The results presented here provide two kinds of information: a) The dynamics of eigenvalues and eigenvectors of the matrix of synchronisation coefficients. b) The dynamics (empirical change) of the network structure that can be extracted from this matrix. Matrix and network are different interpretations or graphical ways to display the information content in the set of synchronisation coefficients. The calculation of eigenvalues and eigenvectors is one possibility for data reduction. Whether it is complementary and not just redundant can be judged only on the basis of the data and a posteriori whether distinct “patterns” are detectable. This is indeed the case here. The interpretation of such patterns is a second step based on model assumptions.

All networks and matrices are calculated for suitable time slices and one may try to observe correspondences with the actual sleep stage. To handle the amount of data we produced a crude graphical interface that can cycle through either time or eigenvalues displaying the respective graphs in a convenient way. Fig. 9.2 is a screen shot of this interface running in time cycle mode and displaying time markers. In this mode network and matrix are displayed on the fly.

9.3 Procedure and results

We show results from one night out of three available sleep data sets obtained from different persons. The data shown here represents healthy sleep. The classification and numbering of sleep stages is given in Table 9.1. Fig. 9.1 shows an example of a matrix of synchronisation coefficients. This picture colour-codes the magnitude of the coefficient. The matrix is symmetric, i.e. there shall be no distinction between positive and negative synchronisation, while this depends on the direction of view in time. Furthermore, the matrices and respective networks are calculated for time slices of 100 seconds. Since we have to assume that the synchronisation pattern is sufficiently stable during this period, the distinction between positive and negative synchronisation is not necessarily important. This information may be displayed separately. However, so far we have found much less negative than positive synchronisation and, have no opinion on the significance of the distinction.

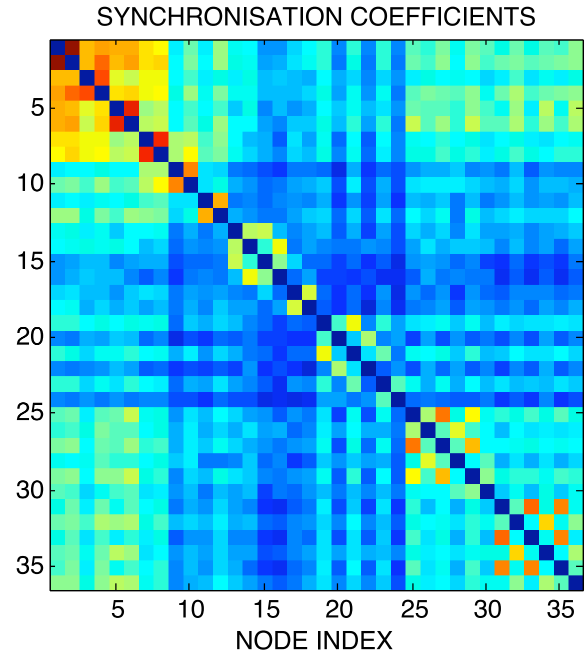


Figure 9.1: Typical matrix of synchronisation coefficients of EEG signals measured during wake.

As with any symmetric matrix we obtain a set of real eigenvalues and eigenvectors for each time slice. The top graph Fig. 9.2 shows the time dependent eigenvalues for the entire sleep. The overall procedure and characteristics of the data analysis can be described quickly:

The time slices in this example are windows of 100 seconds width. The window is moved along time in steps of 40 seconds, i.e. the windows overlap. Each data point on the time axis in the eigenvalue plot or a colour column in the eigen-

stage number	description
0	wake
1	non-REM 1
2	non-REM 2
3	non-REM 3
4	non-REM 4
5	REM
6	artifact

Table 9.1: Typical classification of sleep stages. The non-REM stages are not clearly distinct and continuous in transition as opposed to REM and wake which in turn are similar with respect to typical classification criteria.

frequency band [Hz]	band name
8–13	α
14–30	β
4–7	θ
0.5–3	γ
0–3	δ

Table 9.2: Classification of brain waves into frequency bands (in this order). There are different definitions in neurology and the number of bands differs between 4 to 7 with flexible ranges or overlap.

vector plot represents (the middle of) such a time slice. Each time slice gives a synchronisation matrix, a network, a set of 36 eigenvalues and 36 eigenvectors.

The choice of the above numbers seems appropriate for the time scales of consideration. Remember, we deal with sleep (stage) dynamics that take place on the scale of few seconds to a minute. Higher time resolution is prohibited by statistical uncertainty in the measured quantity, but may be increased with more sophisticated methodology. Fig 9.2, center plot, shows the eigenvector of the largest eigenvalue as a column in colour-code. The colour range is normalised to the min and max of the entire picture. The bottom plot indicates the sleep stages. The time axis for all three plots correspond. We chose the eigenvector of the largest eigenvalue (numbered 36) for this demonstration. The other 35 are also partially significant but are too many to display at once.

Figures 9.3 to 9.10 show the networks at the labelled (green lines) times in Fig 9.2. The names of the nodes correspond to the usual nomenclature in sleepology (posh: somnology). Of course the arrangement is arbitrary. This choice puts left-right next to each other because we observe almost constant, thus possibly meaningless, synchronisation. Deviations from this may be subject to higher order interpretations.

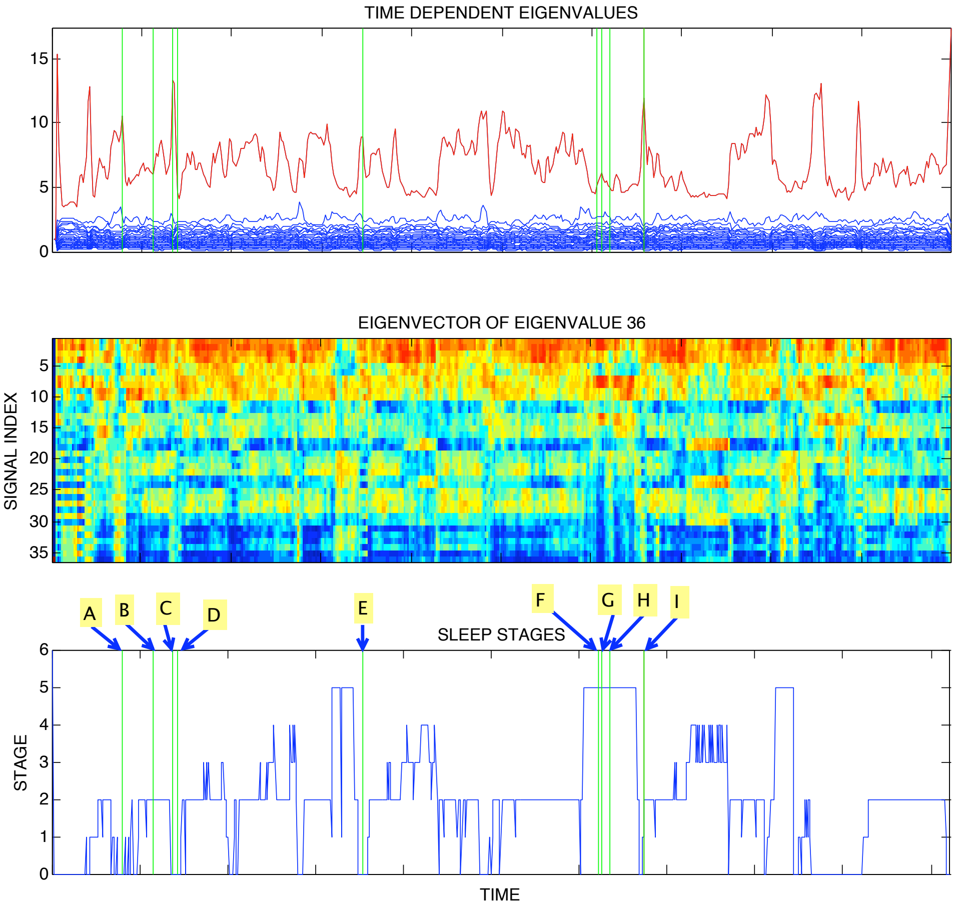


Figure 9.2: Analysis of one night sleep of a healthy subject. BOTTOM: The sleep stages as found by inspection by medical doctors. TOP: Eigenvalues sorted by size. The red curve marks the larges eigenvalue whose eigenvector (MIDDLE plot) is shown colour coded as columns with the width of the time slice. The green lines and labels indicate selected points whose network interpretations are shown in the remaining pictures below.

