Habilitation à diriger les recherches

Spécialité : Mathématiques

Random holonomy,
large unitary matrices

Thierry LÉVY

Rapporteurs : M. Bruce DRIVER
M. David ELWORTHY
M. Wendelin WERNER

Soutenue le 20 novembre 2009 devant le jury compos de :

M. Philippe BIANE
M. Philippe BOUGEROL
M. David ELWORTHY
Mme Alice GUIONNET
M. Yves LE JAN
M. Wendelin WERNER
Introduction

This text is a presentation of my research in mathematics since the beginning of my PhD thesis. The structure of this text has been strongly influenced by the nature of this work and I shall first explain how.

There is one problem which has been constantly at the centre of my circle of interests, namely the construction and the study of the two-dimensional Yang-Mills measure. This measure is a mathematically tractable simplification of one of the fundamental ingredients of the formulation with Feynman integrals of gauge theories, which have been, since almost four decades, the best tool available to physicists for the description, at infinitesimal scale, of the ultimate constituents of matter and their interactions.

My PhD thesis was devoted to giving a second rigorous construction of this measure, after Ambar Sengupta had given the first one. Since then, I have used this construction to study several aspects of the measure and at the same time I have tried to improve it and to generalise it. This has led me to define a class of stochastic processes, which I call Markovian holonomy fields, which contains the Yang-Mills measure pretty much in the same way as the class of Lévy processes contains the Brownian motion. In addition to this, during the last few years, the study of the so-called large $N$ limit of the Yang-Mills measure has driven my attention towards questions about large unitary matrices the significance of which is, to a large extent, independent of my original motivation for considering them.

The thread running through my research to this date is therefore that which leads from the study of the Yang-Mills measure to the definition of Markovian holonomy fields. Nevertheless, neither the Yang-Mills measure nor these Markovian holonomy fields are classical or simply well-known objects. Thus, instead of writing a list of results with a short description of their context, I have preferred to conceive a large part of this text as an introduction to the theory of random holonomy fields.

The first two chapters of these notes are precisely that. They can be seen as an overview of the monograph [L7] or, even better, as a guide to reading it. The first chapter is probabilistic in nature and the second more geometric. Indeed, one of the peculiarities of the work that is described here is that it uses both the language of probability and that of differential geometry. Among many other options, I have chosen to write this text for a reader familiar with probability and not hostile to differential geometry. The most important geometric objects are thus described in detail, presumably in much more detail.

\footnote{This monograph will be published in the course of the year.}
than an expert would like. Moreover short reviews of various subjects are included in the form of inserts, so that the reader can easily refer to the ones which are useful to him and, just as easily, ignore those which are not.

The third chapter presents my work, partially in collaboration with Florent Benaych-Georges and Mylène Maïda, on large random unitary matrices. In this field, which is likely to be more familiar to many potential readers, I have reckoned that a shorter presentation of the context would be sufficient. Nevertheless, in this chapter too, the objects are carefully introduced and the progression of the text is punctuated by inserts.

Finally, as an appendix to these notes, I have added a text which I would have loved to have at hand when I was starting my own research on the Yang-Mills measure. In this text, I sketch a path, albeit superficial, from Maxwell’s equations to gauge theories, with an emphasis on the reason why it is natural to introduce principal bundles in the quantum mechanical formulation of electromagnetism. This text is a patchwork of fragments which I have gathered here and there, and its only originality may be the fact that it is written by a mathematician and for mathematicians.

**Presentation of the results**

I will now describe in greater detail the subject of my research, and in the first place how it is issued from theoretical physics.

**The mathematical difficulties of gauge theories**

The Yang-Mills measure bears the names of the two physicists Chen Ning Yang and Robert L. Mills who, in 1954, considered for the first time the possibility of gauge fields taking their values in non-Abelian Lie algebras in order to account for certain effects of the strong interaction [YM54]. In doing so, they opened the way to non-Abelian gauge theories, which have provided physicists with the framework needed to progressively build the so-called *standard model* which has been, since the beginning of the 1970’s the best available theory for the nature and interaction of the fundamental constituents of matter, with a conceptually unified description of the electromagnetic, weak, and strong interactions [CG07].

In the standard model, the interactions are conveyed by *gauge bosons*, like photons, which convey the electromagnetic interaction. These bosons are represented by fluctuations of a *gauge field*, whose mathematical nature is that of a *connection on a principal bundle*. The theory of connections was precisely formalised by Charles Ehresmann in the 1950’s [Ehr51], for internal reasons in mathematics. It is said that Chen Ning Yang and Shiing-Shen Chern, in a discussion around 1970, both wondered why the other found so natural to consider such objects in their own field.

From the physical point of view, a gauge theory is characterised by its *Lagrangian*, which expresses how the various fields under consideration interact. The Yang-Mills

\[^{2}\text{In order to imagine how two electrons can repulse each other by exchanging photons, one can think of two ice- skaters moving forward on parallel tracks and throwing a very heavy ball to each other, one after the other, thus each time inducing their trajectories to diverge slightly.}\]
Lagrangian, in a space without matter and with only gauge bosons, is given by the expression $L(A) = \text{Tr}(F \wedge \ast F)$, where $F$ is the strength field of the gauge field $A$, that is the curvature of the connection $A$. This Lagrangian being given, one defines the Yang-Mills action as the integral

$$S(A) = \int_{\mathbb{R}^4} L(A),$$

so that $S$ is a real-valued functional on the space, traditionally denoted by $\mathcal{A}$, of all connections. In accordance with Feynman’s formulation of quantum mechanics, one can then compute the probability that a certain physical event occurs by computing the square of the modulus of an integral of the form

$$\int_{\mathcal{B}} \Phi(A) e^{\frac{i}{\hbar} S(A)} dA,$$

where $\mathcal{B} \subset \mathcal{A}$ is a certain subset of connections, $\Phi : \mathcal{A} \to \mathbb{C}$ is a functional, and $dA$ a reference measure on $\mathcal{A}$.

Expressions like $[1]$ have no mathematical meaning. The most obvious problem is the fact that the space $\mathcal{A}$ of connections is not locally compact and does not carry any $\sigma$-finite Radon measure invariant by translations, although that is what $dA$ should be, according to the use that physicists make of it. Nevertheless, the incredible precision of the predictions which physicists extract from such integrals, and more generally from quantum field theories, indicate that it is possible and important to uncover the mathematical nature of the objects involved in these theories.

Generally speaking, finding a rigorous mathematical framework for the formulation of the standard model is still a widely open problem but in the case of a two-dimensional space-time and in a Euclidean situation (that is, replacing in $[1]$ the $i$ of the exponential with a $-1$), it is possible to give a rigorous meaning to integrals of the form $[1]$ which is consistent with their physical meaning. One can, in this way, give a meaning to the expression

$$\mu(dA) = \frac{1}{Z} e^{-\frac{1}{2} S(A)} dA,$$  \hspace{1cm} (2)

which is the most common heuristic description of the Yang-Mills measure, as a probability measure on the space $\mathcal{A}$ of all connections.

In the context which we will consider here, with a two-dimensional space-time, one can, as a first approximation, think of a connection as a differential 1-form on a surface. A typical functional of a connection is thus its integral along a path, or along a loop. In order to be less approximate, it is better to think of a connection as a multiplicative

\footnote{One could think that the gauge symmetry of the theory solves this problem, since it suffices to work on the quotient of $\mathcal{A}$ under the action of the gauge group. But, contrarily to a wrong idea which I have heard expressed several times, the quotient of the space of connections by the action of the gauge group is not locally compact nor finite-dimensional in any sense. However, the moduli space of flat connections, quotiented by gauge transformations on a principal bundle over a reasonable space (for example a space whose fundamental group is finitely generated) is a finite-dimensional object.}

\footnote{It is one of the six remaining problems for the resolution of which the Clay institute offers one million dollars. See \url{www.claymath.org/millenium/Yang-Mills_Theory}.}

\footnote{For more details on the discussion so far, one can consult Section 2.3 and in particular Definition 2.3.1.}
version of a differential form. A group $G$ is given, called the \textit{structure group}, which could for example be SU(2), and the connection determines, by integration along each path, an element of this group. This element is called the \textit{holonomy} of the connection along the path. Of course, the holonomy along the concatenation of two paths is the product of the holonomies along each path.

It happens that only holonomies along \textit{loops} carry meaningful information in gauge theories, because of the symmetry of these theories which is precisely called \textit{gauge symmetry}. The structure group $G$ being fixed, one can thus think of a connection as an object which to each loop traced on the surface associates an element of $G$. This can be written in a concise (but slightly incorrect) way as follows: if $M$ is a surface and $\text{L}(M)$ is the set of loops on $M$, and if $G$ is a compact Lie group, then we have the inclusion\footnote{For a definition of the gauge symmetry and a correct statement of this inclusion, see \ref{2.2.4} and what precedes it. For a discussion of the relation between gauge symmetry and the choice of a gauge in electromagnetism, see, in the appendix, Sections \ref{A.3} and \ref{A.4}.} $\mathcal{A} \subset \{\text{fonctions de } \text{L}(M) \text{ dans } G\}$.

From this point of view, defining a probability measure on the space of connections amounts to defining a probability on the set of functions from the set of loops on our surface and with values in the structure group $G$. This idea is absolutely fundamental for all that follows, as we are going to describe objects of the latter type: not random differential forms, but stochastic processes indexed by loops traced on surfaces and with values in a compact Lie group.

In order to help the imagination, let us say that the processes which we are going to consider are \textit{Lévy\footnote{This is \textit{Paul} Lévy, 1886-1971.} processes} indexed by loops and with values in a group. If the surface which we consider is for example a plane, and if for each $t \geq 0$ we consider the random variable associated with a circle which goes once positively around a disk of area $t$ (in such a way that the interiors of these circles increase with $t$, see Proposition \ref{1.1.3}) then we obtain a stochastic process indexed by $\mathbb{R}_+$ with values in $G$, whose multiplicative increments are independent and stationary. If the surface is a sphere, then our family of loops is going, after a time equal to the total area of the sphere, to shrink back to a constant loop after “going around the sphere”. In this case, we will rather obtain a \textit{bridge} of a Lévy process with values in $G$. Surfaces with different topologies will lead to variants of this bridge with values in $G$.

Part of the problems which we are going to consider are thus related to these processes, their definition, through discretisation, their characterisation, their classification, their geometrical meaning, and the study of their asymptotic behaviour in certain limits. We are in particular going to consider the limit in which the total area of the surface on which one works is multiplied by a constant which tends to zero (the \textit{semiclassical} limit of the theory), for which we present a large deviations principle, proved in collaboration with James R. Norris. The study of another limit, when the structure group is the unitary group $\text{U}(N)$ and $N$ tends to infinity, will not be discussed directly here, but is the motivation for tackling some questions related to large random unitary matrices which are presented below and will be exposed in detail in the last chapter. Some of these questions have been solved in collaboration with Florent Benaych-Georges and Mylène Maïda.
The construction of the Yang-Mills measure

After the seminal work of Yang and Mills in 1954 [YM54], one traditionally refers to a paper of Alexander A. Migdal written in 1975 [Mig75] as the first notable step towards a mathematically rigorous description of the discrete Yang-Mills measure in two dimensions. In this paper, Migdal identifies indeed, maybe for the first time, the fundamental role played by the heat kernel on the structure group. This kernel is not named, but written as the sum of its Fourier series.

In the 1980’s, Sergio Albeverio, Raphael Høegh-Krohn and Helge Holden examined, in an essentially discrete framework, the idea of a random measure on a surface taking its values in a Lie group [AHKH86, AHKH88]. In particular, they understood the importance for this problem of convolution semi-groups. The Markovian holonomy fields which are presented here are, as we shall see, in correspondence with convolution semi-groups of measures on the structure group and this can be seen as a continuation of the ideas of these authors.

The problem of the construction of the continuous Yang-Mills measure in two dimensions was first considered in the 1980’s by Leonard Gross [Gro85, Gro88], then studied by Bruce Driver [Dri89, Dri91] in the case where the surface is a plane, and was finally resolved by Ambar Sengupta on an arbitrary compact surface, around 1994 [Sen92, Sen97a]. The approach of Gross, Driver and Sengupta is based on the observation of the fact that, provided one works in an appropriate gauge, a connection taken under the Yang-Mills measure can, according to (2), be understood as a connection whose curvature has the distribution of a white noise, possibly conditioned in a way which depends on the topology of the surface on which one works. It is thus by giving himself a white noise and by making sense of a singular conditioning of this noise that Ambar Sengupta has achieved the construction of a sort of Brownian motion indexed by the loops traced on a compact surface and with values in a connected compact Lie group.

In the early 1990’s, Edward Witten published two papers on two-dimensional Yang-Mills theory [Wit91, Wit92], which of course raised great interest. In one of these papers [Wit91], Witten describes the main features of the discrete Yang-Mills theory and uses it to study the geometry of the moduli space of flat connections on a Riemann surface, and more specifically to compute its symplectic volume. The measure of symplectic volume of the moduli space of flat connections can indeed be understood as the semi-classical limit of the Yang-Mills measure, that is, its limit as the total area of the space-time tends to zero. I have not studied this aspect of the measure and I will not discuss it in this text.

Nevertheless, the large deviation principle James Norris and I have established, and to which I will come back in a moment, studies a limit, but of another nature, as the total area of the space-time tends to zero.

Around the same time, other more or less partial approaches of the problem of the

---

8This white noise takes its values in the Lie algebra of the structure group.
9This construction is analogous to the construction of the usual Brownian motion on [0, 1] using a white noise W on [0, 1], that is, an isometry of $L^2([0, 1], dt)$ into a linear space of centred Gaussian variables. One sets, for all $t \in [0, 1]$, $B_t = W(1, 0)$. Conditioning the noise by the equality $W(1, 0) = 0$, one gets a Brownian bridge.
10For more detail on this aspect of the measure, the reader is kindly invited to consult the works of Michael Atiyah and Raoul Bott [AB83], William Goldman [Gol84], Robin Forman [For93], Kefeng Liu [Liu96, Liu97], Ambar Sengupta and Christopher King [Sen97b, KS94].
construction of the Yang-Mills measure were proposed which, although not all mathematically completely satisfactory, have shed light on various aspects of this measure. I think in particular to the work of Dana Fine [Fin90, Fin91], Claas Becker [Bec95], Doug Pickrell [Pic96], John Baez and Stephen Sawin [BS97, BS98], Abhay Ashtekar et al. [AI92, AL94, AL95, AMMa94].

In my first research [L1, L2], I gave another construction of the Yang-Mills measure, and this constituted my PhD thesis, under the direction of Yves Le Jan. My approach was to rigorously ground the discrete theory described by Witten, and then to prove that it is possible to pass it to the continuous limit. More precisely, in order to construct a process of random holonomy indexed by loops traces on a surface, one describes its finite-dimensional marginals and then applies a version of Kolmogorov’s theorem. However, the fact that these processes are indexed by paths creates specific problems, as for example the fact that it is not possible to describe in a simple way all its marginals. Approximations are thus required, the quality of which determines the class of loops along which one will eventually be able to define the random holonomy. In [L2], I was able to define a process indexed by piecewise smooth paths and associated, in the sense alluded to a few pages ago, with the Brownian motion on a connected compact Lie group.

In a later work [L4], I generalised the discrete part of the construction in order to be able to include some topological refinements. This allowed James Norris and I, during our joint work in Cambridge [LN5], to prove a large deviation principle. This result is the first, and, to this day, the unique rigorous result which relates the Yang-Mills measure to the Yang-Mills energy functional. In this work, it was necessary to consider genuine principal bundles and genuine connections on these bundles, and to confront them with the functions of loops which were supposed to be their holonomy.

Besides, following Bergfinnur Durhuus [Dur80] and d’Ambar Sengupta [Sen94], I studied an algebraic aspect of the discrete Yang-Mills theory and improved a density result which is of some theoretical importance in order to justify the almost exclusive use that physicists make of Wilson loops as observables in gauge theories [L3]. More precisely, it is not obvious that the class of observables that is usually considered by physicists, namely the algebra generated by Wilson loops, which are the traces of the holonomies along loops, is complete, in the sense that it separates the points of the configuration space. The cases where the structure group is unitary or odd orthogonal, or a product of such groups, had been solved positively by Durhuus and then Sengupta. Using some classical invariant theory and another natural class of observables, the spin networks, invented in

\[\text{\cite{L1, L2}}\]

\[\text{\cite{L4}}\]

\[\text{\cite{Dur80, Sen94}}\]

\[\text{\cite{L3}}\]

\[\text{\cite{L5}}\]

\[\text{\cite{AI92, AL94, AL95, AMMa94}}\]

\[\text{\cite{Fin90, Fin91, Bec95, Pic96, BS97, BS98, AI92, AL94, AL95, AMMa94}}\]
the 1970’s by Roger Penrose and introduced in the 1990’s by John Baez in gauge theories (see [Bae96]), I have proved that the algebra generated by Wilson loops is complete when the structure group is any product of unitary, orthogonal and symplectic groups. The case of spin groups is still open.

Despite the existence of two rigorous approaches, that of Ambar Sengupta and mine, for the construction of the Yang-Mills measure, and despite the generalisations that I had added to mine, it seemed to me that several points needed to be clarified and that the construction had not at all reached its final form.

Large unitary matrices

Around 2005, I started to study an aspect of the Yang-Mills measure that I had ignored until then, namely its behaviour when the gauge group is the unitary group $U(N)$ and $N$ tends to infinity. This limit is usually simply called the large $N$ limit of the theory. Isadore Singer has published a short paper [Sin95] on this subject, in which he opens vast perspectives where essentially everything is still to be done. Michael Anshelevich developed and explained this paper as far as possible [Ans97]. Nevertheless, the literature in this field is still essentially the physical literature [GG95, GM95, GT93b, GT93a, Kaz81, KK80].

The first questions which arise in this direction of research can in fact be formulated quite independently from the Yang-Mills measure and are related to large unitary matrices taken under the heat kernel measure. These random matrices have been studied mainly by Philippe Biane [Bia97], Feng Xu [Xu97], Steve Zelditch [Zel04]. More recently, Ambar Sengupta [Sen08] has clarified the work of F. Xu.

In [L6], I established a relation between the Brownian motion on the unitary group and a very simple random walk on the symmetric group, which can be understood as a consequence of the Schur-Weyl duality or of the Itō formula. Understanding this relation has allowed me to give a combinatorial expression of the moments of the distribution of the eigenvalues of a unitary matrix of size $N$ taken under the heat kernel measure. This expression is a convergent power expansion in non-positive even powers of $N$, which can be interpreted rigorously as an expansion according to the genus of a certain ramified covering of a disk. This interpretation gives a rigorous proof of the simplest of the wonderful formulas given by David Gross and Washington Taylor in [GT93b, GT93a]. The expression that I have given allows one also to recover the moments of the limiting distribution of the eigenvalues, already computed by P. Biane, and results of asymptotic freeness (in the sense of free probability).

The familiarity with the unitary Brownian motion gained during this work allowed me, in collaboration with Florent Benaych-Georges, to construct in [LBGS] a family of structures of dependence (or independence) in a non-commutative probability space which interpolate between independence and freeness. It was known that, if $A$ and $B$ are two large diagonal real matrices whose eigenvalues are approximately distributed according to two probability distributions $\mu$ and $\nu$, and if $U$ is a permutation matrix chosen uniformly at random (resp. a unitary matrix chosen under the Haar measure), then the eigenvalues of $A + UBU^{-1}$ are distributed according to the measure $\mu * \nu$, the classical convolution of $\mu$ and $\nu$ (resp. according to $\mu \boxplus \nu$, the free convolution of $\mu$ and $\nu$). Giving to $U$ the distribution of a Brownian motion at time $t$ on the unitary group whose distribution at
time $0$ is the uniform measure on the group of permutation matrices, we have defined an operation of convolution $*_{t}$ for all non-negative real $t$, which for $t = 0$ is the classical convolution and for $t$ tending to infinity tends to the free convolution. In fact, we defined the structure of dependence between two sub-algebras of a non-commutative probability space which underlies this convolution. Our initial hope was to identify cumulants associated with this $t$-free convolution, that is, universal multilinear forms the cancellation of which would characterise the $t$-freeness of some of their arguments. We thought that they might interpolate between classical cumulants, which are intimately connected to the combinatorics of the partitions of a set, and free cumulants, related to the non-crossing partitions of a set endowed with a cyclic order. This hope has been turned down [*] and we have in fact proved that there are no $t$-free cumulants.

In a work in collaboration with Mylène Maïda [LM9], we established a central limit theorem for functions on the unitary group of the form $U \mapsto \text{tr} f(U)$, where $f$ is a sufficiently regular function on the unit circle of the complex plane and $\text{tr}$ is the normalised trace. The function $f$ being given, the expression $\text{tr} f(U)$ has a meaning for $U$ element of $U(N)$ for all $N$. The results of P. Biane [Bia97] describe the limit as $N$ tends to infinity of this function when $U$ has the distribution of a Brownian motion at time $t$. This limit is a deterministic and almost-sure limit. We have studied the fluctuations associated with this convergence, proved that they are Gaussian, and determined the quadratic form on the space of regular functions which gives the covariance of these fluctuations. This covariance is naturally expressed in the framework of free probability, in terms of the free unitary Brownian motion. We have been able to relate our result to that obtained by Persi Diaconis and Steven Evans [DE01] in the case of unitary matrices taken under the Haar measure, by showing that the covariance which they had identified, namely the scalar product in the Sobolev space $H^{1}$, is the limit of our covariance as time tends to infinity.

**Markovian holonomy fields**

In parallel to these works on large unitary matrices, I realised during the summer 2006 that it was possible to prove a certain property of continuity for the Yang-Mills measure, and that this allowed me to improve the apparently secondary and fairly technical point that is the class of loops with which one is working. In this respect, I was then able to work in a much more natural framework than ever before. The point was to extend the definition of random holonomy to the set of all loops with finite length, which is probably the largest class which one can deal with using only the techniques I have at my disposal [*]. The possibility of this extension has pushed me to reconsider the construction from the start and, progressively, I have been led to define Markovian holonomy fields [L7], thus completely generalising the construction that I had given in my PhD thesis. The Markov property of the random holonomy fields has become a central part of their very definition, which is a mixture of the definition of well-behaved classical Markov processes and topological quantum field theories. The class of objects which I have thus defined contains the Yang-Mills measure, which is its prototype, essentially as the class of Lévy processes contains the Brownian motion.

One of the important aspects of the greater generality brought by Markovian holonomy
fields is the possibility to consider non-connected structure groups and in particular finite groups. In the case of the symmetric group, a work by Alessandro d’Adda and Paolo Provero [DP04] showed a link between a theory analogous to Yang-Mills theory and a model of random ramified coverings. I have obtained a similar result in the general case, which allows one to realise explicitly a large class of Markovian holonomy fields with values in a finite group as monodromies of a random ramified covering\(^\text{14}\) whose ramification locus is essentially a Poisson point process.

**Localisation of the main statements**

To state the main results presented in these notes with sufficient precision would necessitate long introductions which, in fact, constitute most of the text to follow. I will thus simply indicate where in the text the main statements are located.

The main statements of the first chapter are:

- the definition of Markovian holonomy fields (Definition 1.3.3, p.37),
- a theorem of existence for these fields (Theorem 1.3.7, p.40),
- a theorem of extension which is the key for passing from the discrete theory to a continuous one (Theorem 1.2.17, p.35),
- a partial classification result of these fields (Theorem 1.3.6, p.39),
- the construction of the Yang-Mills fields starting from a white noise (Theorem 1.3.9, p.42),
- a density result for Wilson loops in the algebra of observables in a discrete gauge theory, for a certain class of structure groups (Theorem 1.2.11, p.31).

The main statements of the second chapter are:

- the geometric realisation of Markovian holonomy fields with values in finite groups by random ramified coverings (Theorem 2.1.5, p.51),
- the disintegration of the Yang-Mills field according to the possible isomorphism classes of principal \(G\)-bundles over a surface (Theorem 2.4.6, p.64),
- two large deviation results for the Yang-Mills measure as the total area of the surface tends to zero, proved in collaboration with James Norris (Theorems 2.4.8 and 2.4.9, p.65).

The main statements of the third chapter are:

- a theorem which expresses in combinatorial terms the moments of a unitary matrix taken under the heat kernel measure (Theorem 3.1.4, p.73),

\(^{14}\)Instead of a ramified covering, one should rather consider a ramified principal bundle, that is, a ramified covering on the regular fibres of which the structure group acts simply transitively.
• a formula due to Gross and Taylor, which I have been able to prove thanks to the previous result (Theorem 3.1.13, p.79).

Then, in collaboration with Mylène Maïda:

• a central limit theorem for the repartition of the eigenvalues of a unitary matrix under the heat kernel measure (Theorem 3.2.13, p.84).

• the convergence, as time tends to infinity, of the covariance which appears in the previous theorem to that identified by Persi Diaconis and Steven Evans in the case of uniformly distributed unitary matrices [DE01] (Theorem 3.2.14, p.84).

Finally, in collaboration with Florent Benaych-Georges:

• the definition of the $t$-free convolution of two probability measures on $\mathbb{R}$ (Theorem 3.3.1, p.86).

• the definition of $t$-freeness of two sub-algebras of a non-commutative probability space, interpolating between the notions of freeness and independence (Definition 3.3.5, p.89).

• a statement which encapsulates the differential system allowing one to compute the $t$-free convolution of certain distributions (Proposition 3.3.7, p.90).

• a result of non-existence of cumulants for the notions of $t$-freeness (Theorem 3.3.14, p.93).
Chapter 1

Markovian holonomy fields

1.1 Introduction

The aim of this introduction is to provide us with an overview of Markovian holonomy fields, before we take the subject from the beginning with an order of exposition close to the logical order. We are going to study a simple and fundamental example which will give us a sense of the kind of Markov property which is involved here. We will then give a temporary definition of these holonomy fields, and discuss their “transition kernels”, whose properties are intimately related to the surgery of surfaces and can be nicely drawn. We will finish this section by stating a result of existence and classification, which is one of the main achievements of our work on these processes.

Virtually all that is said in this introduction will be repeated in more detail later in this chapter.

1.1.1 The Poisson process modulo $n$ indexed by loops

Consider the Euclidean plane $\mathbb{R}^2$. Let us temporarily call loop a continuous mapping $l : [0, 1] \to \mathbb{R}^2$ such that $l(0) = l(1)$ and whose range is Lebesgue negligible. Let $l$ be a loop and let $z$ be a point of $\mathbb{R}^2$ identified with $\mathbb{C}$, not located on the range of $l$. The index of $l$ with respect to $z$ is the integer $n_l(z)$ defined by

$$n_l(z) = \frac{1}{2i\pi} \int_{\tilde{l}} \frac{dw}{w - z}, \tag{1.1}$$

where $\tilde{l}$ is any piecewise $C^1$ loop whose uniform distance to $l$ is smaller than the distance of $z$ to the range of $l$, so that $l$ and $\tilde{l}$ are homotopic in $\mathbb{R}^2 \setminus \{z\}$. Since we are assuming that the range of $l$ is negligible, the function $n_l$ is defined and finite almost everywhere on $\mathbb{R}^2$, and has compact support.

Let $n \geq 2$ be an integer. We are going to associate to each loop $l$ a random variable with values in $\mathbb{Z}/n\mathbb{Z}$. For this, we are going to add modulo $n$ the indices of $l$ with respect to the points of one or more Poisson point processes.
CHAPTER 1. MARKOVIAN HOLOMONY FIELDS \([L_2, L_3, L_7]\)

**Definition 1.1.1** Let \(\lambda_1, \ldots, \lambda_{n-1}\) be non-negative reals. Let \(\Pi_1, \ldots, \Pi_{n-1}\) be independent Poisson point processes on \(\mathbb{R}^2\) such that for all \(k \in \{1, \ldots, n-1\}\), the intensity of \(\Pi_k\) is \(\lambda_k\) times the Lebesgue measure. To each loop \(l\) we associate the random variable

\[
N_l = n \sum_{k=1}^{n-1} k \int_{\mathbb{R}^2} n_k(z) d\Pi_k(z) \pmod{n}.
\]

Figure 1.1: Take \(n = 4\). The processes \(\Pi_1, \Pi_2\) and \(\Pi_3\) are respectively represented by black squares, blue crosses and red disks. So, the contribution of \(\Pi_1\) is 2, that of \(\Pi_2\) is 6 and that of \(\Pi_3\), 6. Hence, in this example, \(N_l = 2 \pmod{4}\).