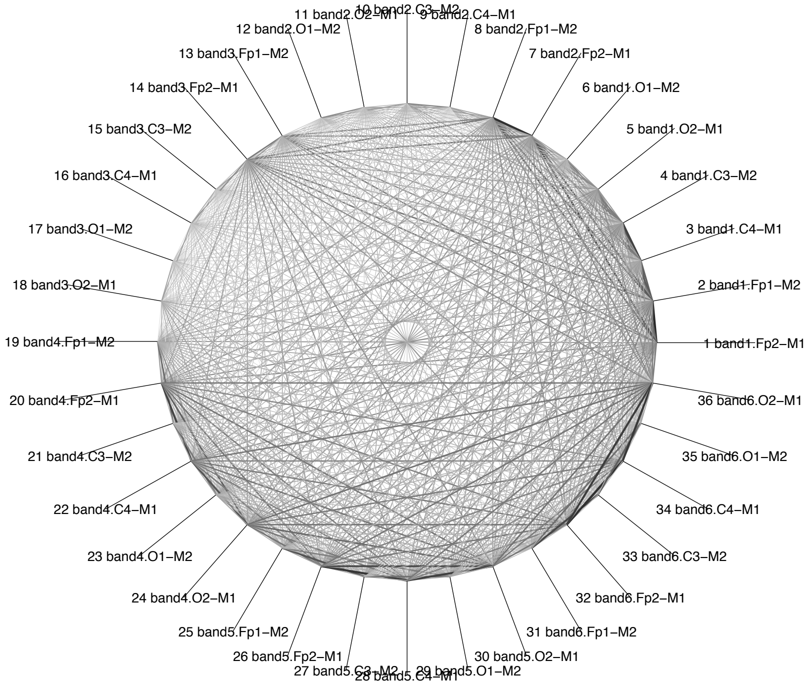


Figure 9.3: Network representation at A

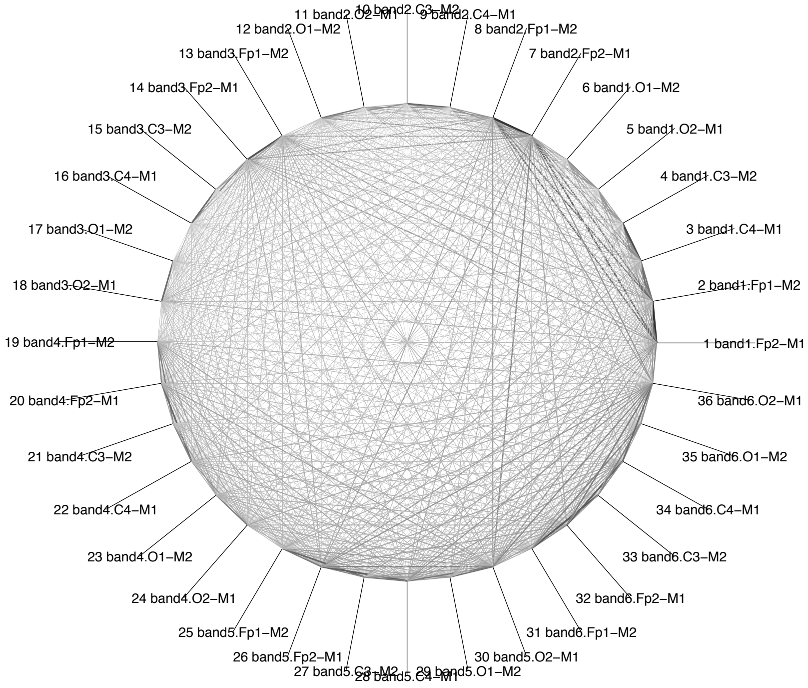


Figure 9.4: Network representation at B

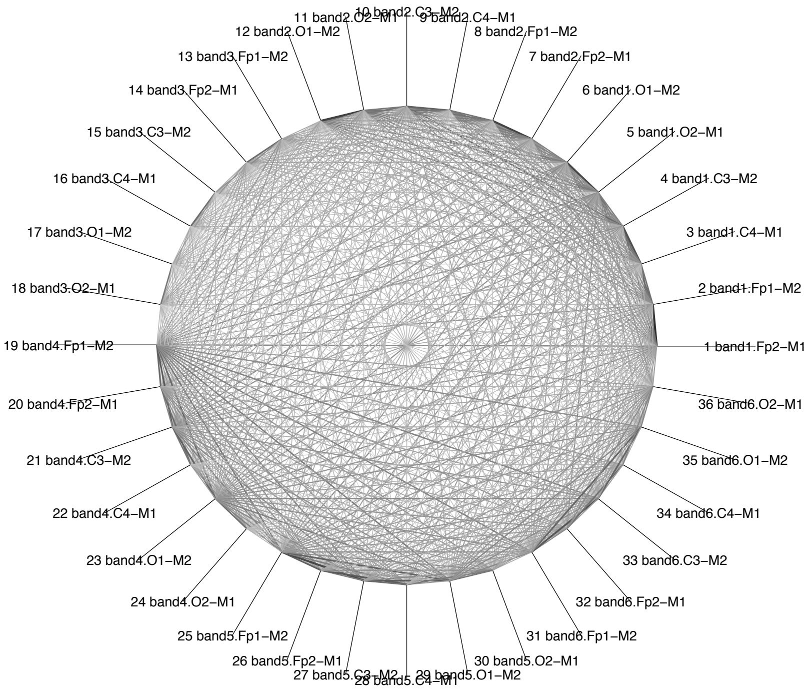


Figure 9.5: Network representation at C. This pattern appeared for a brief moment just at the transition to wake. This sudden emergence of connectivity occurs quite regularly at changes between states, mostly towards wake.

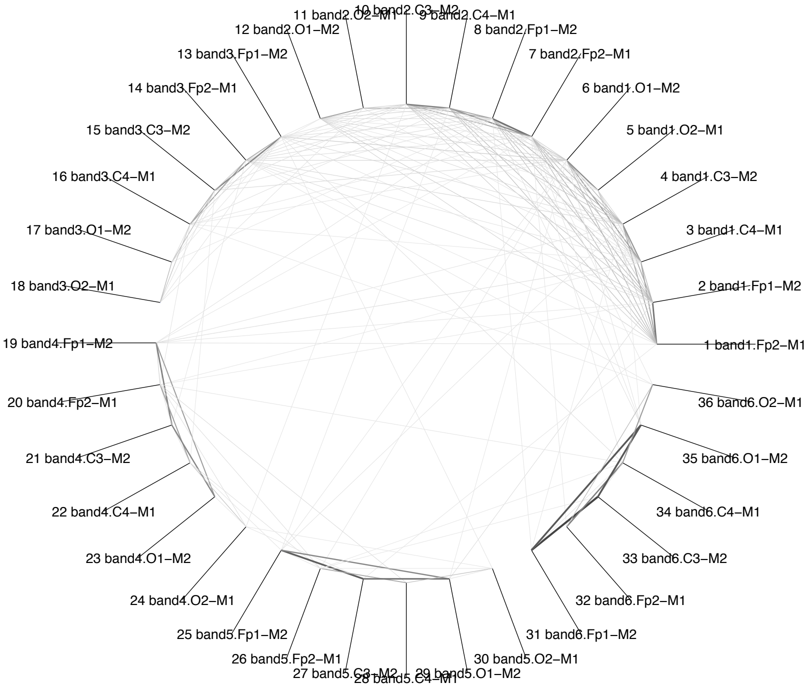


Figure 9.6: Network representation at D. The three triangles are very typical for wake.

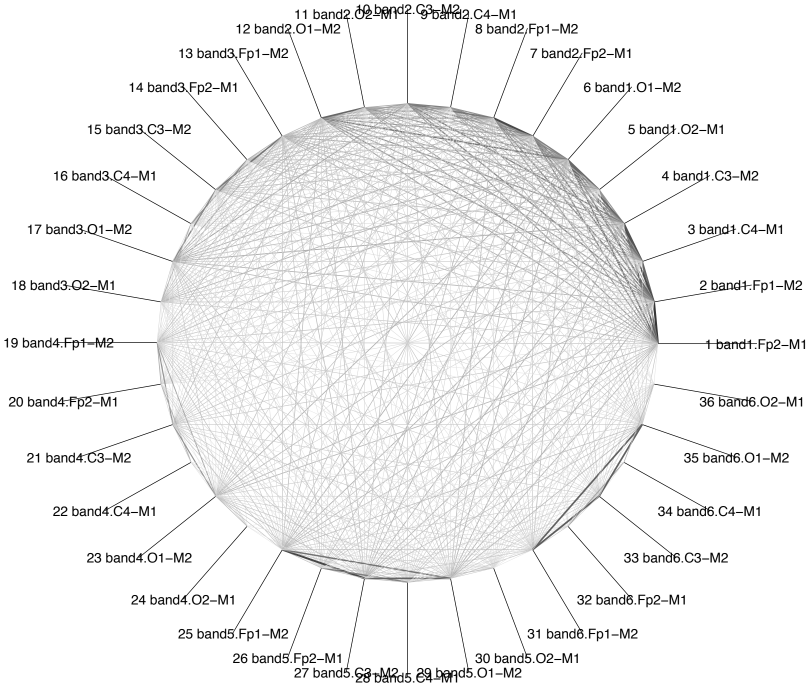


Figure 9.7: Network representation at E. We often observe a short burst of connectivity just after the drop from REM.

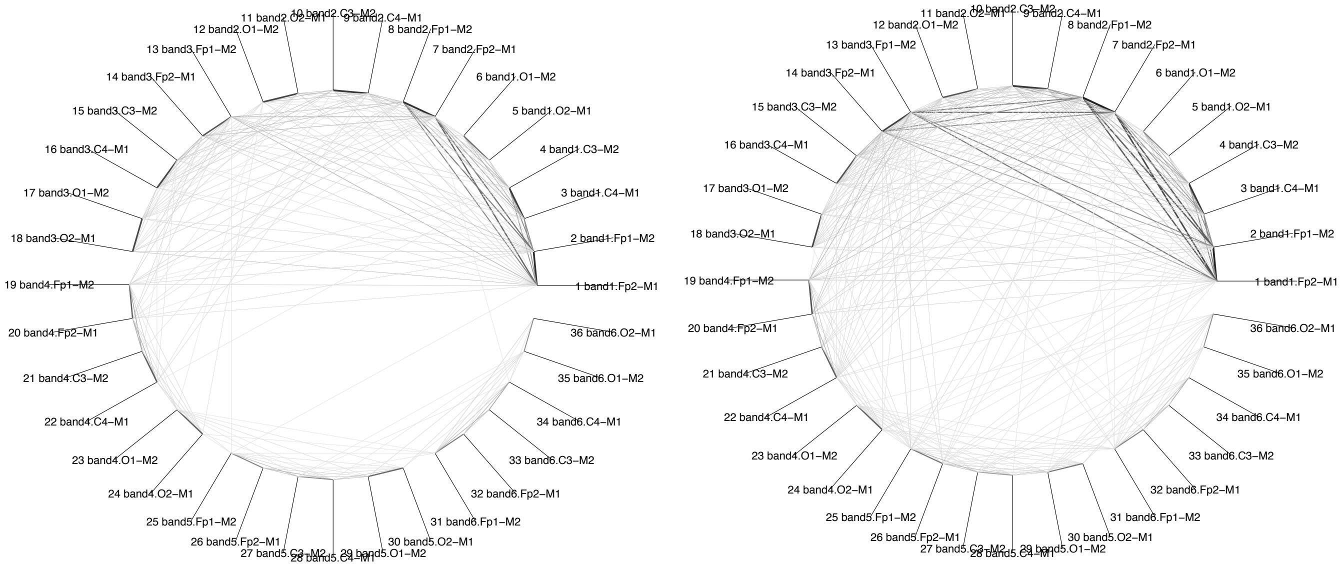


Figure 9.8: Network representations at F and G. Examples of a more connected REM.