Let us temporarily denote by \(L(\mathbb{R}^2)\) the set of loops on \(\mathbb{R}^2\). We have just defined a stochastic process indexed by \(L(\mathbb{R}^2)\) and taking its values in \(\mathbb{Z}/n\mathbb{Z}\). Let us state a few properties of this process. Given a loop \(l\), we denote by \(l^{-1}\) the same loop traced backwards. Also, given two loops \(l_1\) and \(l_2\) based at the same point, we denote their concatenation by \(l_1l_2\).

**Proposition 1.1.2** The process \((N_l)_{l \in L(\mathbb{R}^2)}\) is multiplicative. This means that for all loops \(l, l_1, l_2 \in L(\mathbb{R}^2)\) such that \(l_1\) and \(l_2\) are based at the same point, we have

\[
N_{l^{-1}} = -N_l,
\]

\[
N_{l_1l_2} = N_{l_1} + N_{l_2}.
\]

Let us now consider a very simple one-parameter family of loops. For each \(t \geq 0\), let us denote by \(l_t\) a loop which bounds once and positively a disk of area equal to \(t\); these disks being chosen in such a way that for all \(s \leq t\), the disk of area \(t\) contains the disk of area \(s\). Then \((N_{l_t})_{t \geq 0}\) is an ordinary stochastic process, indexed by \(\mathbb{R}_+\), with values in \(\mathbb{Z}/n\mathbb{Z}\).

**Proposition 1.1.3** The process \((N_{l_t})_{t \geq 0}\) is a Markov chain on \(\mathbb{Z}/n\mathbb{Z}\). Its generator \((Q_{a,b})_{a,b \in \mathbb{Z}/n\mathbb{Z}}\) is given, for all \(a \neq b\), by \(Q_{a,b} = \lambda_c\), where \(c\) is the unique integer in \(\{1, \ldots, n-1\}\) such that \(b = a + c \pmod{n}\).

The increments of this process are independent and stationary: it is a Lévy process (that is, a continuous time random walk), whose jump measure, which is a positive measure on \((\mathbb{Z}/n\mathbb{Z}) \setminus \{0\}\), gives the mass \(\lambda_k\) to the class of \(k\) for all \(k \in \{1, \ldots, n-1\}\).
1.1. INTRODUCTION

We say that the Lévy process described in the proposition above is associated with the process \((N_l)_{l \in \mathbb{L}(\mathbb{R}^2)}\). Its distribution does not depend on the way in which we have chosen the loops \((l_t)_{t \geq 0}\). From the Markov property of this process follows a Markov property of the process \(N\), which can be expressed as follows. If \(J\) is a loop which is also a Jordan curve, we will denote respectively by \(\text{int}(J)\) and \(\text{ext}(J)\) the closures of the bounded and unbounded connected components of the complement in \(\mathbb{R}^2\) of the range of \(J\).

**Proposition 1.1.4** Let \(J\) be a Jordan curve. Then the two \(\sigma\)-fields \(\sigma(N_l : l \subset \text{int}(J))\) and \(\sigma(N_l : l \subset \text{ext}(J))\) are independent given \(N_J\).

More generally, let \(J, J_1, \ldots, J_n\) be pairwise disjoint Jordan curves such that the interiors of \(J_1, \ldots, J_n\) are pairwise disjoint and contained in the interior of \(J\). Define

\[
M_1 = \text{int}(J) \cap \text{ext}(J_1) \cap \ldots \cap \text{ext}(J_n) \quad \text{and} \quad M_2 = \text{ext}(J) \cup \text{int}(J_1) \cup \ldots \cup \text{int}(J_n).
\]

Then the two \(\sigma\)-fields \(\sigma(N_l : l \subset M_1)\) and \(\sigma(N_l : l \subset M_2)\) are independent conditional on \(\sigma(N_J, N_{J_1}, \ldots, N_{J_n})\).

Figure 1.3: Left : loops located in the exterior of \(J\) may wind around the interior of \(J\). Right : the domain \(M_2\) is shaded.
1.1.2 A first definition of Markovian holonomy fields

The proposition 1.1.4 expresses a Markov property where the notions of past and future are replaced by the notions of interior and exterior, with respect to one or more curves which split the plane into several connected components. Even applying this Markov property on such a simple domain as $\mathbb{R}^2$, we find ourselves dealing with more complicated domains such as the domain $M_1$ depicted (in white) in the right part of Figure 1.3 homeomorphic to a sphere with five holes.

Along with the cutting operation which has for instance lead us from the plane to the five-holed sphere, gluing operations are also natural. If for example we start from Poisson processes indexed by loops traced on two disjoint disks, both conditioned by the fact that they have a certain value $x \in \mathbb{Z}/n\mathbb{Z}$ on the boundary of their respective disk, then it is natural to glue these disks in order to define the Poisson process indexed by loops traced on a sphere, or at least the conditional version of this process determined by the event that its value along a certain equator is $x$.

In order to formulate the Markov property in full generality, one should thus not restrict oneself to the plane $\mathbb{R}^2$ nor to a disk, but consider a class of domains which is stable under the operations of cutting along Jordan curves and gluing along components of the boundary. Such a class is easy to identify: as soon as one starts with disks and three-holed spheres, one can, by gluings, construct any compact surface. It is thus the class of compact surfaces which constitutes the natural class of domains on which the loops which index our processes will be traced.

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
\textbf{INSERT 1 – Topology of surfaces} [Mas77, Moi77] \\
\hline
\end{tabular}
\end{table}

A compact surface, or simply a surface, is a compact smooth 2-dimensional real manifold which may or may not be orientable and may have a boundary. As a topological space, it is a Hausdorff compact space of which each point admits a neighbourhood homeomorphic to $\mathbb{R}^2$ or to $\mathbb{R} \times \mathbb{R}_+$. The differentiable structure carries no information which is not already contained in the topological structure: two surfaces are diffeomorphic if and only if they are homeomorphic.

The boundary of a surface, if not empty, is a finite disjoint union of circles. A surface whose boundary is empty is called a closed surfaces. To each surface one can associate a closed surface by gluing a disk along each connected component of the boundary. Two connected surfaces with the same number of connected components of boundary are diffeomorphic if and only if the associated closed surfaces are diffeomorphic. Finally, connected closed surfaces are classified up to diffeomorphism by their orientability and their genus, which can be any non-negative integer for an orientable surface, and any positive integer for a non-orientable surface. When one is working at the same time with orientable and non-orientable surfaces, it is more convenient to consider their reduced genus, equal to the genus for non-orientable surfaces and twice the genus for orientable ones. The main advantage of this reduced genus is that it is additive with respect to the operation of connected sum.

The closed surface of genus 0 is the sphere, the closed oriented surface of genus $g$ (and reduced genus $2g$) is the connected sum of $g$ tori, and the non-orientable surface of genus $g$ is the connected sum of $g$ projective planes. The projective plane can be obtained from the unit sphere of $\mathbb{R}^3$ by identifying all pairs of antipodal points.

Unlike in the 1-dimensional case, an orientation of an orientable surface does not allow one to distinguish among the connected components of its boundary between incoming and outcoming components. In fact, any permutation of the boundary components of a surface, even an oriented one, can be realised by a diffeomorphism, actually by an orientation-preserving diffeomorphism if the surface is oriented.
Just as the Poisson process described in the previous section was taking its values in \( \mathbb{Z}/n\mathbb{Z} \), Markovian holonomy fields will take their values in a group, which in general will not be Abelian. It will rather be a compact Lie group, that is to say, a compact matrix group. We will also insist that holonomy fields satisfy a multiplicativity property analogous to that expressed by Proposition 1.1.2, namely the following (see also Definition 1.2.1 below) : if \( G \) is a group, \( M \) a surface and \( m \) a point of \( M \), a family \( (H_l)_{l \in \Lambda_m(M)} \) of random variables indexed by the set of loops on \( M \) based at \( m \) is multiplicative if

\[
\forall l_1, l_2 \in \Lambda_m(M) \ , \ H_{l_2}^{-1} H_{l_1}^{-1} \text{ and } H_{l_1} H_{l_2} = H_{l_2} H_{l_1} \text{ almost surely.}
\]

The reason why the order of \( l_1 \) and \( l_2 \) is reversed in the last equality will be explained in Section 2.2.2.

Another important aspect of the Poisson process indexed by loops is the role played by the Lebesgue measure on \( \mathbb{R}^2 \) in its construction. It was indeed, up to multiplicative constants, the intensity of the Poisson point processes \( \Pi_1, \ldots, \Pi_{n-1} \). In the general case, we need to endow the surfaces which we consider with a measure of area, which as we shall see plays a role analogous to a measure of lengths in the usual case of Markov processes indexed by intervals of time.

We can now give a temporary and loose definition of Markovian holonomy fields.

**Definition 1.1.5 (Markovian holonomy fields : temporary definition)** Let \( G \) be a compact Lie group. A bidimensional \( G \)-valued Markovian holonomy field is a collection of \( G \)-valued stochastic processes, one for each compact surface endowed with a measure of area and a choice of boundary conditions along some of its boundary components. For each such surface, the process is indexed by the set of loops traced on the surface and it is multiplicative. Finally, the collection of all these processes behaves with respect to the surgery of surfaces (cutting along Jordan curves and gluing along boundary components) in a way which is governed by the Markov property.

### 1.1.3 Transition kernels and surgery of surfaces

In order to understand the role of the boundary conditions which are mentioned in the definition above, we are going to make an analogy with the more familiar case of Markov processes indexed by intervals of time. The data of such a process, taking its values in a state space \( \mathcal{X} \), allows one to associate to each positive real \( t \) and each initial condition \( x \in \mathcal{X} \), a probability measure \( P_t(x, dy) \) on \( \mathcal{X} \). The transitions of this process occur along intervals, which can be thought of, albeit pedantically, as topological transitions (or cobordisms) between pairs of points. Similarly, the transitions of a Markovian holonomy field occur along surfaces seen as cobordisms between families of circles (see for example Figure 1.4 below).

Nevertheless, according to the remark made at the end of Insert 1, and contrarily to what happens on intervals, the boundary of a surface, even an oriented one, does not split canonically into incoming and outgoing connected components. Besides, a surface may have an arbitrary non-negative number of boundary components. Looking at a particular
surface as realising a transition between certain connected components of its boundary and the others is thus the result of an arbitrary decision. The analytic counterpart of this higher symmetry between connected components of the boundary as between the initial and final point of an interval is the necessity to abandon the distinction which makes the transition kernel $P_t(x, dy)$ of a usual Markov process a function in $x$ and a measure in $y$. In the usual case, this distinction can be overlooked when the process is regular enough and, for all $t > 0$ and all $x \in X$, the measure $P_t(x, dy)$ admits a density, usually denoted by $p_t(x, y)$ with respect to some reference measure on $X$. The theory of Markovian holonomy fields as we develop it restricts itself to fields which are regular enough to ensure that we are in an analogous situation, where all transition kernels have a density.

It is time to describe these kernels. Just as the transition of a usual process along an interval depends only on the length of this interval and not on its position in the real line, the transition of a Markovian holonomy field along a surface depends on this surface essentially only up to diffeomorphism. A surface is characterised by its orientability, which we denote by a sign, $+$ or $-$, by the number of connected components of its boundary, which we denote by $p$, and by its reduced genus, which we denote by $g$ (see the definition of this reduced genus in Insert 1). This topological structure is completed by what plays the role of the length of an interval, namely a total area, denoted by $t$, of a measure of area with which we endow the surface. A theorem of Moser asserts that the only invariant under diffeomorphisms of such a measure of area is precisely its total area. The quadruple $(\pm, p, g, t)$ characterises completely the surface up to diffeomorphism and the transition kernels of a Markovian holonomy field are a family of functions

$$Z_{\pm}^{\pm}_{p, g, t}: G^p \to \mathbb{R}_+,$$

analogous to the densities $p_t(x, y)$ alluded to earlier.

Let us now consider a surface whose boundary has $p$ connected components, with reduced genus $g$ and total area $t$. Let us choose $q$ connected components of its boundary and declare them incoming, whereas the $p - q$ others are declared outgoing. If we fix “initial” conditions, namely $q$ elements $x_1, \ldots, x_q \in G$, along the $q$ incoming boundary components, for each of which we have chosen an orientation, then the field conditioned to satisfy these initial conditions determines a process indexed by the loops traced on our surface and we can in particular consider the distribution of the random variables associated with the $p - q$ outgoing connected components, seen as $p - q$ special loops. This distribution is given by

$$Z_{p, g, t}^{\pm}(x_1^{-1}, \ldots, x_q^{-1}, y_1, \ldots, y_{p-q}) \ dy_1 \ldots dy_q.$$

The reference measure $dy$ on the group $G$ is its normalised Haar measure, that is, its unique Borel probability measure which is invariant under translations. The exponents which appear take care of our decision to consider some components as incoming and the others as outgoing, and make the function $Z_{p, g, t}^{\pm}$ symmetric in all its arguments.

The usual Markov property, for processes indexed by time, can be expressed by the fact that the transition kernels form a convolution semigroup. From the point of view of intervals seen as cobordisms between pairs of points, this property governs the behaviour of the process on a time interval which results from gluing together two intervals which
Figure 1.4: Here, $p = 7$, $q = 4$ and $g = 6$ (it is the reduced genus).

Figure 1.5: Here, $\varepsilon = \varepsilon' = +$, $p = 3$, $g = 0$, $p' = 2$ and $g' = 2$. The most natural orientation of the surfaces being understood, only the boundary component of the two-holed torus labelled by $z$ is negatively oriented. This accounts for the exponent $-1$ which it carries in the integral.

have one point in common. In the 2-dimensional context, gluing surfaces along some of their boundary components, or gluing two components of the boundary of a single surface, or even gluing with itself a component of the boundary of a surface, are operations which produce new surfaces and each of them gives rise to a relation between the transition kernels of a Markovian holonomy field, analogous to the Chapman-Kolmogorov equation.

If $\varepsilon$ and $\varepsilon'$ are two signs, we denote by $\varepsilon \wedge \varepsilon'$ the sign which is equal to $+$ if and only if both $\varepsilon$ and $\varepsilon'$ are $+$.

**Proposition 1.1.6** Let $(Z_{p,g,t}^\varepsilon)$ be the partition functions of a Markovian holonomy field. For all meaningful choice of the parameters, one has the following relations.

$$
\int_G Z_{p,g,t}^\varepsilon(x_1, \ldots, x_{p-1}, z) Z_{p',g',t'}^{\varepsilon'}(y_1, \ldots, y_{p'-1}, z^{-1}) \, dz = Z_{p+p'-2,g+g',t+t'}^{\varepsilon \wedge \varepsilon'}(x_1, \ldots, x_{p-1}, y_1, \ldots, y_{p'-1}).
$$

(1.4)

This relation corresponds to the gluing of two distinct surfaces along a boundary component of each.

This relation corresponds to the gluing of two boundary components of the same surface. Note that our convention on the reduced genus allows us not to distinguish between the orientable case (where the usual genus is increased by 1) and the non-orientable case (where the genus is increased by 2).
\[
\int_G Z_{p,g,t}^\varepsilon (x_1, \ldots, x_{p-1}, z^2) \, dz = Z_{p-1,g+1,t}^- (x_1, \ldots, x_{p-1}). \tag{1.6}
\]

This last relation corresponds to the gluing of one boundary component of a surface with itself, according to a fixed point free involutive orientation-preserving diffeomorphism. The surface thus obtained is always non-orientable, homeomorphic to the surface which one would have obtained by gluing a Möbius band along the boundary component.

The infinity of transition kernels of a Markovian holonomy field thus satisfy an infinity of relations, one for each operation of surgery. The fact, which we have mentioned earlier, that every compact surface can be obtained by gluing enough disks and three-holed spheres (see Figure 1.8), thus implies that the whole set of transition kernels is in fact determined by a small number of them.

The orientable surface of reduced genus 6 with 5 holes is realised by gluing several three-holed spheres and one disk. Such a decomposition is not unique and the one we have represented does not use a minimal number of elementary blocks.
Proposition 1.1.7 All transition kernels of a Markovian holonomy field are determined by the kernels \((Z^+_{1,0,t})_{t>0}\) and \((Z^+_{3,0,t})_{t>0}\), associated respectively to disks and three-holed spheres of all areas.

Moreover, under an assumption of regularity which we will make more explicit later, the kernels \((Z^+_{1,0,t})_{t>0}\) suffice to determine all the others.

In the case of the Poisson process indexed by loops, the kernels \((Z^+_{1,0,t})_{t>0}\), associated with disks, are the 1-dimensional marginal distributions of the process indexed by \(\mathbb{R}_+\) which we have described in Proposition 1.1.3. In the general case, we have the following result, under certain assumptions of regularity for the Markovian holonomy field.

Proposition 1.1.8 Consider a Markovian holonomy field with values in a group \(G\). The measures \((Z^+_{1,0,t}(x)\ dx)_{t>0}\) on \(G\) form a convolution semi-group on \(G\). They are the 1-dimensional marginal distributions of a \(G\)-valued Lévy process, which is said to be associated with the field under consideration.

With the unpleasant restriction that it is not proved in all cases that a Markovian holonomy field is completely determined by its transition kernels, the last result establishes a correspondence between Lévy processes on a compact Lie group and holonomy fields with values in this group. The next result is one of our fundamental results on Markovian holonomy fields. We will state it again, more precisely, at the end of this chapter, after we have introduced the necessary definitions.

Theorem 1.1.9 ([L7, Theorem 4.3.1]) Let \(G\) be a compact Lie group. For all \(G\)-valued Lévy process which satisfies certain technical assumptions, there exists a Markovian holonomy field which takes its values in \(G\) and to which this Lévy process is associated.

The proof of this result of existence is constructive and we will spend most of this chapter explaining how one manages to construct a Markovian holonomy field.

1.2 Multiplicative processes indexed by paths

1.2.1 The measureable space

Let us start by describing more precisely the objects which we consider. We give ourselves a surface \(M\). It is endowed with its differentiable structure, with an orientation if it is orientable and a Borel measure which allows us to measure areas. We will assume that this measure admits, in each coordinate chart, a smooth positive density with respect to the Lebesgue measure. We will denote it by \(\text{vol}\). If \(M\) is oriented, one can identify \(\text{vol}\) with a volume 2-form, that is, a non-vanishing differential 2-form. The class of paths which we consider is the class of rectifiable paths, that is, continuous paths with finite length. The fact that a path has finite length is not a metric notion, but indeed a differentiable one: it is invariant under diffeomorphisms. We will denote by \(P(M)\) the set of paths with finite length, taken up to increasing reparametrization. When \(c\) is a path, we will denote by \(c^{-1}\) the same path traced backwards, and \(c_0\) and \(c_1\) the origin and finishing point of \(c\). If \(c_1\) and \(c_2\) are two paths such that \(c_1 = c_2\), we denote their concatenation by \(c_1 c_2\).
The group which we are going to consider and in which the field is going to take its values is a compact Lie group, which we denote by $G$.

**INSERT 2 – Compact Lie groups**

A compact Lie group is a compact topological group which is also a differentiable manifold in such a way that product and inversion are smooth mappings. Such a group is isomorphic, by a group isomorphism which is also a diffeomorphism, to a closed subgroup of the unitary group $U(N)$ for some $N$. One loses thus only some intrinsic character of the description, but no generality, in thinking of a compact Lie group as a compact group of matrices with complex coefficients. This is what we do from now on, usually without mentioning it, as soon as we feel that it allows us to simplify the presentation. Observe that any finite group endowed with the discrete topology is a compact Lie group.

Let $G$ be a compact Lie group. It has two properties which are fundamental for us. Firstly, it carries a unique Borel probability measure which is invariant by all translations: the Haar measure. We will not denote this measure, meaning that we will simply write $\int_G f(g) \, dg$ for the integral of a continuous function $f : G \to \mathbb{R}$ with respect to this measure. Secondly, the Lie algebra of $G$ carries a scalar product which is invariant by adjunction. This Lie algebra can be defined as the linear subspace of $M_N(\mathbb{C})$ given by

$$g = \{ A \in M_N(\mathbb{C}) : \forall t \in \mathbb{R}, e^{tA} \in G \}.$$  

The real linear space $g$ carries a scalar product (and in general, more than one, even up to multiplication by positive scalars) invariant under the action of $G$ by adjunction, that is, by conjugation: $\text{ad}(g)A = gAg^{-1}$. The data of a scalar product on $g$ is equivalent to the data of a Riemannian metric on $G$ for which right and left translations are isometries. The Riemannian volume of such a metric is thus a multiple of the Haar measure.

In the case of the unitary group $U(N)$, the Lie algebra $u(N)$ is the space of skew-Hermitian matrices. The scalar product $\text{Tr}(A^*B) = -\text{Tr}(AB)$ is then an invariant scalar product. The bilinear form $\text{Tr}(A^*B) + \alpha \text{Tr}(A^*) \text{Tr}(B)$ is also invariant for all real $\alpha$, so that it is a scalar product as soon as it is a scalar product, which is the case for $\alpha > -\frac{1}{2}$.

We will later need some simple topological considerations on compact Lie groups. If a compact Lie group $G$ is not connected, then the connected components of the unit element is a normal subgroup of $G$ denoted by $G^0$ such that the quotient $G/G^0$ is finite.

Let us assume that $G$ is connected. The fundamental group of $G$, as the fundamental group of a topological group, is Abelian. In fact, given two continuous loops $\gamma_1, \gamma_2 : [0,1] \to G$ based at the unit element, the concatenation $\gamma_1 \gamma_2$ of these two loops is homotopic to their pointwise product $t \mapsto \gamma_1(t) \gamma_2(t)$, as well as to their pointwise product in the other order: $t \mapsto \gamma_2(t) \gamma_1(t)$. The pointwise multiplication endows free homotopy classes, without any conditions on the endpoints, with the structure of an Abelian group canonically isomorphic to $\pi_1(G,1)$, where 1 denotes the unit element. This group is finite if and only if the centre of $G$ is finite, in which case $G$ is said to be semi-simple. In this case, $G$ admits a universal covering by another compact Lie group. For example, the fundamental group of $SO(3)$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ and $SO(3)$ admits a universal covering, of degree 2, by $SU(2)$. In the case where the centre of $G$ is not finite, for example if $G = U(N)$, then a universal covering of $G$ is isomorphic to the product of a compact Lie group and $\mathbb{R}^n$ for a certain $n \geq 1$.

Let us describe the canonical space of a multiplicative process indexed by a subset of $\mathcal{P}(M)$ and with values in $G$.

---

1Since $U(N)$ can be seen as a subgroup of $SO(2N)$, one can just as well think of matrices with real coefficients.
Definition 1.2.1 Let $P$ be a subset of $\mathcal{P}(M)$. Let $G$ be a compact Lie group. A function $h : P \to G$ is said to be multiplicative if
1. $\forall c \in P, c^{-1} \in P \Rightarrow h(c^{-1}) = h(c)^{-1}$.
2. $\forall c_1, c_2 \in P, c_1c_2 \in P \Rightarrow h(c_1c_2) = h(c_2)h(c_1)$.

We denote by $M(P, G)$ the set of multiplicative functions from $P$ into $G$.

We are going to endow $M(P, G)$ with the $\sigma$-field $C$ generated by the evaluation mappings $h \mapsto h(c)$, for $c$ running through all of $P$. A random holonomy field on $M$ is a finite or probability measure on $(M(P, M), G, C)$.

1.2.2 Graphs

In order to construct a measure on $(M(P, M), G, C)$, we would like to follow the usual procedure for constructing a stochastic process, by first describing the finite-dimensional marginals and then applying a Kolmogorov theorem. This is unfortunately not possible directly here, because we are not able to write down in a simple way all the finite-dimensional distributions of the measures which we want to construct. Before we explain where this problem comes from, let us first describe those finite-dimensional marginals for which it is possible.

We will say that a loop $l : [0, 1] \to M$ is simple if it is injective on $[0, 1)$.

Definition 1.2.2 A graph on $M$ is a triple $\mathbb{G} = (V, E, F)$ such that
1. $E$ is a set of edges, that is, injective paths or simple loops, which do not meet the boundary of $M$ or are contained in it; we assume that $E$ contains the edge $e^{-1}$ as soon as it contains the edge $e$, and that two distinct edges which are not each other's inverse meet, if at all, only at some of their endpoints,
2. $V$ is the set of endpoints of the edges of $E$,
3. $F$ is the set of connected components of the complement in $M$ of the union of the ranges of the edges of $E$,
4. each element of $F$ is homeomorphic to an open disk of $\mathbb{R}^2$.

The elements of $V$ are called vertices of the graph, and the elements of $F$ its faces.

Let $\mathbb{G}$ be a graph on $M$. Let us denote by $P(\mathbb{G})$ the set of paths that can be constructed by concatenating edges of $\mathbb{G}$. We are going to describe a probability measure on $\mathcal{M}(P(\mathbb{G}), G)$. Since $P(\mathbb{G})$ is infinite, this is not properly speaking a finite-dimensional marginal of the holonomy field, but it is easy to check that the restriction mapping $\mathcal{M}(P(\mathbb{G}), G) \to \mathcal{M}(E, G)$ is a bijection. In other words, a multiplicative mapping on paths traced in a graph is completely determined by its values on the edges. If we choose moreover an orientation of $\mathbb{G}$, that is a subset $E^+$ of $E$ such that $E$ is the disjoint union of $E^+$ and $(E^+)^{-1}$, then it suffices to know a multiplicative function on $E^+$: there is also a bijective restriction mapping $\mathcal{M}(E, G) \to \mathcal{M}(E^+, G)$. Finally, any function from $E^+$ to $G$ is multiplicative, because $E^+$ contains the inverse of none of its elements, nor any of their concatenations. Thus, $\mathcal{M}(E^+, G) = G^{E^+}$.

The measure $dg_{E^+}$, which is the Haar measure on the compact Lie group $G^{E^+}$, determines through the bijections which we have just described measures on $\mathcal{M}(E, G)$ and
\[ \mathcal{M}(P(G), G), \] which we call the uniform measures and denote simply by \( dh \). The measure which we are going to define on \( \mathcal{M}(P(G), G) \) has a density with respect to this uniform measure, which depends on the choice of a Lévy process on \( G \).

### 1.2.3 Lévy processes and marginals associated with graphs

Let us start by briefly describing Lévy processes on \( G \).

#### Insert 3 – Lévy processes on a compact Lie group

A process \( X = (X_t)_{t \geq 0} \) with values in \( G \) is said to be a Lévy process if it is càdlàg and if its increments are independent and stationary. Since the multiplication in \( G \) is not commutative, one needs to choose a side in order to defined increments, and we decide that the increment of \( X \) between \( s \) and \( t \) is \( X_t^{-1}X_s \) (one speaks of a left Lévy process). Let \( X \) be a Lévy process. Then the formula

\[ P_t f(x) = \mathbb{E}[f(xX_0^{-1}X_t)] \]

defines a left-invariant Feller semigroup, in the sense that for all \( x, y \in G \) and all Borel subset \( \Gamma \) of \( G \), one has

\[ P_t(yx, y\Gamma) = P_t(x, \Gamma). \]

Conversely, any homogeneous Feller process on \( G \) whose transition kernel is left-invariant is a left Lévy process. Finally, the data of a Lévy process is equivalent to the data of weakly continuous convolution semigroup of probability measures on \( G \), namely the semigroup of the 1-dimensional marginal distributions of \( X \).

The generator of a Lévy process is the sum of a diffusive term, a drift term and a jump term, the latter characterised by a measure denoted by \( \Pi \), which does not charge the unit element and satisfies the condition

\[ \int_G d_G(1, x)^2 \Pi(dx) < +\infty, \]

where \( d_G \) is a bi-invariant Riemanian distance on \( G \). We are going to consider processes issued from the unit element and invariant by conjugation, that is to say such that for all \( g \in G \), the processes \( X \) and \( gXg^{-1} \) have the same distribution. For such processes, the three parts of the generator are invariant by conjugation. This constraint is very strong on the diffusion and drift part : if the group \( G \) is simple, for example if \( G = SU(N) \), the the drift must be zero and the diffusive part of the generator is fully prescribed by the condition of invariance, up to multiplication by a non-negative constant.