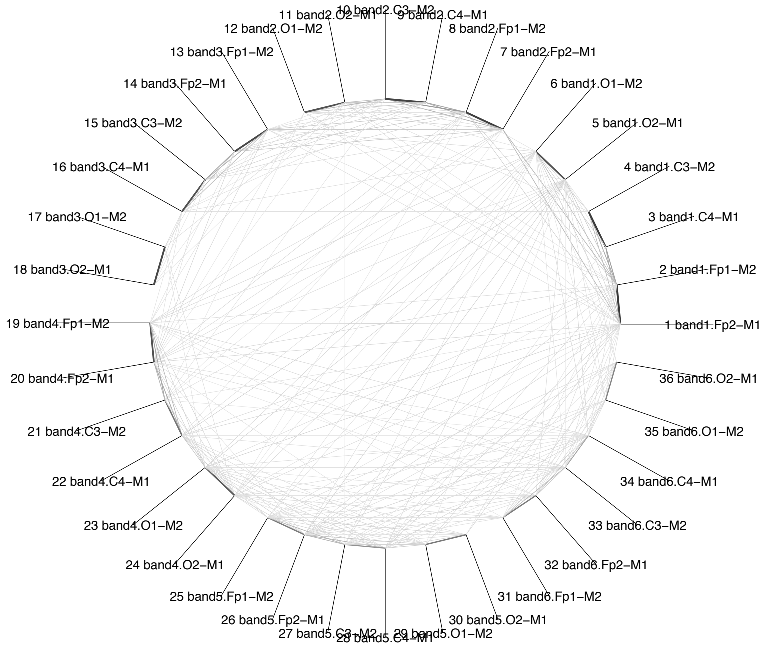


Figure 9.9: Network representation at H. Example for a less connected REM which is more persistent as opposed to F and G.

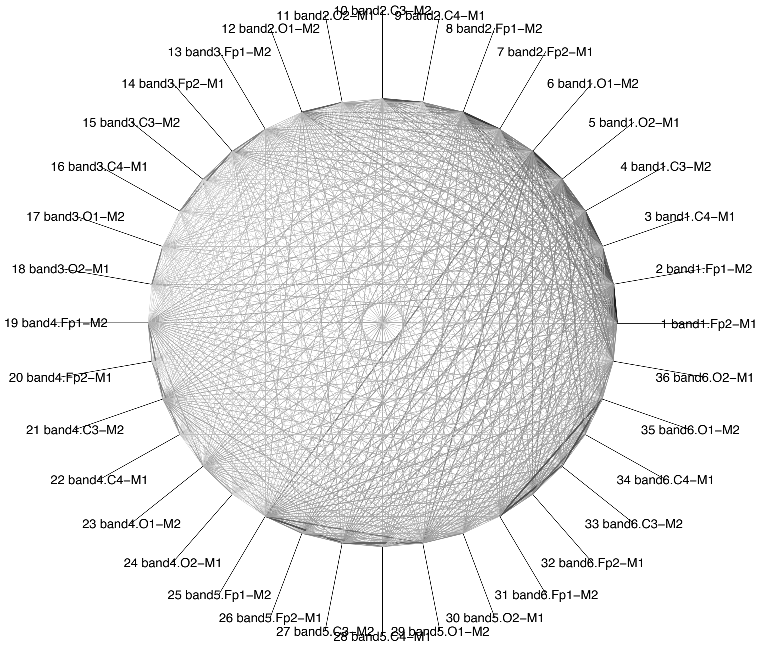


Figure 9.10: Network representation at I. This situation occurred just after the drop from REM to wake. The features are the seemingly wake-typical triangular connections around 16:30 and 18:30, speaking in terms of the clock dial. We also observe the sudden overall connectivity which often occurs just after the drop from one higher sleep stage to wake. This connectivity disappears quickly and concurs with the peak in the largest eigenvalue.

9.4 Some remarks on eigenvalues of coefficient matrices

This topic is a field of study of its own as we have seen in previous capters. To enable the reader to glance into possible interpretations we make a few remarks, nevertheless you may revisit previous chapters in this document.

The calculation of eigenvalues and eigenvectors is a data reduction technique that has a similar motivation as the principle component analysis with singular value decomposition. The information that is condensed or "reduced out" of the data is different and the choice of method depends on the expectation and educated guess of the scientist. In the case of eigenvalues form synchronisation coefficients (being a more general measure but similar to correlation under certain conditions) we extract or measure so called modes. A mode could be, for example, a common behaviour of a sub-group of time series (if this is where the coefficients come from) that is detected by the instrument of choice. This instrument is, for example, the correlation estimator or the synchronisation coefficient as defined in Eq. (5.2). The influence of such modes onto the (realised) eigenvalue spectrum is non-trivial. Most studies in literature take unexpectedly large eigenvalues as an indicator for a strong mode. This can be some parallel co-movement of a few time series or a loose correlated behaviour of a large sub set. So far, it seems to have gone unnoticed that such modes produce large and small eigenvalues leaking out of the spectrum for the null-situation. These small eigenvalues carry similar, partly redundant, information. At least this is the case for correlation measurement for which a mathematically clean theory exists. For example, it is also unclear why the matrix of synchronisation coefficients must always be positive definite.

Furthermore, one must always compare the results with the null-hypothesis because the random or no-information-case produces eigenvalues and vectors that do posses a structure. But what is the null-hypothesis situation? Indeed, the issue of surrogate data is not trivial. As rehearsed also elsewhere in several places in this document, the mere test on some random artificial wiggles does not prove insensitivity for false positive results, i.e. apparent correspondence to sleep stages. Neither is the combination of signals from 36 sleepers a suitable null-hypothesis because it is obvious to the trained eye already that the channels cannot fit together. This issue has to be picked up once the research goal is an improved sleep stage finder.

The magnitude of the eigenvector elements, usually, indicate the time series that contribute to the mode that is "detected" by the respective eigenvalue. Since eigenvalues can only be referenced by their size, the mode cannot always be traced across time if the eigenvalues cross. Then more sophisticated methods are needed. This situation is very similar to electronic band structure calculation, where band crossings are not distinguishable from small band gaps [60]. It cannot be overstressed: This also requires a model assumption about the system. Here we just show the eigenvector for the largest eigenvalue. Thus one must not expect to get a sleep stage finder by classifying the patterns of colours. Also note that measurement errors affecting all cables for example (electric machine running on 50Hz somewhere) might induce a mode in the spectrum that is not visible in the data directly, not even via a power spectrum analysis.

9.5 Patterns in data and remarks on the significance

The common knowledge that REM sleep goes together with an active visual cortex is not expected to be observed in our results and with this type of data analysis: 1) There are only two electrodes in the relevant region. 2) Activity is not necessarily expressed in stronger synchronisation, in particular not necessarily with other regions of the brain.

REM may directly affect the read-out of the electrodes. To what extent this poses a problem or can be seen in the plots is not yet analysed and non-trivial. However, eye movement or any other activity of the sleeper is most likely not the cause of most of the detected patterns because seemingly sleep stage typical patterns persist for a long time or occur quite regularly at certain transitions between stages.

Some of the selected data refers to peaks of the largest eigenvalue. During REM the analysis of the available other person's sleep shows stronger than in this example a plateau in the largest eigenvalue.

The reader may observe different patterns in REM. The conclusion is that REM or any other state is not the only relevant state variable of the brain during sleep. If unrelated measurement errors are excluded, like technical problems or movement of the sleeper, the analysis algorithm is not corruptible. We cannot comment or judge on the accuracy of the sleep stage classification. Overall it seems that more is going on beyond fluctuation around some stationary state.

We are inclined to observe in the eigenvector of the largest eigenvalue that during REM only few nodes contribute to the dynamics of this eigenvalue. Overall the intervals of REM are visible by inspection. However, a thorough quantification is needed before significance is proven or even some automatic classification is thinkable. So far these pictures may assist the standard procedure by medical doctors.

At this point one more result may be appropriate to discuss. The whole point about eigenvector and eigenvalue analysis is principle component selection. In the data presented here it appears that only one eigenvalue significantly departs from the (seemingly) informationless bulk. We may now average the main contributing channels as given by the eigenvector of eigenvalue 36 in a sleep stage specific colour-coded manner. The result is shown in Fig. 9.11. Note that relatively few lines per sleep stage appear. This means that they are repetitive. All stages are therefore rather distinct, at least in the inspection by eye. What this plot does not reveal is the frequency of modes which requires an additional visual variable.