We need to consider Lévy processes whose distribution at each positive time has a continuous density with respect to the Haar measure. The following result shows that this happens under an apparently much weaker assumption. Recall that \( G^0 \) is the connected component of the unit element in \( G \) and \( G/G^0 \) is finite.

**Proposition 1.2.3 (L7, Proposition 4.2.3)** Let \( X \) be a Lévy process invariant by conjugation on \( G \), issued from the unit element. If there exists \( p > 1 \) such that for all \( t > 0 \) the distribution of \( X_t \) admits a density which is in the \( L^p \) space of the Haar measure on \( G \), then this density, which we will denote by \((t, x) \mapsto Q_t(x)\),
is continuous on \((0, +\infty) \times G\) and positive on \((0, +\infty) \times G^0\). If moreover the support of the image measure of \(\Pi\) by the projection \(G \to G/G^0\) generates \(G/G^0\), then \(Q\) is positive on \((0, +\infty) \times G\).

**Definition 1.2.4** We will say that a Lévy process \(X\) on \(G\) is admissible if it is issued from the unit element, invariant in distribution by conjugation and if for all \(t > 0\) the distribution of \(X_t\) admits a density \(Q_t\) with respect to the Haar measure, such that \((t, x) \mapsto Q_t(x)\) is continuous and positive on \((0, +\infty) \times G\).

The group \(G\) admits a Riemannian metric such that the translations are isometries (see insert 2). Such a metric is not unique, but we choose one and we denote by \(d_G\) the associated Riemannian distance. Most of the time, we will write distances between the unit element and another element of \(G\). This distance can be replaced without damage in all formulas by a Euclidean distance in \(M_N(\mathbb{C})\) between, the identity matrix and the element under consideration.

Let us give examples of admissible processes.

- If \(G\) is a finite group, then the admissible processes are exactly the random walks whose jump measure is invariant by conjugation and has a generating support.
- If \(G\) is connected, the the Brownian motion on \(G\) is an admissible Lévy process.
- If \(X\) is an admissible Lévy process, one can add or subtract to its Lévy measure any finite measure supported by \(G^0\) without changing the fact that it is admissible.
- There exist pure jump admissible processes, like for example the process on \(SU(2)\) whose jump measure is \(\Pi(dx) = d_G(1, x)^\alpha \, dx\) for \(\alpha \in (-5, -3)\).

Let \(X\) be an admissible Lévy process on \(G\), of which we denote by \(Q\) the density. Let \(M\) be a surface endowed with a graph \(G\). We are now going to define a measure on \((M(P(G), G))\) associated to \(X\).

Observe that each face of \(G\) is bounded by a loop in \(G\). This apparently trivial remark is not completely straightforward: firstly, the loop is not unique, since it is defined up to orientation and choice of an origin, and secondly it is not necessarily a simple loop, since it can take twice the same edge, possibly twice in the same direction if the surface is not orientable (see Figure 1.10). In order to simplify the exposition, we will ignore the non-orientable case in most of the forthcoming statements. In an orientable and oriented surface, the boundary of a face \(F\) is thus a loop \(\partial F\) defined up to the choice of its origin. Thus, if \(h : P(G) \to G\) is a multiplicative function, the value of \(h(\partial F)\) is defined up to conjugation. This is enough for the next definition to make sense.

**Definition 1.2.5** Let \(M\) be a surface endowed with a measure of area denoted by \(\text{vol}\). Let \(G\) be a graph on \(M\). One defines the probability measure \(DF_{M, \text{vol}}^{X,G}\) on \((M(P(G), G), C)\) by

\[
DF_{M, \text{vol}}^{X,G}(dh) = \frac{1}{Z_{M, \text{vol}}^{X,G}} \prod_{F \in \mathcal{F}} Q_{\text{vol}(F)}(h(\partial F)) \, dh.
\]
It is possible to incorporate boundary conditions in the definition of $DF_{M,\text{vol}}^{X,G}$ but we prefer not to do so in order to keep the exposition as simple as possible. Let us simply indicate that the density is not affected by the boundary conditions, only the uniform measure $dh$ is conditioned to satisfy the desired boundary conditions. The normalisation constant $Z_{M,\text{vol}}^{X,G}$ is then modified and becomes a function of the boundary conditions.

Let us also mention that, in the non-orientable case, it is also necessary to assume that the process $X$ is invariant by inversion, that is, that it has the same distribution as $X^{-1}$.

![Figure 1.10: This graph on a Möbius band has two vertices, three edges and a single face, whose boundary goes twice through the equator (the bolder edge), in the same direction.](image)

We are now going to justify, to the extent possible, the definition 1.2.5 by finding a sense in which it is natural. For this, and in the first place, we discuss gauge symmetry.

### 1.2.4 Gauge symmetry

**Definition 1.2.6** The gauge group is the group $\mathcal{J} = \mathcal{F}(M, G)$ of all functions from $M$ to $G$. If $P$ is a subset of $\mathcal{P}(M)$, then $\mathcal{J}$ acts on $\mathcal{M}(P, G)$ in the following way: the element $j$ of $\mathcal{J}$ transforms the element $h$ of $\mathcal{M}(P, G)$ in the element $j \cdot h$ defined by

$$\forall c \in P, \quad (j \cdot h)(c) = j(c)^{-1} h(c) j(c).$$

For geometric of physical reasons which we will discuss later, one should consider on $\mathcal{M}(\mathcal{P}(M), G)$ only objects, measures or functions, which are invariant under the action of the gauge group. This restricts a great deal the space of observables, that is the space of permitted functions. For example, if $c$ is a path whose extremities are distinct, then the only functions $f : G \to \mathbb{R}$ such that $h \mapsto f(h(c))$ is invariant under the action of $\mathcal{J}$ are constant functions. Indeed, the value of $h(c)$ can be transformed into any element of $G$ by an appropriate gauge transformation. On the other hand, if $c$ is a loop, the any function $f : G \to \mathbb{R}$ which is invariant by conjugation, that is such that $f(yxy^{-1}) = f(x)$ for all $x, y$, gives rise to a function $h \mapsto f(h(c))$ which is invariant under gauge transformations.

A closer examination indicates that gauge transformations conjugate the values of $h$ on loops, but in fact conjugates by the same element the values of $h$ on loops based at the same point. Let us define the diagonal action of $G$ on $G^n$ for all integer $n \geq 1$ by $y \cdot (x_1, \ldots, x_n) = (yx_1y^{-1}, \ldots, yx_ny^{-1})$. Then for all function $f : G^n \to \mathbb{R}$ invariant by
diagonal conjugation and all family \( l_1, \ldots, l_n \) of loops based at the same point, the function 
\( h \mapsto f(h(l_1), \ldots, h(l_n)) \) is invariant under gauge transformations and it is essentially the most general one.

Before we state a result, let us give an example of diagonal conjugacy classes. In the group \( \text{SO}(3) \), each rotation which is neither the identity nor an axial symmetry (let us call such rotations proper rotations) has an axis, which is an oriented line, and an angle, which belongs to \((0, \pi)\). This angle characterises the conjugacy class of the rotation. Let us now consider two collections \((r_1, \ldots, r_n)\) and \((r'_1, \ldots, r'_n)\) of proper rotations. Then \( r_i \) is conjugated to \( r'_i \) for all \( i = 1, \ldots, n \) if and only if \( r_i \) and \( r'_i \) have the same angle for all \( i = 1, \ldots, n \). However, there exists a single rotation \( r \) such that \( r'_i = rr_ir^{-1} \) for all \( i = 1, \ldots, n \) if and only if, in addition to the previous condition, there exists a rotation (which is none other than \( r \)) which, for all \( i = 1, \ldots, n \), sends the oriented axis of \( r_i \) to the oriented axis of \( r'_i \). Thus, the diagonal conjugacy class of a family of rotations is given by the angles of these rotations and the relative position of their axes.

**Proposition 1.2.7** Let \( P \) be a subset of \( P(M) \) stable by inversion and by concatenation. Assume that each extremity of a path of \( P \) is joined to any other by a path of \( P \). Let \( m \) be an extremity of a path of \( P \). Let \( L_m \) be the set of loops based at \( m \) which belong to \( P \). Then the following two assertions hold.

1. The composed mapping \( \mathcal{M}(P,G) \rightarrow \mathcal{M}(L_m,G) \rightarrow \mathcal{M}(L_m,G)/G \), where the first mapping is the restriction and the second is the quotient by the action of \( G \) by diagonal conjugation on all factors, can be factored into an injection

\[
\mathcal{M}(P,G)/J \rightarrow \mathcal{M}(L_m,G)/G.
\]

2. Let \( c_1, \ldots, c_n \) be elements of \( P \). Let \( f : G^n \rightarrow \mathbb{R} \) be a continuous function such that the function \( h \mapsto f(h(c_1), \ldots, h(c_n)) \) is invariant under gauge transformations. Then there exists \( n \) loops \( l_1, \ldots, l_n \) based at \( m \) and a continuous function \( f : G^n \rightarrow \mathbb{R} \) invariant under the diagonal action of \( G \) and such that for all \( h \in \mathcal{M}(P,G) \) one has \( f(h(c_1), \ldots, h(c_n)) = f(h(l_1), \ldots, h(l_n)) \).

When one restricts to observables invariant by gauge transformations, one loses nothing in restricting oneself to loops, nor even to loops based at a single arbitrary point.

The proposition above makes hopefully the next definition understandable, of a \( \sigma \)-field on \( \mathcal{M}(P,G) \) which is smaller than the cylinder \( \sigma \)-field which we have considered so far.

**Definition 1.2.8** Let \( P \) be a subset of \( P(M) \) stable by inversion and by concatenation. Assume that each extremity of a path of \( P \) is joined to any other by a path of \( P \). One calls invariant \( \sigma \)-field, and one denotes by \( \mathcal{I} \), the smallest \( \sigma \)-field which makes all functions \( h \mapsto f(h(l_1), \ldots, h(l_n)) \) measurable, where \( l_1, \ldots, l_n \) are loops in \( P \) based at the same point and \( f : G^n \rightarrow \mathbb{R} \) is invariant under the diagonal action of \( G \) by conjugation.

Let us come back to the probability measure that we have defined on \( \mathcal{M}(P(G),G) \). The next result should come as no surprise.

**Lemma 1.2.9** The measure \( DF_{M,vol}^{X,G} \) is invariant under the action of the gauge group.
Let us conclude this paragraph by an observation on the nature of the boundary conditions which can be imposed to a holonomy field. When one puts boundary conditions to the measure $DF_{M,\text{vol}}^{X,G}$, one prescribes the value of multiplicative functions along connected components of the boundary of $M$. These connected components have on one hand no privileged origin and, on the other hand, would they have one, that these would necessarily be distinct. There is thus no meaning, up to gauge transformation, to prescribe more than the conjugacy class of the value of the functions along each component of the boundary. The boundary conditions thus consist of the data of one conjugacy class of $G$ for each oriented connected component of the boundary of $M$ (see Definition 1.3.2).

1.2.5 A random representation of the group of loops

The gauge invariance of the measure $DF_{M,\text{vol}}^{X,G}$ allows us to see it as a good object from the geometrical and physical point of view, but it does not suffice to make it look natural. In order to understand it better, let us study the structure of the set of loops based at a point in a graph.

---

**Insert 4 – Group of loops in a graph**

Let $G$ be a graph. Let $m$ be a vertex of $G$ and let $L_m(G)$ be the set of loops in $G$ based at $m$. The concatenation endows $L_m(G)$ with a structure of monoid. Let us declare that two loops are equivalent if one can transform one into the other by a sequence of insertions or erasures of paths of the form $ee^{-1}$ where $e$ is an edge of $G$. This relation is compatible with concatenation and the quotient of $L_m(G)$ by this equivalence relation is a group. Let us called reduced loop a loop which contains no sub-path of the form $ee^{-1}$. Then one proves that each equivalence class contains a unique reduced loop, which is also the unique element of shortest length in this class. One can thus choose to represent each class by its unique reduced representative, and consider as group operation the operation of concatenation-reduction. One obtains in this way a group, denoted by $RL_m(G)$, which is canonically isomorphic to the fundamental group $\pi_1(M \setminus Y, m)$ where $Y$ is a finite subset of $M$ which contains exactly one point in each face of $G$.

The group $RL_m(G)$ is free and the simplest way to exhibit a base of this group is to choose a spanning tree of $G$ and to associate to each edge not covered by this tree a loop of $RL_m(G)$. For this, to each edge $e$ which is not in the spanning tree, one associates the loop which starts from $m$, goes to the origin of $e$ by the unique injective path which does so in the tree, then traverses $e$ and finally goes back from the endpoint of $e$ to $m$ injectively along the tree.

Let $v = \#V$, $e = \#E/2$ and $f = \#F$ be the number of vertices, edges and faces of $G$. Let also $g$ denote the reduced genus of $M$ (that is twice the usual genus of the surface obtained from $M$ by gluing a disk along each component of the boundary of $M$) and $p$ the number of boundary components of $M$. Then there are $e - v + 1$ edges not covered by a spanning tree of $G$. The Euler characteristic of $M$ is equal to $2 - g - p$ on one hand and to $v - e + f$ on the other hand. Thus, the number of edges not covered by a spanning tree of $G$ is $g + p + f - 1$.

The group $RL_m(G)$ is thus a free group of rank $g + p + f - 1$. On a disk for example, this number equals $f$, the number of faces of $G$. It is possible to give one, indeed many bases of this group which are explicitly in one-to-one correspondence with the set of faces of $G$. Let us call lasso a loop which is of the form $sbs^{-1}$, where $s$ is a path issued from $m$ and $b$ is a simple loop based at the endpoint of $s$. The loop $b$ is called the meander (or buckle) of the lasso. We say that a lasso is facial if its meander is the boundary of a face of $G$. For a graph on a disk, it is possible to find bases of $RL_m(G)$ formed by facial lassos, one for each face of $G$, in such a way that the product of these lassos in a certain order is a lasso whose meander is the boundary of the disk. In the general case, the result is the following. Although $RL_m(G)$ is free, it is more natural to give a presentation of it which involves one relation, just as the space of finite real sequences $(a_0, \ldots, a_n)$ with sum equal to 0, although $n$-dimensional, is more naturally described as the kernel of a linear form in $\mathbb{R}^{n+1}$ than as the range of a linear mapping $\mathbb{R}^n \to \mathbb{R}^{n+1}$.
Proposition 1.2.10 Let $M$ be a connected surface with reduced genus $g$, whose boundary has $p$ connected components. Let $G$ be a graph on $M$ with $f$ faces. Let $m$ be a vertex of $G$. There exists $g$ lassos $a_1, \ldots, a_g$, a word $w$ in $g$ letters and their inverses, $p$ lassos $c_1, \ldots, c_p$, the meanders of which are the $p$ connected components of the boundary of $M$, and $f$ facial lassos $l_1, \ldots, l_f$ bounding the $f$ faces of $G$; such that the group $\text{RL}_m(G)$ admits the presentation
\[(a_i, c_j, l_k \mid w(a_1,\ldots,a_g)c_1\ldots c_p l_1\ldots l_f = 1),\]
and, for all continuous function $f : G^{g+p+f} \to \mathbb{R}$, one has
\[
\int_{\mathcal{M}(P(G),G)} f(h(a_1),\ldots,h(a_g),h(c_1),\ldots,h(c_p),h(l_1),\ldots,h(l_f)) \, dh
= \int_{G^{g+p+f}} f(x_1,\ldots,x_g,y_1,\ldots,y_p,z_1,\ldots,z_{f-1},z_f) \, dx_1 \cdots dy_p \cdots dz_{f-1},
\]
where we have set $z_i = (w(x_1,\ldots,x_g)c_1\ldots c_p z_1\ldots z_{i-1})^{-1}$.

The lassos $a_1,\ldots,a_g$ generate in particular, with the relation $w(a_1,\ldots,a_g) = 1$, the fundamental group of surface obtained from $M$ by gluing a disk along each boundary component.

With boundary conditions specified, the variables $h(c_i)$ would be uniform on the conjugacy classes which are prescribed to them. In the present case, we see that, under the uniform measure, the variables $h(l_1),\ldots,h(l_f)$ are just as independent as the topology of $M$ allows them to be.

Consider a multiplicative function $h : P(G) \to G$. We know by Proposition 1.2.7 that, up to gauge transformation, $h$ is characterised by its restriction to $L_m(G)$, where $m$ is a vertex of $G$ chosen arbitrarily. It follows immediately from the multiplicativity property of $h$ that two loops which are equivalent in the sense defined in the last insert have the same image by $h$. Thus, $h$ determines a multiplicative mapping from $\text{RL}_m(G)$ to $G$, which is nothing but a group homomorphism from $\text{RL}_m(G)$ into $G$. We can write, somewhat abstractly,
\[
\mathcal{M}(P(G),G)/\mathcal{J} \cong \text{Hom}(\text{RL}_m(G),G)/\text{Int}(G),
\]
where $\text{Int}(G)$ is the group of inner automorphisms of $G$, those which are given by the conjugation by a fixed element, which acts on the left by composition on $\text{Hom}(\text{RL}_m(G),G)$.

Thus, we see that the measure $DF_{M,\text{vol}}^{X,G}$ determines a random homomorphism from $\text{RL}_m(G)$ into $G$ whose distribution is invariant by composition by an inner automorphism. The distribution of such a random homomorphism is characterised by the distribution of the image of a generating set of $\text{RL}_m(G)$, which can be chosen to be a basis in order to avoid any kind of algebraic constraint on the image of this generating set, apart from the fact that it must be invariant by diagonal conjugation.

With the description that we have given of $\text{RL}_m(G)$, and given the Lévy process $X$ on $G$, the random homomorphism determined by the measure $DF_{M,\text{vol}}^{X,G}$ is as natural as it can be. On a disk for instance, it sends the lassos $l_1,\ldots,l_f$ of a basis described by Proposition 1.2.10 on independent random variables with respective distributions equal to those of $X_{\text{vol}(F_1)},\ldots,X_{\text{vol}(F_f)}$, where of course the faces have been numbered in such a way that the lasso $l_i$ bounds the face $F_i$. This description of the distribution of the random homomorphism depends on the basis of $\text{RL}_m(G)$ which one chooses and it is not
obvious that two different bases give rise, in distribution, to the same homomorphism. This can be checked directly as a consequence of a theorem of E. Artin on the action of the braid group on the free group \[ \text{Art47}. \]

On surfaces of higher genus, or with boundary conditions specified, the distributions of the images of the facial lassos are modified in the most natural possible way. For example, on a disk with boundary condition given by an element \( x_1 \) of \( G \), one conditions the Lévy process \( X \) to reach the conjugacy class of \( X \) at a time equal to the total area of the disk, and the images of \( l_1, \ldots, l_t \) have the distribution of the increments of this process. Another simple example is that of a sphere, where the process \( X \) is conditioned to come back to the unit element at a time equal to the total area of the sphere. One sees incidentally that one finds something very similar to what happens on a disk with boundary condition given by the unit element. Finally, one can for example describe what happens on a closed surface with reduced genus equal to 4: if the graph \( G \) is fine enough, one can choose the basis so that the word \( w \) in 4 letters and their inverses is \( aba^{-1}b^{-1}cdc^{-1}d^{-1} \). Then the facial lassos are sent to variables which have the distribution of the increments of the process \( X \) conditioned to have, at the time equal to the total area of the surface, the distribution of \( UVU^{-1}V^{-1}WZW^{-1}Z^{-1} \), where \( U, V, Z, W \) are uniform and independent on \( G \). For some graphs, the word \( w \) can be different, for instance \( abcd\overline{a}b^{-1}c^{-1}d^{-1} \), but one checks that the law of \( w(U_1, U_2, U_3, U_4) \) where \( U_1, U_2, U_3, U_4 \) are uniform and independent on \( G \) does not depend on the graph \( G \), but only on the topology of \( M \).

1.2.6 Wilson loops \[ \text{L3} \]

We are going to make a pause in the progression towards the definition of Markovian holonomy fields and to present a result which is concerned with the discrete theory that we have just described. This section can be skipped without damage for the understanding of the main construction.

Let \( G \) be a graph on a surface \( M \), although really the surface plays no role here. On the space \( \mathcal{M}(\mathcal{P}(G), G) \), considerations of gauge symmetry have led us to introduce functions of the form \( h \mapsto f(h(l_1), \ldots, h(l_n)) \) where \( l_1, \ldots, l_n \) are loops based at the same point and \( f \) is invariant by diagonal conjugation. Propositions \[ \text{1.2.7 and 1.2.10} \] allow us to check that these functions separate the points of the quotient space \( \mathcal{M}(\mathcal{P}(G), G)/J \). Thus, from the physical point of view, they generate a good algebra of observables.

However, in the study of discrete gauge theories, physicists often consider a much smaller algebra of observables, generated by simpler functions called Wilson loops. They are the functions of the form \( h \mapsto f(h(l)) \) where \( l \) is a loop and \( f \) is invariant by conjugation. The natural question which arises is whether these observables suffice to generate all continuous functions on the quotient space \( \mathcal{M}(\mathcal{P}(G), G)/J \), that is, according to the theorem of Stone-Weierstrass, whether they separate the points.

Let \( h \) be a multiplicative function. The question is the following : knowing the conjugacy class of \( h(l) \) for all loop \( l \), does one know \( h \) up to gauge transformation, that is, does one know the class of diagonal conjugation of \( (h(l_1), \ldots, h(l_n)) \) for all \( n \)-tuple of loops based at the same point ? Given \( l_1, \ldots, l_n \), we have access to the conjugacy classes of \( h(l_1), \ldots, h(l_n) \), but also to those of \( h(l_1l_2) = h(l_2)h(l_1), h(l_1l_2l_3) = h(l_3)h(l_2)h(l_1) \), or actually of the image by \( h \) of any other loop that can be formed by concatenating copies
of $l_1, \ldots, l_n$ and their inverses. We have thus access to the conjugacy class of any product of the elements $h(l_1), \ldots, h(l_n)$ and their inverses of $G$.

The question is now purely algebraic, in the group in which $h$ takes its values. If this group is $SO(3)$, one can put the question in the following way: given $n$ rotations $r_1, \ldots, r_n$ of $\mathbb{R}^3$, does the data of the angle of any rotation that can be formed by multiplying elements taken, with possible repetitions, in the set $\{r_1, \ldots, r_n, r_1^{-1}, \ldots, r_n^{-1}\}$ suffice to determine the relative position of the axes of $r_1, \ldots, r_n$?

In this case, the answer is positive.

**Theorem 1.2.11 ([L3, Proposition 3.4])** Let $G$ be a finite product of groups taken in the following list: unitary, special unitary, orthogonal, special orthogonal, symplectic. Let $r \geq 1$ be an integer. Let $g = (g_1, \ldots, g_r)$ and $g' = (g'_1, \ldots, g'_r)$ be two elements of $G^r$. Assume that for all word $w$ in $r$ letters and their inverses, the elements $w(g)$ and $w(g')$ of $G$ are conjugated. Then there exists $k \in G$ such that for all $i$ between 1 and $r$ one has $g'_i = kg_ik^{-1}$.

This theorem has been proved by B. Durhuus for unitary and special unitary groups [Dur80] and by A. Sengupta for unitary and odd orthogonal groups [?]. The proof of the theorem above relies on the use of spin networks, which are a natural set of invariant functions on $\mathcal{M}(P(G), G)$ of a different nature and for which it is straightforward to check that they generate a dense subalgebra of functions [Bae96]. One then proves, using the fundamental theorems of invariant theory, that these spin networks can be expressed in terms of Wilson loops.

### 1.2.7 Invariance by subdivision

Let us come back to the mainstream of the construction of Markovian holonomy fields. The normalisation constant $Z_{X,G,M,\text{vol}}$ which appears in Definition 1.2.5 and which, in presence of boundary conditions, is a function of these boundary conditions, will be, as the notation suggests, one of the transition kernels of the holonomy field associated to $X$. It seems however to depend on the graph $G$. A first important results states that it is not the case.

**Proposition 1.2.12 ([L7, Prop. 3.2.5, Prop. 4.1.10])** The number $Z_{X,G,M,\text{vol}}$ which, in presence of boundary conditions is a function $Z_{X,G,M,\text{vol}}(x_1, \ldots, x_p)$ of these conditions, does not depend on the graph $G$ and is denoted by $Z_{X,M,\text{vol}}^X(x_1, \ldots, x_p)$. If $M$ is orientable with reduced genus $g$, the one has the following expression:

$$Z_{X,M,\text{vol}}^X(x_1, \ldots, x_p) = \int_G Q_{\text{vol}(M)}(\left[a_1, a_2, \ldots, a_{g-1}, a_g\right] c_1 x_1 c_1^{-1} \cdots c_p x_p c_p^{-1}) \, da_idc_k,$$

where we have used the notation $[a, b] = aba^{-1}b^{-1}$.

If $M$ is non-orientable with genus $g$, then

$$Z_{X,M,\text{vol}}^X(x_1, \ldots, x_p) = \int_G Q_{\text{vol}(M)}(\left[a_1^2, a_2^2, c_1 x_1 c_1^{-1} \cdots c_p x_p c_p^{-1}\right] \, da_idc_k.$$
The expressions which appear here are reminiscent of the classical presentations of the fundamental groups of surfaces. This is not surprising since these presentations can be computed using polygonal models for these surfaces, the edges of which are identified by pairs, and that the same polygonal models produce graphs with a single face on these surfaces, in which the partition function can be computed to give almost directly the expressions above.

Now, we would like to bind all the measures $DF_X, GM, vol$ into a single measure on $M$ ($P(M), G$).

We have said at the beginning of Section 1.2.2 that this does not work directly. On one hand, the measures $DF_X, GM, vol$ do not suffice to describe all the finite-dimensional marginals of a measure on $M$ ($P(M), G$): in general, a finite number of paths, or even one single path, are not paths which can be realised by concatenating the edges of a graph. For example, even a very regular path can have a range whose complement in $M$ has infinitely many connected components. Another aspect of the same problem is the fact that the natural partial order on the set of graph does not have the required property in order to apply a theorem of the kind of Kolmogorov’s theorem. The partial order is the following.

**Definition 1.2.13** A graph $G_2$ is said to be finer than a graph $G_1$ if $E_1 \subset P(G_2)$ or, equivalently, if $P(G_1) \subset P(G_2)$.

The set of all graphs endowed with this partial order is not directed, which means that given two graphs, there does not always exist a third graph which is finer than the first two.

One solves these problems by temporarily restricting the class of paths which one considers on $M$. The point is to consider paths which, like piecewise affine or real analytic paths on $\mathbb{R}^2$ for instance, have a rigidity property which prevents them from intersecting or self-intersecting in a pathological way. The only differentiable structure on $M$ does not allow one to define such a class. It is thus necessary to enrich the structure of $M$, keeping in mind that the final object that we construct must be independent of this enrichment.

We choose to endow $M$ with a Riemannian metric, which we choose in such a way that the Riemannian volume coincides with the measure of area. We denote by $A(M)$ the set of piecewise geodesic paths on $M$, or by $A_\gamma(M)$ if we want to refer explicitly to the metric.

Given two graphs $G_1$ and $G_2$ such that $G_2$ is finer than $G_1$, one has a restriction mapping $M(P(G_2), G) \to M(P(G_1), G)$ induced by the inclusion $P(G_1) \subset P(G_2)$. The fundamental result of the discrete theory is the following property of invariance by subdivision. It is proved in [?], mainly at the proposition 4.3.4.

**Proposition 1.2.14** 1. Let $G_1$ and $G_2$ be two graphs such that $G_2$ is finer than $G_1$. Then the restriction mapping $M(P(G_2), G) \to M(P(G_1), G)$ sends the measure $DF_{M,vol}$ on the measure $DF_{M,vol}^{X,G_1}$.

2. Let $\gamma$ be a Riemannian metric on $M$. The family of probability spaces

$$\left\{ (M(P(G), G), I, DF_{M,vol}^{X,G}) : G \text{ piecewise geodesic graph} \right\}$$

is a consistent family. One can take its inverse limit and obtain a probability space $(M(A_\gamma(M), G), I, DF_{M,vol}^{X,G})$, so that the marginal distribution of the canonical process on this space associated to a graph with piecewise geodesic edges $G$ is the measure $DF_{M,vol}^{X,G}$. 