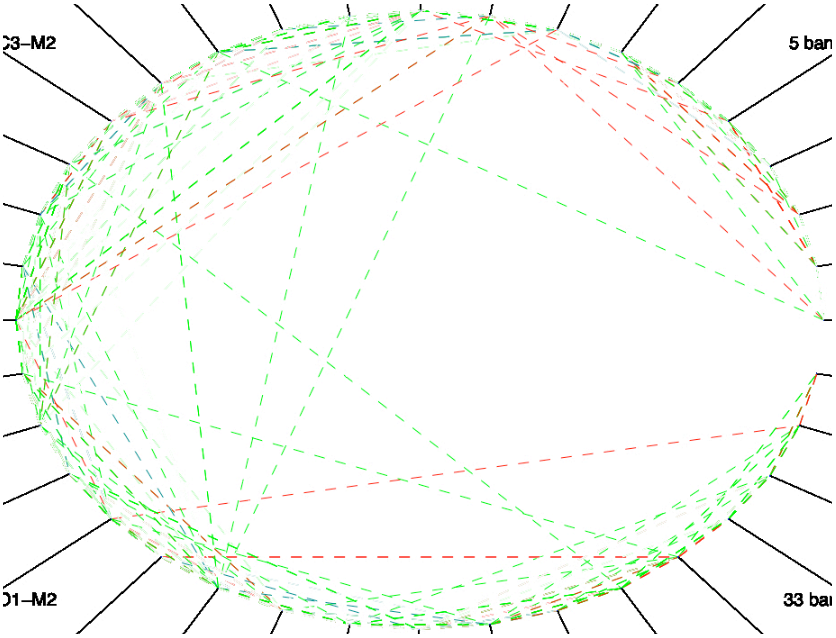


Figure 9.11: The lines connect all nodes contributing to the mode coded in eigenvector 36 which represents the most dominant mode by eigenvalue 36. The node labels are clipped off in order to enhance the details. The colours are: red–REM, blue–deep-sleep, green–wake. Blue lines are only seen on the right hemisphere.

9.6 Summary

We showed preliminary results from the time series analysis of EEG data using the synchronisation measure to detect patterns. These patterns, for example the network, are extracted from the matrix of synchronisation coefficients of the amplitudes of the signals and the eigenvalues and -vectors. In doing so, we can also compare between different bands and detect synchronised behaviour in the activity.

Already in this preliminary stage of the analysis the extracted patterns react very sensitively to the sleep stage and the patterns seem sufficiently reproducible for sleep stages to attempt a sleep stage finder on this basis in future work with additional data from other measurements. We conclude that the synchronicity of activity in EGG data is significant. To elaborate more on this one must take many nights of the same individual and analyse the individuality of the patterns. This requires additional algorithms for pattern recognition as well as more scientific insight.

We also need high resolution data to be able to filter non-physiological correlation and to calculate network properties that are independent of the arbitrary reduced display that mostly pleases the human eye.

Chapter 10

Some quintessence

Instead of one single topic this document contains a chain of topics with overlap. The contributions to science are explicitly listed at the beginning in chapter 1 along with a summary at the end of the respective chapter or section. Nevertheless, there are some global conclusions and lessons learned for the author of this document mostly in dealing with representatives of the community, literally.

Lesson one is reflected in the quotation by Albert Einstein in Chap. 8. It has to be understood in the context of its time and needs to be enhanced. He talks about certainty which is possibly misunderstandable as problems with the accuracy of certain laws, but this is not what it is about. Since the advent of the so called complex systems science (some) people have become more sensitised to questions of "meaning". The first sign of it being a bigger issue was probably Max Born's Nobel Prize for the interpretation of the wave function which piqued experimentalists like Aharonov and Bohm plus Ehrenberg and Siday to test whether the wave function and the probability cloud are really a physical entity. So it seemed at least, and people saw no problem because nothing seemed to be missing in the model. Schrödinger's equation still describes all (stylised) facts of the world, if we were able solve it for slightly larger system than a few atoms. With the attempt to model systems that under no circumstances could be described bottom-up one had to decide on which aspect to model and what the equations (or the computer code) actually mean. The reproduction of a stylised fact could not be traced back to the microscopic structure of the system anymore as opposed to the diffusion equation for example that, in the limit of scales, is rigorously connected to the diffusing random walk particle which in turn is a limiting case of the most accurate atomistic model one can set up. It is at this point where things fall apart. People truly think that rough model equations for a certain stylised fact of a mass-psychological system have to be consistent with physical laws. And the most infuriating proposed example is the conservation of some quantity in a social model that looks like the Ising Hamiltonian. This issue and the definition or meaning of a "model" is truly not understood and typical debates can be held and observed with many people.

> *[..] Most important is the ability to understand the underlying dynamics of a complex system. These insights are needed to assess whether the assumptions of a model are correct and complete. The modeler must be able to recognize whether a model reflects reality, and to identify and deal with divergences between theory and data.*

From William Silvert, Modelling as a discipline (2001), Ref. [161]

This guideline can hardly be stated clearer but is often disregarded in well-financed international projects.

Lesson two is a bit more methodological and in extension of the above quotation. It refers to the often encountered fixation on the number of sigmas in confidence testing of model predictions where three sigma, or a sufficiently high number, is considered enough to discriminate between chance and the prediction of a model. First of all, statistical testing must be possible and sensible. Along with this misunderstanding goes the neglect of null-hypothesis testing which often reveals that the most stupid random "model" can be as good as three sigma. And it often turns out that a reasonable null-model is the biggest problem. One can often observe refusal to realise this because of the following attitude: The problem is the problem and not how to simulate the random system. At this point we can refer back to the previous lesson. A not so fictitious example: We claim to have a model of the brain. Proof: 95% of the times the model correctly predicts the direction of walking $\Delta t =$ five seconds later of any randomly chosen person in the world that can more or less be considered to be walking. One does not have to play devil's advocate to find that most people do not tend to do a random walk. So if someone goes somewhere it is usually rather straightforward. This might be enforced by the street layout in Manhatten or simply by the fact that the shortest connection between two points is a straight line. Now which is the appropriate city in the world to test our model? The most appropriate answer to this question is not enough since we miss other parameters as how long do we test, and how large must be Δt. In Manhatten Δt must be longer than the time to cross a typical block. Then we might be able to observe a random walk on a grid. Thus no textbook gives us *the* test on how to decide how "good" a model is. Any testing is based on the model and our understanding of its connection to reality. This guideline is often disregarded in well-financed international projects.

Number three, what about the non-local and non-causal integrals of the fractional diffusion equation? Non-locality is usually disliked because of the limited speed of light. This is a physical issue. We have learned, however, that in complex systems we are satisfied with the reproduction of stylised facts regardless of the justification of the microscopic parts of the model. In this case we must be content with the Ptolemaeic planetary model as shown in Fig. 10.1. It reproduces the loops of the planets in the sky but are we unnecessarily pedantic in having a problem with this model? Assuming we had not Kepler's analytic solution of the orbits, then we would need numeric integration to calculate the orbits. As with all such methods they produce errors of different types. Which is the right integrator? A symplectic one that turns around the axis of the orbit ellipse but conserves energy, or a high-order Runge-Kutta scheme that traces the true orbit longer but lets the planet fall into the sun after a while. Unless somebody checks if there are hinges with disks on the edges of which the planets turn there is no proper way to decide if Ptolemaeus or Gallileo is right. Meanwhile, both models can be used to predict the lights in the sky and the non-local integral in the fractional diffusion equation predicts the probability density for a continuous-time random walker.

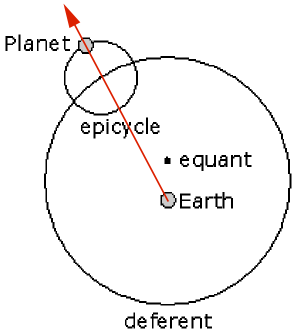


Figure 10.1: The Ptolemaeic solar system with the line of sight from us to a planet.

Number four is a case for one of the previous general lessons on modelling and situated in the context of financial time series analysis. As usual, data analysis goes hand in hand with model assumptions on the source of the data and also the questions asked. Therefore we often encounter statements about the underlying system in treatises that are entirely focused on measurement and analysis. A statement that is simply wrong is that high-frequency financial time series are irregular observations of a continuous process. To contradict such statements is not yet part of the lesson. But it is an example of where methodologists tweak reality to argue for (their) nice theories and methods. Another example in the financial context is the notion of noise. It is is very tempting to introduce noise and the continuous process postulated above because we can apply smoothing models and then argue for methods and models that are proven to work best on smooth continuous processes with noise that is possibly Gaussian and independent. Likewise, at a microscopic level the partial differential equation of diffusion does not work anymore for several applications and questions asked. This is physically obvious. In the field of so-called complex systems the physical normative often does not exist or is simple to talk away. And again, data analysis makes assumptions on the behaviour of the data and on the important feature that has to be extracted. The latter decides on which analysis method is appropriate and is connected to model assumptions. This guideline is often disregarded in well-financed international projects.

And finally, for the geeks: The future belongs to high level programming with all thinkable methods and algorithms already implemented. Even if the language is primitive like Matlab without strict typing the size of the code is so small that it makes up for the strict typology in a language like C++ where it takes a page of code to do what 4 lines of Matlab can do. Furthermore, the inclusion of libraries costs enormous time which no speed-up of the resulting code can recover, not to speak of the errors which have more room to thrive and spread than in the 4 lines above. Secondly, a low level language like C++ can not easily guess what the programmer wants to do. Specialised pre-fabricated functions that are part of the language package can, thus making better use of the hardware. This is particularly useful with multi-core processors.

Appendix A

Numerical calculation of the Lévy probability density

Prior to the calculation we give a short remark on two alternative ways to express the Fourier transform: the frequency and pulsation variants (using angular speed ω). Physicists tend to use ω and all related nuisance factors. Practitioners however, like engineers, use the frequency variant for good reasons and nearly all implementations of the FT perform the latter. In Formulas:

$$H(\omega) = FT\left[h(z)\right](\omega) = \int h(z)e^{i\omega z}dz \tag{A.1}$$

$$h(z) = FT^{-1}\left[H(\omega)\right](z) = \frac{1}{2\pi}\int H(\omega)e^{-i\omega z}d\omega. \tag{A.2}$$

With

$$\omega = 2\pi f \tag{A.3}$$

we get the frequency variant

$$H(f) = FT\left[h(z)\right](f) = \int h(z)e^{2\pi i f z}dz \tag{A.4}$$

$$h(z) = FT^{-1}\left[H(f)\right](z) = \int H(f)e^{-2\pi i f z}df. \tag{A.5}$$

Furthermore, implementations provide the FT of an array that the scientist has to fill. This input is assumed to comprise the right sight of a symmetric function. This means data symmetry is assumed. If anything else is the case, this must be accounted for in a post processing or in the interpretation of the resulting Fourier coefficients. The above mentioned nuisance factors come in due to the different worlds the implementation on one side and the theories on the other usually live in. Since the algorithm in the FT expects values $H(f)$ but gets $H(\omega)$ the argument values have to be divided by 2π or alternatively the x-axis in the resulting plot has to be multiplied by 2π. Below in Eq. (A.8) we choose the latter.

The Lévy α-stable distribution is a prime example of a non-analytic density function. It does not possess any moments above the second due to the divergence of the respective integrals except for $\alpha = 2$ (Gaussian case). For $\alpha \leq 2$ the mean diverges also. Values for $L(x)$, Eq. (4.10), can be computed directly as in Ref. [135, 136] or using the Fourier Transform (FT). Ref. [126] presented an FT-based procedure and report that with 2^{13} data points and more than 100 required samples of $L(x)$ the setup overhead of doing a Fast FT (FFT) is more advantageous. We consider 100 required samples or less as a rarely occurring small amount in applications of stable distributions. For our purposes 2^{13} data points turn out to be unnecessarily restrictive since the FFT is sufficiently fast to make up a negligible part of the total setup time. Therefore we employ the FFT.

Too keep it simple for didactic purposes: the characteristic function for the symmetric case

$$\phi(q) = \exp(-q^{\alpha}) \tag{A.6}$$

is evaluated at N points at intervals Δq giving tabular values $(q_i, \phi(q_i))$, $i = 1, \ldots, N$. Both parameters must be specified according to Eq. (A.8) to fit the required tails. Due to the even symmetry only positive arguments of the characteristic function and the real part of the resulting transform are required:

$$L(x) = \mathcal{R}e\left\{\frac{1}{\pi\Delta x}\,\mathrm{FFT}^{-1}\left[(q_i, \phi(q_i))\right]\right\}. \tag{A.7}$$

Despite the arbitrary choice of sign for the exponent in the Fourier integral FFT^{-1} denotes the inverse Fourier transform for consistency with the FFT algorithm and sign convention taken from [149, Chap. 13]. The spacing of the abscissa for the resulting plot is calculated by

$$\Delta x = 4\pi\,\frac{f_{\mathrm{Nyquist}}}{N} = \frac{2\pi}{N\Delta q}, \tag{A.8}$$

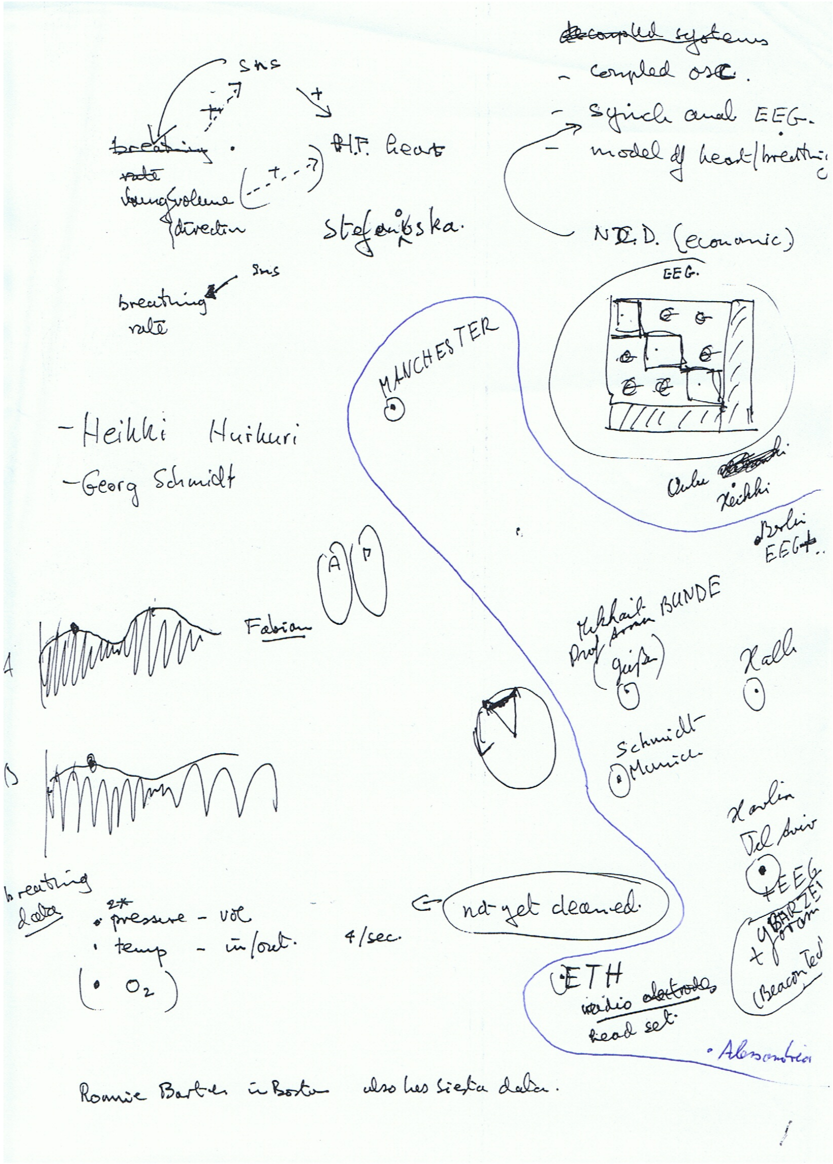
where N is the number of samples and $f_{\mathrm{Nyquist}} = 1/(2\Delta q)$. The time on our machine for the FFT part within the setup using $N = 2^{16}$ data points is 0.08 seconds. The resulting tabular values for x and $L(x)$ are used within the interpolation routine to provide a smooth curve for the tiling algorithm. The drawn random number is multiplied by the scaling factor γ.

Special attention is required if the resulting tabular values are used in the general Ziggurat implementation attempt by [91]. Since the iterative setup of the Ziggurat requires strictly decreasing $L(x)$, the tabular data $(x_i,\, L(x_i))$ must be truncated where numerical fluctuations from the FFT start to accumulate to significant flutter effects in the tail. These effects disturb the monotonicity, i.e. a sign change in $\Delta L = L(x_i) - L(x_{i-1})$, and a breakdown of the procedure. Also of concern is the interpolation as such. Typically, few surrounding values of x are chosen to interpolate for $L(x)$. i.e. if $x \in [x_i,\, x_{i+1}]$ we use data points $i = i-3, \ldots, i+4$. Due to the sudden shifting of these data points for very close subsequent values x and $x+\epsilon$, i.e. $x_i, \cdots, x_{i+1}$ becomes $x_{i+1}, \cdots, x_{i+2}$, significant jumps might occur between $L(x)$ and $L(x+\epsilon)$. Since these fluctuations are exceedingly small on the absolute scale they can be ignored in the tiling method which is inherently resistant to such errors.

Appendix B

Serious research

The following piece of A4 paper was produced at the Martin-Luther University Halle on the 15th of November 2008 together with Prof. emeritus David Breé in the attempt to think of possible scenarios for DAPHNet follow-ups and the partners involved. Please appreciate the representation of the Middle East together with Europe inspired by pre-medieval Mediterranean cartography.



Appendix C

Future research, seriously

The presented work opened up more ideas and produced more half-finished results than could be included here. The following list can be a direct continuation:

1. Direct generation of rotationally invariant matrices without summations.
2. Extended synchronisation analysis on EEG data. Systematic pattern analysis and subsequent null-hypothesis tests.
3. Detailed analysis of eigenvectors from similarity and dissimilarity matrices in the case of less measurements than variables.
4. Spectral clustering on financial and microarray data.
5. Mode reconstruction experiments on singular correlation matrices as in microarray experiments.
6. A working implementation and benchmarking of the suggested clustering method with eigenvectors is pending.