It is in the proof of this property that the property of semigroup of the 1-dimensional distributions of $X$ plays its role. Let us illustrate this on an example. Consider the two graphs represented on Figure 1.11. The graph $G_1$ is obtained from $G_2$ by removing the edge $e$. In $G_2$, this edge is located on the boundary of two faces $F_1$ and $F_2$, one bounded by the loop $e_1e_2e$ and the other by the loop $e^{-1}e_3e_4$. These faces are reunited in $G_1$ to form a single face $F$, bounded by the loop $e_1e_2e_3e_4$.

![Figure 1.11: Removing an edge to a graph is the most important elementary operation which allows one to go from one graph to a less fine one.](image)

Let $f : \mathcal{M}(\mathcal{P}(G_1), G) \rightarrow \mathbb{R}$ be a function. It induces a function $\tilde{f} : \mathcal{M}(\mathcal{P}(G_2), G) \rightarrow \mathbb{R}$, such that the value $\tilde{f}(h)$ of $\tilde{f}$ on a multiplicative function $h$ does not depend on the value of $h(e)$. The invariance by subdivision means that $\int_{\mathcal{M}(\mathcal{P}(G_2), G)} \tilde{f}(h) \, DF_{M,\text{vol}}^{X,G_2}(dh) = \int_{\mathcal{M}(\mathcal{P}(G_1), G)} f(h) \, DF_{M,\text{vol}}^{X,G_1}(dh)$. If we choose an orientation of $G_2$ (as we have partially done on the picture), we can see the integral on the right as an integral on $G_{E_1}^+$ and that on the left as an integral on $G_{E_2}^+$. There is thus one more variable in the integral on the left, which corresponds to the edge $e$. Only the density of $DF_{M,\text{vol}}^{X,G_2}$ depends on this extra variable, through the two terms associated to the two faces bounded by $e$: there is in the integral a term of the form $\int \ldots Q_{\text{vol}(F_1)}(h(e_4)h(e_3)h(e_1))Q_{\text{vol}(F_2)}(h(e_4)h(e_3)h(e))^{-1} \ldots d(h(e_1))$ which, by renaming the variables $h(e_1), h(e_2), h(e_3), h(e_4), h(e)$ as $x_1, x_2, x_3, x_4, x$ and using the invariance by conjugation of $Q$, can be rewritten perhaps more suggestively as

$$\int \ldots Q_{\text{vol}(F_1)}(x_2x_1x)Q_{\text{vol}(F_2)}(x^{-1}x_4x_3) \ldots dx.$$ 

The property of convolution semigroup of $Q$ allows us to integrate with respect to $x$, thus producing the term $Q_{\text{vol}(F_1)+\text{vol}(F_2)}(x_2x_1x_4x_3) = Q_{\text{vol}(F)}(x_4x_3x_2x_1)$, which is exactly the term corresponding to $F$ in the density of $DF_{M,\text{vol}}^{X,G_2}$.

Once the invariance by subdivision is established, one can apply a theorem of the kind of Kolmogorov’s theorem, in a version which is not very usual because of the size of the index space, but actually easier to prove than the more common versions, thanks to the compactness of $G$. 

1.2. MULTIPLICATIVE PROCESSES INDEXED BY PATHS
1.2.8 Extension to rectifiable loops

At the cost of an artificial supplement of structure on \( M \), we have now at our disposal a process with values in \( G \) indexed by piecewise geodesic loops. In order to go from piecewise geodesic loops to rectifiable ones, we are going to proceed by approximation, and this requires that we endow \( P(M) \) with an appropriate topology.

The data of a Riemannian metric on \( M \) makes it a metric space \( (M,d) \) and allows one to measure the length of rectifiable paths. We will denote by \( \ell(c) \) the length of a path \( c \). We define a distance on \( P(M) \) by setting, for all \( c,c' \) parametrized at constant speed,

\[
d_\ell(c,c') = \max \{ d(c(t), c'(t)) : t \in [0,1] \} + |\ell(c) - \ell(c')|.
\]

The topology of the metric space \( (P(M),d_\ell) \) does not depend on the Riemannian metric chosen on \( M \). This distance is easy to manipulate, but it has the unpleasant property that the space \( (P(M),d_\ell) \) is never complete. Indeed, it is easy to construct a sequence of paths of length 2 which converges uniformly while zigzagging towards a geodesic segment of length 1.

In fact, there exists another distance on \( P(M) \), which is complete and metrized the same topology as \( d_\ell \). It is the distance in 1-variation which, on \( \mathbb{R}^2 \) for example reads

\[
d_1(c,c') = \max \{ d(c(t), c'(t)) : t \in [0,1] \} + \int_0^1 |\dot{c}(t) - \dot{c}'(t)| \, dt.
\]

This distance is however much less easy to handle than \( d_\ell \) and it matters to us mainly insofar as it gives us a complete metrisation of the topology that we use.

Most of the time, because of the invariance under gauge transformations, we will consider convergences with fixed endpoints : we will say that a sequence \( (c_n)_{n \geq 0} \) of paths converges towards a path \( c \) with fixed endpoints if \( d_\ell(c_n,c) \) tends to 0 and if, for all \( n \geq 0 \), the path \( c_n \) has the same endpoints as \( c \).

The set \( A(M) \) of piecewise geodesic paths is of course dense in \( P(M) \) for the convergence with fixed endpoints.

In order to define a measure on \( \mathcal{M}(P(M),G) \), we need two things : a property of regularity of the measure \( DF^X_{M,vol} \) on \( \mathcal{M}(A(M),G) \) and a theorem of extension. Let us start with the first thing. Recall that \( d_G \) is the Riemannian distance on \( G \) for a bi-invariant metric.

**Proposition 1.2.15** Consider the measure \( DF^X_{M,vol} \) on \( \mathcal{M}(A(M),G) \). There exists \( K > 0 \) such that for all simple loop \( l \in A(M) \) bounding a disk \( D \) and such that \( \ell(l) < K^{-1} \), one has

\[
\int_{\mathcal{M}(A(M),G)} d_G(1,h(l)) \, DF^X_{M,vol}(dh) \leq K \sqrt{\text{vol}(D)}.
\]

This property, which expresses the fact that the holonomy along the boundary of a small disk is close to the unit element, can only be inherited from a similar property of the Lévy process \( X \). The following property is in fact true for a Lévy process on an arbitrary Lie group.
Proposition 1.2.16 ([Lia04, Lemma 3.5]) Let $X = (X_t)_{t \geq 0}$ be a Lévy process issued from the unit element on $G$. There exists a constant $K$ such that

\[ \forall t \geq 0, \quad \mathbb{E}[d_G(1, X_t)] \leq K \sqrt{t}. \]

In order to use this regularity property of $DF^X_{M,\text{vol}}$, we are going to consider the group $\Gamma$ of all random variables with values in $G$ defined on the probability space $(\mathcal{M}(A(M), G), DF^X_{M,\text{vol}})$. This group is endowed with a distance $\delta$ defined by the formula

\[ \forall S, T \in \Gamma, \quad \delta(S, T) = \mathbb{E}[d_G(S, T)]. \]

The pair $(\Gamma, \delta)$ is a complete metric topological group on which the properties of invariance of $d_G$ imply that the translations and the inversion are isometries.

We are going to consider the canonical process on $(\mathcal{M}(A(M), G), DF^X_{M,\text{vol}})$ as a multiplicative function from $A(M)$ into $\Gamma$: to each path $c$ is associated the random variable $H_c$ defined by $H_c(h) = h(c)$. The regularity property stated in Proposition 1.2.15 can be rewritten as

\[ \delta(1, H_c) \leq K \sqrt{\text{vol}(D)}. \]

We apply now the following result.

Theorem 1.2.17 ([L7, Theorem 3.3.1]) Let $(M, \gamma)$ be a compact Riemannian surface. Let vol denote the Riemannian volume of $\gamma$. Let $(\Gamma, \delta)$ be a complete metric group on which translations and inversion are isometries. Let $H \in \mathcal{M}(A(M), \Gamma)$ be a multiplicative function. Assume that there exists $K > 0$ such that for all simple loop $l \in A_\gamma(M)$ bounding a disk $D$ and such that $\ell(l) \leq K^{-1}$, the inequality $\delta(1, H(l)) \leq K \sqrt{\text{vol}(D)}$ holds.

Then $H$ admits a unique extension to an element of $\mathcal{M}(P(M), G)$, also denoted by $H$, such that if a sequence $(c_n)_{n \geq 0}$ of paths converges with fixed endpoints to a path $c$, then $H(c_n) \xrightarrow{n \to \infty} H(c)$.

The multiplicative function $H : P(M) \to \Gamma$ determines, following backwards what we have done a few paragraphs above, a probability measure on $\mathcal{M}(P(M), G)$, which we denote by $HF^X_{M,\text{vol}}$. The following theorem, the proof of which is unfortunately not very pleasant, guarantees that we have reached our goal.

Theorem 1.2.18 ([L7, Proposition 3.4.1]) The measure $HF^X_{M,\text{vol}}$ does not depend on the choice of the Riemannian metric on $M$. Moreover for all graph $G$ on $M$, whose edges need not be piecewise geodesic for any Riemannian metric, the marginal distribution of $HF^X_{M,\text{vol}}$ associated to $P(G)$ is the measure $DF^X_{M,\text{vol}}$.

Let us give an application of Theorem 1.2.17 in a simpler and more familiar setting. Let us work in the plane $\mathbb{R}^2$. Recall that we have defined, for all continuous loop $l : [0, 1] \to \mathbb{R}^2$, the function $n_l$ outside the image of $l$ (see (1.1)). Consider the usual Euclidian metric on $\mathbb{R}^2$ and denote by $A(\mathbb{R}^2)$ the set of piecewise affine paths. If the loop $l$ is piecewise affine, then the complement of its range has finitely many connected components.
and the function \( n_1 \), locally constant, takes only a finite number of values. It is in particular bounded and, since it has compact support, square-integrable.

To each piecewise affine path \( c \), let us associate the loop obtained by concatenating to \( c \) the two segments which join the origin of \( \mathbb{R}^2 \) to each of its endpoints and let us use the same notation \( n_c \) for the function associated to this loop. With this notation and considering the discussion above, one checks immediately that the mapping \( c \mapsto n_c \) is a multiplicative (or rather, additive) function

\[
 n : A(\mathbb{R}^2) \to L^2(\mathbb{R}^2).
\]

If \( l \) is a simple loop which bounds a disk \( D \), we have the equality \( \| n_l \|_{L^2} = \sqrt{\text{area}(D)} \).

In order to really be in the setting of Theorem 1.2.17, we should restrict ourselves to a compact disk of the plane, but we take this disk large enough to simply ignore that it is not the whole plane. We can thus claim that the function \( c \mapsto n_c \) extends to a continuous function from \( P(\mathbb{R}^2) \) to \( L^2(\mathbb{R}^2) \).

In particular, we deduce that for all rectifiable loop \( l \), the function \( n_l \) is square-integrable. Moreover, it depends continuously (in the \( L^2 \) sense) on the loop \( l \) (in the sense of the convergence for the distance \( d_l \) or the convergence in \( 1 \)-variation).

The fact that \( n_l \) is square-integrable for all rectifiable loop is known, if not well-known, and it is quantified by the Banchoff-Pohl inequality, which generalises the usual isoperimetric inequality by stating the following:

\[
\| n_l \|_{L^2}^2 \leq \frac{1}{4\pi} \ell(l)^2. \tag{1.7}
\]

1.3 Markovian holonomy fields

1.3.1 Definition

The construction which leads to the definition of the measure \( \text{HF}^X_{M,\text{vol}} \) considers only one surface, with possibly a set of boundary conditions which are determined from the beginning. In order to go from this point of view to the definition of Markovian holonomy fields, we need only one more step, in which we are going to consider all possible surfaces at once, and relate them thanks to the operations of cutting and gluing. In order to give a precise definition, we need to give a more formal framework to boundary conditions and gluing operations.

When one glues together two surfaces along two components of their boundary, along which compatible boundary conditions had been specified, one gets a new surface on which stays, as a remnant of the operation of surgery just performed, a curve along which the field is constrained. This leads us to define marked surfaces.

**Definition 1.3.1** A marked surface is a pair \((M, \mathcal{C})\), where \( M \) is a surface and \( \mathcal{C} \) is a finite subset of pairwise disjoint smooth 1-dimensional submanifolds of the interior of \( M \). The elements of \( \mathcal{C} \) are oriented and we assume that each element of \( \mathcal{C} \) appears in \( \mathcal{C} \) with the two possible orientations.

We denote by \( \mathcal{B}(M) \) the set of connected components of the boundary of \( M \), each taken with the two possible orientations. We denote by \( \text{Conj}(G) \) the set of conjugacy classes of \( G \). Observe that the inverse of a conjugacy class is still a conjugacy class.
1.3. MARKOVIAN HOLONOMY FIELDS

Definition 1.3.2 By a set of \(G\)-constraints on \((M, C)\) we mean a mapping \(C : C \cup B(M) \to \text{Conj}(G)\) such that for all \(N \in C \cup B(M)\), and with the notation \(N^{-1}\) for the same curve oriented in the opposite way, one has \(C(N^{-1}) = C(N)^{-1}\).

Let \((M, C)\) be a marked surface and let \(l \in C\) be a mark. We denote by \(\text{Spl}_l(M)\) the surface, not necessarily connected, which one obtains by cutting \(M\) along \(l\). This surface is naturally endowed with a gluing map \(\psi : \text{Spl}_l(M) \to M\). The marks of \(M\) other than \(l\) (resp. a measure of area on \(M\), a set of \(G\)-constraints on \((M, C)\)) determine marks on \(\text{Spl}_l(M)\), denoted by \(\text{Spl}_l(C)\) (resp. a measure of area denoted by \(\text{Spl}_l(\text{vol})\) on \(\text{Spl}_l(M)\), a set of \(G\)-constraints \(\text{Spl}_l(C)\) on \((\text{Spl}_l(M), \text{Spl}_l(C))\).

Let \((M, C, C)\) be a marked surface endowed with a set of \(G\)-constraints, let \(l \in C\) be a mark and let \(x\) be an element of \(G\). We denote by \(C_{l \to x}\) the set of \(G\)-constraints which coincides with \(C\) on all marks except \(l\) and \(l^{-1}\) and such that \(C(l)\) is the conjugacy class of \(x\).