Acknowledgements

D.F. has been supported by

- Land Hessen, Germany (principal investigator Prof. G. Germano)
- an Italian research grant PRIN 2006 (principal investigator Prof. E. Scalas)
- a British Council grant within the Researcher Exchange Programme (RXP) 2008 (principal investigators Prof. G. Iori and Prof. G. Germano)
- the EU grant 018474-2 via the Institute of Scientific Interchange in Torino, Italy (principal investigators Prof. S. Solomon and Prof. David Breé)

Bibliography

[1] Mathematica, version 6.0, Wolfram Research, 2007. [46]

[2] Matlab, version 7.4, The Mathworks, 2007. [46, 67]

[3] J. Ahrens. Sampling from general distributions by suboptimal division of domains. *Grazer Math. Ber.*, 319:1–20, 1993. [49, 50, 51, 52, 53, 55]

[4] J. Ahrens. One-table method for sampling from continuous and discrete distributions. *Computing*, 54:127–146, 1995. [50, 51, 57]

[5] E. Bachmat, D. Berend, L. Sapir, S. Skiena, and N. Stolyarov. Analysis of aeroplane boarding via spacetime geometry and random matrix theory. *J. Phys. A-Math. Gen.*, 39:L453–L459, 2006. [32]

[6] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statist. Sci.*, 9:611–677, 1999. [32]

[7] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207–211, 2005. [20]

[8] E. Barkai. CTRW pathways to the fractional diffusion equation. *Chem. Phys.*, 284:13–27, 2002. [11]

[9] O. E. Barndorff-Nielsen and S. Thorbjørnsen. Lévy laws in free probability. *P. Natl. Acad. Sci. USA*, 99:16568–16575, Dec. 2002. [36, 40]

[10] P. Barthelemy, J. Bertolotti, and D. S. Wiersma. A Lévy flight for light. *Nature*, 453:495–498, 2008. [11]

[11] E. Barucci and R. Renò. On measuring volatility and the garch forecasting performance. *J. Int. Finan. Markets, Inst. Money*, 12:182–200, 2002. [80]

[12] D. ben Avraham and S. Havlin. *Diffusion and Reactions in Fractals and Disordered Systems.* Cambridge University Press, Cambridge, 2000. [10]

[13] H. Bercovici and V. Pata. Stable laws and domains of attraction in free probability theory. *Ann. Math.*, 149:1023–1060, 1999. [39, 40]

[14] H. Bercovici and D. Voiculescu. Free Convolution of Measures with Unbounded Support. *Indiana Univ. Math. J.*, 42:733–773, 1993. [37]

[15] G. Biroli, J.-P. Bouchaud, and M. Potters. Extreme value problems in random matrix theory and other disordered system. *J. Stat. Mech.*, page P07019, 2007. [34]

[16] G. Biroli, J.-P. Bouchaud, and M. Potters. On the top eigenvalue of heavy-tailed random matrices. *Europhys. Lett.*, 78:10001, 2007. [34, 82]

[17] J.-P. Bouchaud. Power-laws in economy and finance: some ideas from physics, 2000. [35]

[18] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: The subtle nature of random price changes. *Quant. Financ.*, 4:176–190, 2004. [32]

[19] J.-P. Bouchaud and A. Georges. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.*, 195:127–293, 1990. [10]

[20] J.-P. Bouchaud, L. Laloux, M. A. Miceli, and M. Potters. Large dimension forecasting models and random singular value spectra. *Eur. Phys. J. B*, 55:201–207, Jan. 2007. [34]

[21] J.-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing.* Cambridge University Press, Cambridge, 2000. [32]

[22] R. P. Brent. Note on Marsaglia's XOR shift random number generators. *J. Stat. Software*, 11:4, 8 2004. [61, 63]

[23] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006. [10]

[24] Z. Burda, A. Goerlich, A. Jarosz, and J. Jurkiewicz. Signal and noise in correlation matrix. *Physica A*, 343:295–310, 2004. [34, 40, 41, 45, 95]

[25] Z. Burda, A. Goerlich, and B. Waclaw. Spectral properties of empirical covariance matrices for data with power-law tails. *Phys. Rev. E*, 74:041129, 2006. [34, 35, 40, 45, 95]

[26] Z. Burda, R. A. Janik, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Free random Lévy matrices. *Phys. Rev. E*, 65:021106, Jan 2002. [34, 39, 40, 45, 95]

[27] Z. Burda, A. Jarosz, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Applying free random variables to random matrix analysis of financial data. 2006. [34, 39]

[28] Z. Burda and J. Jurkiewicz. Signal and noise in financial correlation matrices. *Physica A*, 344:67–72, 2004. [34, 39]

[29] Z. Burda, J. Jurkiewicz, and M. A. Nowak. Conference of applications of random matrices to economy and other complex systems, Sep. 2005. [34, 39]

[30] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Free random Lévy variables and financial probabilities. *Physica A*, 299:181–187, Oct. 2001. [34, 39, 40, 45, 95]

[31] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Lévy matrices and financial covariances. *Acta Phys. Polon. B*, 34:4747, Aug. 2003. [33, 34, 39]

[32] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Free Lévy matrices and financial correlations. *Physica A*, 343:694–700, Nov. 2004. [34, 39]

[33] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed. Random Lévy matrices revisited. 2007. [34, 39, 40, 45, 95]

[34] M. Caputo and F. Mainardi. Linear models of dissipation in anelastic solids. *Riv. Nuovo Cimento (Ser. II)*, 1:161–198, 1971. [18]

[35] A. Cartea and D. del Castillo-Negrete. Fractional diffusion models of option prices in markets with jumps. *Physica A*, 374:749–763, 2007. [10, 51]

[36] A. Cartea and D. del Castillo-Negrete. On the fluid limit of the continuous-time random walk with general Lévy jump distribution functions. *Phys. Rev. E*, 76:041105, 2007. [10, 51]

[37] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *J. Amer. Statist. Assoc.*, 71:340–344, 1976. [24, 26, 39, 51, 61, 64, 65, 67]

[38] P. Cizeau and J.-P. Bouchaud. Theory of Lévy matrices. *Phys. Rev. E*, 50:1810–1822, Sep 1994. [33]

[39] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data, 2007. arXiv:0706.1062. [33]

[40] F. Comellas and J. Diaz-Lopez. Spectral reconstruction of complex networks. *Physica A*, 387:6436–6442, 2008. [109, 114]

[41] J. Daly, M. Crane, and H. J. Crane. Random matrix theory filters in portfolio optimisation: A stability and risk assessment. *Physica A*, 387:4248–4260, Mar. 2008. [32]

[42] L. Debnath and P. Mikusinski. *Introduction to Hilbert Spaces with Applications.* Academic Press, San Diego, 3rd edition, 2005. [34]

[43] D. del Castillo-Negrete. Fractional diffusion models of nonlocal transport. *Phys. Plasmas*, 13:082308, 2006. [20]

[44] D. del Castillo-Negrete, B. A. Carreras, and V. E. Lynch. Nondiffusive transport in plasma turbulence: A fractional diffusion approach. *Phys. Rev. Lett.*, 94:065003, 2005. [10]

[45] L. Devroye. *Non-Uniform Random Variate Generation.* Springer, New York, 1986. [49, 50, 51, 52, 55, 57, 64, 65, 66, 67]

[46] L. Devroye. Random variate generation in one line of code. In J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, editors, *Proceedings of the 1996 Winter Simulation Conference*, pages 265–272. IEEE Press, Piscataway, NJ, USA, 1996. [24, 25, 26, 64]

[47] P. Diaconis and M. Shahshahami. The subgroup algorithm for generating uniform random variables. *Prob. in Eng. and Info. Sci.*, 1:15–32, 1987. [40]

[48] U. Dini. *Serie di Fourier e altre rappresentazioni analitiche delle funzioni di una variabile reale.* T. Nistri, Pisa, 1880. [67]

[49] J. Doornik. An improved Ziggurat method to generate normal random samples. Unpublished manuscript, http://www.doornik.com/research/ziggurat.pdf, 2005. [61, 63]

[50] J. L. A. Dubbeldam, A. Milchev, V. G. Rostiashvili, and T. A. Vilgis. Driven polymer translocation through a nanopore: A manifestation of anomalous diffusion. *Europhys. Lett.*, 79:18002, 2007. [10]

[51] J. L. A. Dubbeldam, A. Milchev, V. G. Rostiashvili, and T. A. Vilgis. Polymer translocation through a nanopore: A showcase of anomalous diffusion. *Phys. Rev. E*, 76:010801(R), 2007. [10]

[52] A. Edelman. Eigenvalues and conditon numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9:543–560, October 1988. [32]

[53] A. Edelman. Random matrix theory. *Acta Numer.*, 14:233–297, 2005. [32]

[54] M. Evans and T. Swartz. Random variable generation using concavity properties of transformed densities. *J. Comput. Graph. Statist.*, 7:514–528, 1998. [51]

[55] Y. Fang, A. Brass, D. C. Hoyle, A. Hayes, A. Bashein, S. G. Oliver, D. Waddington, and M. Rattray. A model-based analysis of microarray experimental error and normalisation. *Nucleic Acids Res.*, 31(16), August 2003. [141]

[56] W. Feller. *An Introduction to Probability Therory and its Applications.* John Wiley, New York, 1957. [24]

[57] O. Flomenbom and J. Klafter. Closed-form solutions for continuous time random walks on finite chains. *Phys. Rev. Lett.*, 95:098105, 2005. [11]

[58] N. J. Ford and J. A. Connolly. Comparison of numerical methods for fractional differential equations. *Comm. Pure Appl. Anal.*, 5:289–307, 2006. [20]

[59] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Pers. Commun.*, 6:311–335, 1998. [32]

[60] D. Fulger. Adaptive brillouin zone integration (diplomarbeit). *Department of Physics, University of Dortmund and Technical Faculty, University Bielefeld*, 1999. [136, 156]

[61] D. Fulger and G. Germano. Automatic generation of non-uniform random variates for arbitrary pointwise computable probability densities by tiling, 2009. arXiv:0902.3088. [74]

[62] D. Fulger, E. Scalas, and G. Germano. Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation. *Phys. Rev. E*, 77:021122, 2008. [64, 82]

[63] G. Germano, M. Engel, and E. Scalas. City@home: Monte Carlo derivative pricing distributed on networked computers. In S. Cozzini, S. d'Addona, and R. Mantegna, editors, *Proceedings of the 1st International Workshop on Grid Technology for Financial Modeling and Simulation*, 2006. PoS(GRID2006)011, http://pos.sissa.it. [24, 26, 64]

[64] G. Germano, D. Fulger, and E. Scalas. mlrnd.m: Mittag-Leffler pseudo-random number generator, 2008. MATLAB Central File Exchange, file ID #19392, http://www.mathworks.com/matlabcentral/fileexchange. [21, 64, 74]

[65] W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, 41:337–348, 1992. [51]

[66] A. G. Glen, L. M. Leemis, and J. H. Drew. Computing the distribution of the product of two continuous random variables. *Comput. Statist. Data Anal.*, 4:451–464, 2004. [104]

[67] R. Gorenflo, J. Loutchko, and Yu. Luchko. Computation of the Mittag-Leffler function $E_{\alpha,\beta}(z)$ and its derivative. *Fract. Calc. Appl. Anal.*, 5:491–518, 2002. [74]

[68] R. Gorenflo and F. Mainardi. Fractional calculus: Integral and differential equations of fractional order. In A. Carpinteri and F. Mainardi, editors, *Fractals and Fractional Calculus in Continuum Mechanics*, pages 223–276. Springer, New York, 1997. Volume 378 of the series CISM Courses and Lectures. [18, 25]

[69] R. Gorenflo, F. Mainardi, and A. Vivoli. Continuous-time random walk and parametric subordination in fractional diffusion. *Chaos Soliton Fract.*, 34:87–103, 2007. [11, 21, 24]

[70] R. Gorenflo, A. Vivoli, and F. Mainardi. Discrete and continuous random walk models for space-time fractional diffusion. *Nonlinear Dynam.*, 38:101–116, 2004. [24]

[71] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Phys. Rep.*, 299:189–425, 1998. [31]

[72] E. Heinsalu, M. Patriarca, I. Goychuk, G. Schmid, and P. Hänggi. Fractional Fokker-Planck dynamics: Numerical algorithm and simulations. *Phys. Rev. E*, 73:046133, 2006. [11, 20]

[73] R. Hilfer and L. Anton. Fractional master equations and fractal time random walks. *Phys. Rev. E*, 51:R848–R851, 1995. [20, 74, 82]

[74] R. Hilfer and H. J. Seybold. Computation of the generalized Mittag-Leffler function and its inverse in the complex plane. *Integr. Transf. Spec. F.*, 17:637–652, 2006. [19, 64, 74]

[75] W. Hörmann and J. Leydold. Random-number and random-variate generation: automatic random variate generation for simulation input. In J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, editors, *Proceedings of the 32nd Winter Simulation Conference*, pages 675–682, San Diego, CA, USA, 2000. Society for Computer Simulation International. [49, 51]

[76] W. Hörmann and J. Leydold. Continuous random variate generation by fast numerical inversion. *ACM Trans. Mod. Comput. Simul.*, 13:347–362, 2003. [51, 52]

[77] W. Hörmann, J. Leydold, and G. Derflinger. *Automatic nonuniform random variate generation.* Springer, Berlin, 2004. Series: Statistics and Computing. [49, 51]

[78] R. A. Janik, M. A. Nowak, G. Papp, J. Wambach, and I. Zahed. Non-Hermitian random matrix models: Free random variable approach. *Physical Review E*, 55(4):4100–4106, 1997. [38, 41]

[79] K. Jayakumar. Mittag-Leffler process. *Math. Comput. Model.*, 37:1427–1434, 2003. [24, 25, 26, 64]

[80] B. Kaldor. Marginal productivity and macro-economic theories of distribution. *Rev. Econ. Stud.*, 33:309–319, 1966. [2]

[81] M. Kanter. Stable densities under change of scale and total variation inequalities. *Ann. Probab.*, 3:697–707, 1975. [26]

[82] A. J. Kindermann and J. F. Monahan. Computer generation of random variables using the ratio of uniform deviates. *ACM Trans. Math. Software*, 3:257–260, 1977. [77]

[83] S. Kotz, T. J. Kozubowski, and K. Podgorski. *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance.* Birkhäuser, Boston, 2001. [25]

[84] T. J. Kozubowski. Mixture representations of Linnik distribution revisited. *Statist. Probab. Lett.*, 38:157–160, 1998. [24, 25, 64]

[85] T. J. Kozubowski. Computer simulation of geometric stable distributions. *J. Comput. Appl. Math.*, 116:221–229, 2000. [24, 26, 64]

[86] T. J. Kozubowski. Fractional moment estimation of Linnik and Mittag-Leffler parameters. *Math. Comput. Model.*, 34:1023–1035, 2001. [24, 25, 64]

[87] T. J. Kozubowski and S. T. Rachev. Univariate geometric stable laws. *J. Comput. Anal. Appl.*, 1:177–217, 1999. [24, 26, 64]

[88] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 83:1467–1470, Aug 1999. [32, 33, 39, 93]

[89] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices, Aug 1999. [109]

[90] T. A. M. Langlands. Solution of a modified fractional diffusion equation. *Physica A*, 367:135–144, 2006. [11]

[91] M. Leccardi. Comparison of three algorithms for Lévy noise generation. Unpublished manuscript, 2005. [63, 164]

[92] P. L'Ecuyer. Uniform random number generators: A review. In S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, editors, *Proceedings of the 1997 Winter Simulation Conference*, pages 127–134, Piscataway, NJ, USA, 1997. IEEE Press. [49]

[93] P. L'Ecuyer. Random number generation. In J. E. Gentle, W. Haerdle, and Y. Mori, editors, *Handbook of Computational Statistics.* Springer, Berlin, 2004. [49, 51, 56]

[94] N. Lehmann. Principal components selection given extensively many variables. *Stat. Probab. Lett.*, 74:51–58, 2006. [130, 132]

[95] P. H. W. Leong, G. Zhang, D.-U Lee, W. Luk, and J. D. Villasenor. A comment on the implementation of the Ziggurat method. *J. Stat. Software*, 12:7, 2 2005. [61, 63]

[96] J. L. Leva. A fast normal random number generator. *ACM Trans. Math. Software*, 18:454–455, 1992. [49]

[97] P. Lévy. *Calcul des Probabilités.* Gauthier–Villars, Paris, 1925. [64]

[98] J. Leydold. Automatic sampling with the ratio-of-uniforms method. *ACM Trans. Math. Software*, 26:78–98, 2000. [51, 52]

[99] J. Leydold. A note on transformed density rejection. *Computing*, 65:187–192, 2000. [51]

[100] J. Leydold. Short universal generators via generalized ratio-of-uniforms method. *Math. Comput.*, 72:1453–1471, 2003. [49, 51]

[101] J. Leydold, G. Derflinger, W. Hörmann, and G. Tirler. An automatic code generator for non-uniform random variate generation. *Math. Comput. Simulation*, 62:405–412, 2003. [49, 51]

[102] J. Leydold and G. Tirler. Automatic non-uniform random variate generation in $\mathbb{R}$. In K. Hornik, F. Leisch, and A. Zeileis, editors, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, pages 1–6, 2003. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/. [50, 52, 61]

[103] S. Lin and B. W. Kernighan. An effective heuristic algorithm for the traveling-salesman problem. *Operations Res.*, 21:498–516, 1973. [130]

[104] R. J. Little and D. B. Rubin. *Statistical analysis with missing data.* John Wiley and Sons, 1987. [82, 106]

[105] M. Magdziarz and A. Weron. Competition between subdiffusion and Lévy flights: A Monte Carlo approach. *Phys. Rev. E*, 75:056702, 2007. [11, 20, 24]

[106] M. Magdziarz, A. Weron, and K. Weron. Fractional Fokker-Planck dynamics: Stochastic representation and computer simulation. *Phys. Rev. E*, 75:016708, 2007. [11, 20, 24]

[107] F. Mainardi and R. Gorenflo. On Mittag-Leffler-type functions in fractional evolution processes. 118:283–299, 2000. [10]

[108] F. Mainardi, M. Raberto, R. Gorenflo, and E Scalas. Fractional calculus and continuous time finance ii: the waiting-time distribution. *Physica A*, 287:468–481, 2000. [10, 20]

[109] P. Malliavin and M. Mancino. Fourier series method for measurement of multivariate volatilities. *Financ. Stoch.*, 6:49–61, 2002. [80]

[110] R. N. Mantegna and H. E. Stanley. Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight. *Phys. Rev. Lett.*, 73:2946–2949, 1994. [26, 51]

[111] H. M. Markowitz. Portfolio selection. *J. Finance*, 7:77–91, 1952. [32]

[112] G. Marsaglia. Expressing a random variable in terms of uniform random variables. *Ann. Math. Stat.*, 32:894–898, 1961. [49]

[113] G. Marsaglia. Generating a variable from the tail of the normal distribution. *Technometrics*, 6:101–102, 1964. [49, 65]

[114] G. Marsaglia and M. D. MacLaren. A fast procedure for generating exponential random variables. *Commun. ACM*, 7:298–300, 1964. [49, 50, 65]

[115] G. Marsaglia, M. D. MacLaren, and T. A. Bray. A fast procedure for generating normal random variables. *Commun. ACM*, 7:4–7, 1964. [49, 50, 65]

[116] G. Marsaglia and W. W. Tsang. A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. *SIAM J. Sci. Stat. Comp.*, 5:349–359, 1984. [49, 50, 55, 64, 65]

[117] G. Marsaglia and W. W. Tsang. The Ziggurat method for generating random variables. *J. Stat. Software*, 5:8, 2000. [50, 52, 55, 59, 60, 61, 62, 63, 64, 65, 73, 74]

[118] V. A. Marčhenko and L. A. Pastur. Distribution for some sets of random matrices. *Math. USSR-Sb.*, 1:457–483, 1967. [32, 82, 109]

[119] J. H. McCulloch. stabrnd.m: Stable random number generator, 1996. MATLAB script, http://economics.sbs.ohio-state.edu/jhm/jhm.html. [39]

[120] M. M. Meerschaert, David A. Benson, Hans-Peter Scheffler, and Boris Baeumer. Stochastic solution of space-time fractional diffusion equations. *Phys. Rev. E*, 65:041103, 2002. [11, 20, 65]

[121] M. M. Meerschaert, Hans-Peter Scheffler, and Charles Tadjeran. Finite difference methods for two-dimensional fractional dispersion equation. *J. Comput. Phys.*, 211:249–261, 2006. [20]

[122] M. S. Mega, P. Allegrini, P. Grigolini, V. Latora, L. Palatella, A. Rapisarda, and S. Vinciguerra. Power-law time distribution of large earthquakes. *Phys. Rev. Lett.*, 90:188501, 2003. [20]

[123] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: A fractional dynamics approach. *Phys. Rep.*, 339:1–77, 2000. [11]

[124] R. Metzler and J. Klafter. The restaurant at the end of the random walk: Recent developments in the description of anomalous transport by fractional dynamics. *J. Phys. A: Math. Gen.*, 37:R161–R208, 2004. [11]

[125] P. Minicozzi, F. Rapallo, E. Scalas, and F. Dondero. Accuracy and robustness of clustering algorithms for small-size applications in bioinformatics. *Physica A*, 387:6310–6318, 2008. [111]

[126] S. Mittnik, T. Doganoglu, and D. Chenyao. Computing the probability density function of the stable Paretian distribution. *Math. Comput. Modelling*, 29:235–240, 1999. [24, 164]

[127] S. Mittnik and S. T. Rachev. Alternative multivariate stable distributions and their applications to financial modeling. In S. Cambanis, G. Samorodnitsky, and M. S. Taqqu, editors, *Stable Processes and Related Topics*, pages 107–119. Birkhäuser, Boston, 1991. [25]

[128] E. W. Montroll and G. H. Weiss. Random walks on lattices. ii. *J. Math. Phys.*, 6:167–181, 1965. [16]

[129] A. L. Moustakas, S. H. Simon, and A. M. Sengupta. Statistical mechanics of multi-antenna communications: Replicas and correlations,. *Acta Phys. Polon. B*, 36:2719–2732, Sep 2005. [32]

[130] B. Nadler. Design flaws in the implementation of the Ziggurat and Monty Python methods (and some remarks on Matlab randn). 2008. [61, 63]

[131] R. M. Neal. Slice sampling. *Ann. Statist.*, 31:705–767, 2003. [50]

[132] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, 46:323–351, 2005. [26]

[133] M.E.J. Newman. Detecting community structure in networks. *Eur. Phys. J. B Condens. Matter Phys.*, 38:321–330, 2004. [136]

[134] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability.* Cambridge University Press, Cambridge, 2006. [34, 36, 37, 39]

[135] J. Nolan. Numerical calculation of stable densities and distribution functions. *Commun. Statist. — Stochastic Models*, 13:759–774, 1997. [24, 51, 53, 64, 66, 164]

[136] J. Nolan. An algorithm for evaluating stable densities in Zolotarev's (M) parametrization. *Math. Comput. Modelling*, 29:229–233, 1999. [24, 51, 53, 64, 66, 164]

[137] P. Ormerod. Random matrix theory and macro-economic time-series: An illustration using the evolution of business cycle synchronisation, 18862006. *Economics E-Journal*, 2:26, 2008. [32, 82]

[138] A. Ott, J.-P. Bouchaud, D. Langevin, and W. Urbach. Anomalous diffusion in "living polymers": A genuine Lévy flight? *Phys. Rev. Lett.*, 65:2201–2204, 1990. [10]

[139] A. G. Pakes. Mixture representations for symmetric generalized Linnik laws. *Statist. Probab. Lett.*, 37:213–221, 1998. [24, 25, 64]

[140] K. Pang, Z. Yang, S. Houa, and P. Leung. Non-uniform random variate generation by the vertical strip method. *Europ. J. Oper. Res.*, 142:595–609, 2001. [50, 61, 63]

[141] L. A. Pastur and V. Vasilchuk. On the Law of Addition of Random Matrices. *Commun. Math. Phys.*, 214:249–286, 2000. [39]

[142] K. Pearson. The problem of the random walk. *Nature*, 72:317, 1905. [13]

[143] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Phys. Rev. Lett.*, 83:1471–1474, 1999. [32, 33]

[144] I. Podlubny. *Fractional Differential Equations*. Academic Press, San Diego, 1999. [18, 27]

[145] I. Podlubny and M. Kacenak. mlf.m: Mittag-Leffler function —
Calculates the Mittag-Leffler function with desired accuracy, 2005. MATLAB Central File Exchange, file ID #8738,
http://www.mathworks.com/matlabcentral/fileexchange. [v, 22, 26, 27, 74]

[146] O. V. Precup and G. Iori. A comparison of high-frequency cross-correlation measures. *Physica A*, 244:252–256, 2004. [81]

[147] O. V. Precup and G. Iori. Cross-correlation measures in the high-frequency domain. *Europ. J. Finance*, 13:319–331, 2007. [81]

[148] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C++*. Cambridge University Press, Cambridge, 2nd edition, 2003. [iv, 27, 62]

[149] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes*. Cambridge University Press, Cambridge, 3rd edition, 2007. [vi, 49, 78, 164]

[150] M. Raberto, E. Scalas, and F. Mainardi. Waiting times and returns in high-frequency financial data: An empirical study. *Physica A*, 314:749–755, 2002. [20, 27, 82, 84]

[151] R. Renò. A closer look at the epps effect. *Int. J. Theoretical Appl. Finance*, 6:87102, 2003. [80]

[152] H. Rubin and B. Johnson. Efficient generation of exponential and normal deviates. *J. Stat. Comput. Simulation*, 76:509–518, 2006. [49, 50, 62, 65]

[153] A. I. Saichev and G. M. Zaslavsky. Fractional kinetic equations: Solutions and applications. *Chaos Soliton Fract.*, 7:753–764, 1997. [18]

[154] E. Scalas. The application of continuous-time random walks in finance and economics. *Physica A*, 362:225–239, 2006. [11, 21, 101, 102]

[155] E. Scalas, R. Gorenflo, H. Luckock, F. Mainardi, M. Mantelli, and M. Raberto. Anomalous waiting times in high-frequency financial data. *Quant. Financ.*, 4:695–702, 2004. [20, 84]

[156] E. Scalas, R. Gorenflo, and F. Mainardi. Fractional calculus and continuous time finance. *Physica A*, 284:376–384, 2000. [10]

[157] E. Scalas, R. Gorenflo, and F. Mainardi. Uncoupled continuous-time random walks: Solution and limiting behavior of the master equation. *Phys. Rev. E*, 69:011107, 2004. [17, 18, 21]

[158] H. Scher and E. W. Montroll. Anomalous transit-time dispersion in amorphous solids. *PHYS-REV-B*, 12:24552477, 1975. [11]

[159] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of. *J. Finance*, 19:425–442, 1964. [32]

[160] M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter. Strange kinetics. *Nature*, 363:31–37, 1993. [20]

[161] W. Silvert. Modelling as a discipline. *Int. J. General Systems*, 30:261–283, 2001. [160]

[162] J. O. Smith. *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications*. W3K Publishing, second edition edition, 2007. [147]

[163] I. M. Sokolov and J. Klafter. Field-induced dispersion in subdiffusion. *Phys. Rev. Lett.*, 97:140602, 2006. [10]

[164] M. Sokolov, A. Blumen, and J. Klafter. Linear response in complex systems: CTRW and the fractional Fokker-Planck equations. *Physica A*, 302:268–278, 2001. [11]

[165] R. Speicher. Free convolution and the random sum of matrices. *Publ. Res. Inst. Math. Sci.*, 29:731–744, 1993. [39]

[166] J. Steffen. Optimal boarding method for airline passengers. *J. Air Transp. Manag.*, 14:146–150, 2008. [32]

[167] C. Tadjeran, M. M. Meerschaert, and H.-P. Scheffler. A second-order accurate numerical approximation for the fractional diffusion equation. *J. Comput. Phys.*, 213:205–213, 2006. [20]

[168] H. Theil and K. Laitinen. Singular moment matrices in applied econometrics. In P. R. Krishnaiah, editor, *Multivariate analysis*, pages 629–649. North-Holland, 1980. [82, 106]

[169] D. B. Thomas and W. Luk. Non-uniform random number generation through piecewise linear approximations. *IET Comput. Digital Tech.*, 1:312–321, 2007. [51]

[170] D. B. Thomas, W. Luk, P. H. W. Leong, and J. D. Villasenor. Gaussian random number generators. *ACM Comput. Surv.*, 39:11, 2007. [49, 63, 65]

[171] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna. Cluster analysis for portfolio optimization. *J. Econ. Dyn. Control*, 32:235–258, 2008. [32, 93]

[172] P. Vivo, S. N. Majumdar, and O. Bohigas. Large deviations of the maximum eigenvalue in wishart random matrices. *J. Phys. A*, 40:4317–4337, 2007. [34]

[173] D. Voiculescu. Addition of certain non-commuting random variables. *J. Funct. Anal*, 66:323–346, 1986. [37]

[174] D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104:201–220, 1991. [40]

[175] U. von Luxburg. A tutorial on spectral clustering, 2007. [111]

[176] J. von Neumann. Various techniques used in connection with random digits. *NBS Appl. Math. Ser.*, 12:36–38, 1951. [24, 56, 57, 71]

[177] S. N. Ward. Earthquakes — a deficit vanished. *Nature*, 394:827–829, 1998. [20]

[178] R. Weron. Computationally intensive Value at Risk calculations. In J. E. Gentle, W. Haerdle, and Y. Mori, editors, *Handbook of Computational Statistics*, pages 911–950. Springer, Berlin, 2004. [64, 65, 66]

[179] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. Math.*, 67:325–327, 1958. [31]

[180] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–52, 1928. [31]

[181] G. M. Zaslavsky. Chaos, fractional kinetics and anomalous transport. *Phys. Rep.*, 371:461–580, 2002. [11]

[182] Y. Zhang, D. A. Benson, M. M. Meerschaert, E. M. LaBolle, and H.-P. Scheffler. Random walk approximation of fractional-order multiscaling anomalous diffusion. *Phys. Rev. E*, 74:026706, 2006. [11, 20, 105]