We can at last give the main definition.

Definition 1.3.3 (Markovian holonomy field [L7, Definition 3.1.2])

Let \(G\) be a compact Lie group. A \(G\)-valued two-dimensional Markovian holonomy field is the data, for each quadruplet \((M, \text{vol}, C, C)\) consisting of a marked surface endowed with a measure of area and a set of \(G\)-constraints, of a finite measure \(\text{HF}_{M, \text{vol}, \mathcal{C}, C}\) on \((\mathcal{M}(P(M), G), \mathcal{I})\) such that the following properties are satisfied.

A_1. For all \((M, \text{vol}, C, C)\), \(\text{HF}_{M, \text{vol}, \mathcal{C}, C}(\exists l \in C \cup B(M), \text{vol}(l) \notin C(l)) = 0\).

A_2. For all \((M, \text{vol}, C)\) and all event \(\Gamma \in \mathcal{I}\), the function \(C \mapsto \text{HF}_{M, \text{vol}, \mathcal{C}, C}(\Gamma)\) is a measurable function of \(C\).

A_3. For all \((M, \text{vol}, C, C)\) and all \(l \in C\),

\[
\text{HF}_{M, \text{vol}, \mathcal{C} \setminus \{l, l^{-1}\}, C|_{B(M) \cup C \setminus \{l, l^{-1}\}}} = \int_G \text{HF}_{M, \text{vol}, \mathcal{C}, C_{l \to x}} \, dx.
\]

A_4. If \(\psi : (M, \text{vol}, C, C) \to (M', \text{vol}', C', C')\) is a diffeomorphism which preserves all the structures which we consider, then the mapping from \(\mathcal{M}(P(M'), G)\) to \(\mathcal{M}(P(M), G)\) induced by \(\psi\) sends the measure \(\text{HF}_{M', \text{vol}', \mathcal{C}', C'}\) to the measure \(\text{HF}_{M, \text{vol}, \mathcal{C}, C}\).
Chapter 1. Markovian Holonomy Fields \[L2, L3, L7\]

A5. For all \((M_1, \text{vol}_1, \mathcal{C}_1, C_1)\) and \((M_2, \text{vol}_2, \mathcal{C}_2, C_2)\), one has the identity
\[
HF_{M_1 \cup M_2, \text{vol}_1 \cup \text{vol}_2, \mathcal{C}_1 \cup \mathcal{C}_2, C_1 \cup C_2} = HF_{M_1, \text{vol}_1, \mathcal{C}_1, C_1} \otimes HF_{M_2, \text{vol}_2, \mathcal{C}_2, C_2}.
\]

A6. For all \((M, \text{vol}, \mathcal{C}, C)\) and all \(l \in \mathcal{C}\), denoting by \(\psi : \text{Spl}_l(M) \to M\) the gluing map along \(l\), one has
\[
HF_{\text{Spl}_l(M), \text{Spl}_l(\text{vol}), \text{Spl}_l(\mathcal{C}), \text{Spl}_l(C)} = HF_{M, \text{vol}, \mathcal{C}, C} \circ \psi^{-1}.
\]

A7. For all \((M, \text{vol}, \emptyset, C)\) and for all \(l \in \mathcal{B}(M)\),
\[
\int_{G} HF_{M, \text{vol}, \emptyset, C \rightarrow x}(1) \, dx = 1.
\]

The axioms A1, A2 and A3 are concerned with boundary conditions: they grant the fact that the measures \(HF_{M, \text{vol}, \mathcal{C}, C}\), as the marks and the constraints vary, are indeed disintegrations of each other in the sense one expects them to be.

The axiom A4 ensures that the measure \(HF_{M, \text{vol}, \mathcal{C}, C}\) does not depend on more structure than what we have considered. Let us recall that a classical theorem of Moser asserts that if two diffeomorphic surfaces are endowed with measures of area of the same total area, then there exists a diffeomorphism of the one onto the other which preserves areas. This result can be refined in presence of marks (see Proposition 1.4.3 in [L7]).

The axiom A5 takes care of disjoint reunion of surfaces and allows one to restrict oneself to connected surfaces.

The axiom A6 is the Markov property itself. It encompasses all possible situations regarding connectedness and orientability of the surface obtained after cutting \(M\) along \(l\). We mention orientability here because it can happen that a non-orientable surface becomes orientable when it is cut along a curve, for example a Môbius band along an equator.

Finally, the equator A7 is a normalisation axiom. Without it, given a field \(HF\), we could for all real \(\alpha \neq 0\) define another field \(e^{\alpha \text{vol}(M)}HF_{M, \text{vol}, \mathcal{C}, C}\).

1.3.2 Existence and partial classification

We are now going to state our two main results which are a result of existence and a result of classification. The natural order to state and prove these results is to start by the classification result.

It is crucial to remember that the measures \(HF_{M, \text{vol}, \mathcal{C}, C}\) are finite measures rather than probability measures. The transition kernels, which we have already discussed at length, are nothing but the masses of these measures.

Definition 1.3.4 Let \(HF\) be a Markovian holonomy field. Let \(p, g\) be non-negative integers and \(t\) a positive real. Let \(x_1, \ldots, x_p\) be elements of \(G\). Let \(M\) be an oriented surface with reduced genus \(g\) and whose boundary has \(p\) connected components. Let \(\text{vol}\) be a measure of area on \(M\) with total area \(t\). Let finally \(C\) be the set of \(G\)-constraints along the boundary.
of $M$ which to the positively oriented and arbitrarily ordered components of the boundary of $M$ associates the conjugacy classes of $x_1, \ldots, x_p$. We define

$$Z^{+}_{p,g,t}(x_1, \ldots, x_p) = \text{HF}_{M, \text{vol}, \varnothing, C}(1).$$

Suppose now that $g \geq 1$. Let $M'$ be a non-orientable surface of reduced genus $g$ and whose boundary has $p$ connected components. Let $\text{vol}$ be a measure of area on $M$ with total area $t$. Let $C$ be a set of $G$-constraints along the boundary of $M$ which to the arbitrarily oriented and arbitrarily ordered components of the boundary of $M$ associates the conjugacy classes of $x_1, \ldots, x_p$. We define

$$Z^{-}_{p,g,t}(x_1, \ldots, x_p) = \text{HF}_{M', \text{vol}, \varnothing, C}(1).$$

The functions thus defined are called the partition functions of the holonomy field.

One proves that these functions satisfy the first part of Proposition 1.1.7: they are all determined by the functions associated to disks and three-holed spheres.

The axiom $A_7$ guarantees that each partition function is, with respect to each one of its arguments and for all value of the others, the density of a probability measure on $G$. In particular, the family of measures $(Z^{+}_{1,0,t}(x) \, dx)_{t>0}$ is a one-parameter family of probability measures on $G$. In order to study it, and more generally in order to study the holonomy fields which we have just defined, it is necessary to make a few regularity assumptions.

**Definition 1.3.5** Let $\text{HF}$ be a Markovian holonomy field. We say that $\text{HF}$ is regular if the following two conditions are satisfied.

1. For all $(M, \text{vol}, \mathcal{E}, C)$, and for all sequence $(c_n)_{n \geq 0}$ of paths on $M$ which converges with fixed endpoints to $c$, one has

$$\int_{\mathcal{M}(P(M), G)} d_G(h(c_n), h(c)) \, \text{HF}_{M, \text{vol}, \mathcal{E}, C}(dh) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

2. For all $(M, \text{vol}, \mathcal{E})$, the function $(t, C) \mapsto \text{HF}_{M, \text{vol}, \mathcal{E}, C}(1)$ is continuous.

For a regular holonomy field, the second part of Proposition 1.1.7 is true: all partition functions are determined by the $Z^{+}_{1,0,t}$. The next theorem clarifies their structure and classifies partially the random holonomy fields.

**Theorem 1.3.6** ([L7, Proposition 4.2.1]) Let $\text{HF}$ be a regular Markovian holonomy field. Then the probability measures $(Z^{+}_{1,0,t}(x) \, dx)_{t>0}$ on $G$ are the 1-dimensional marginal distributions of a conjugation invariant Lévy process issued from the unit element. This process is said to be associated to $\text{HF}$ and determines all its partition functions. It is admissible in the sense of Definition 1.2.4.

It is tempting to think that this Lévy process determines completely the holonomy field, but, except when $G$ is Abelian, this does not seem easy to prove. On the other hand, the construction which we have explained in the previous sections allows one to prove the next theorem.
Theorem 1.3.7 ([L7, Theorem 4.3.1]) Every admissible Lévy process on a compact Lie group is associated with a regular Markovian holonomy field.

This theorem is probably the most important in this whole chapter and the most significant result in my attempts to improve and generalise the construction of the Yang-Mills measure. Its proof occupies most of the monograph [L7].

1.3.3 The Yang-Mills field

Among all admissible Lévy processes on a connected compact Lie group, the Brownian motion plays a special role. The holonomy field constructed by applying Theorem 1.3.7 to the Brownian motion has a name: it is the Yang-Mills field. It is the first to have been constructed and it is the one which physicists use in quantum field theories. We will come back to this particular field in the next chapter where we will investigate, in the form of a large deviation principle, the links between the object that we have constructed and the expressions which physicists manipulate with such great success.

In this section, we give a simple description of this field, analogous to that of the Poisson process indexed by loops (see Section 1.1.1), when the group $G$ is Abelian, that is, since we also suppose it compact and connected, when $G = U(1)^n$ for some $n \geq 0$. This case has been the first to be understood and the first works devoted to the construction of the Yang-Mills field with more general groups have proceeded by reducing the problem to the Abelian case. For the sake of simplicity, we treat the case $n = 1$.

Let us work, to start with, in the plane $\mathbb{R}^2$. Let $W$ be a white noise on $\mathbb{R}^2$, that is, an isometry of $L^2(\mathbb{R}^2)$ into a linear space of centred Gaussian random variables. Equivalently, $W$ is the Gaussian process indexed by $L^2(\mathbb{R}^2)$ whose covariance is given by the $L^2$ scalar product. We will denote by $W(f)$ the Gaussian variable associated with a square-integrable function $f$ on $\mathbb{R}^2$.

By analogy with Definition 1.1.1, we associate to each loop $l$ the random variable

$$B_l = e^{iW(n_l)}. \quad (1.8)$$

This definition makes sense only if $n_l$ is square-integrable. We have discussed this point earlier and indicated that the Banchoff-Pohl inequality guarantees us that this is the case when $l$ has finite length (see (1.7) and the discussion which precedes it).

If we restrict ourselves to a disk $D$ contained in $\mathbb{R}^2$, the family of variables $(B_l)_{l \in L(D)}$ thus defined has the distribution of $(H_l)_{l \in L(D)}$ under the measure $\int_{U(1)} H F_{D, \text{vol}, \partial D} dH$, where $HF$ is the Yang-Mills field which we integrate here with respect to the boundary condition. According to the axiom A7, this measure is indeed a probability measure.

In order to put a specific boundary condition on the boundary of the disk $D$, let us introduce a variant of the function $n_l$ by setting

$$n^0_l = n_l - \frac{1}{\text{vol}(D)} \int_D n_l \text{dvol},$$

that is by subtracting its mean to $n_l$. If we define, for each rectifiable loop $l$ in $D$, $B^0_l = e^{iW(n^0_l)}$, then we obtain a process indexed by $L(D)$ which resembles the process $B$
and satisfies $B_{0D}^0 = 1$ almost surely. However, the distribution of $(B_{l}^0)_{t \in \mathbb{L}(D)}$ is not equal to that of the canonical process under the normalised measure $Z_{1,0,\text{vol}(D)}^+(1)^{-1} \mathcal{H}_{D,0,0,0D}$.

In order to see this, let us make again in this context the construction which gave rise to Proposition 1.1.3. Suppose to fix the ideas that 

\[ \mathbb{P}(T = 2k\pi) \propto e^{-2k^2 \pi^2}. \]  

A better definition consists then in taking a random variable $T$ whose distribution is that described above, independent of the white noise $W$, to set $\tilde{n}_t = \frac{1}{\text{vol}(D)} \int_D n_t \, d\text{vol}$, and then

\[ B_t^{(1)} = e^{i(W(n_t^0) + \tilde{n}_t T)}. \]

The process $B^1$ thus defined has the distribution of the canonical process under the measure $Z_{1,0,\text{vol}(D)}^+(1)^{-1} \mathcal{H}_{D,0,0,0D}$.

In order to define the field on more general surfaces, it is necessary to extend the definition of the function $n^0_t$ and the number $\tilde{n}_t$. Such an extension is possible, at the price of certain choices. Let $M$ be an oriented surface with reduced genus $g$. Let $N_1, \ldots, N_p$ denote the boundary components of $M$, positively oriented. Let $C_1, \ldots, C_g$ be a family of smooth oriented curves on $M$ (it suffices that they be rectifiable) which determine a basis of the first homology group of the surface obtained from $M$ by gluing a disk along each boundary component. Then the classes of $N_1, \ldots, N_p, C_1, \ldots, C_g$ generate $H^1(M, \mathbb{Z})$ with the relation $[N_1] + \ldots + [N_p] = 0$. Let now $l$ be a loop on $M$. Its homology class decomposes uniquely as

\[ [l] = \sum_{i=1}^g \zeta_i [C_i] + \sum_{j=1}^{p-1} \nu_j [N_j], \]

where the $\zeta_i$ and the $\nu_j$ are integers. The cycle

\[ l^1 = l - \sum_{i=1}^g \zeta_i C_i - \sum_{j=1}^{p-1} \nu_j N_j \]
is thus homologous to 0. There exists then a unique function on $M$ with zero mean, locally constant on the complement of the range of $l^\perp$ and which varies by 1 or $-1$ at each crossing of $l^\perp$, depending on the orientation of $M$ and $l^\perp$. This definition is of course only valid if $l^\perp$ is regular enough but we will not enter the detail of the definition. We will denote this function by $n_0^\perp l$. The values that it takes are all equal modulo 1 and we denote by $\bar{n}^\perp l$ the corresponding class of $\mathbb{R}/\mathbb{Z}$. One checks easily that this notation is consistent with the case of the disk studied above.

We can now define a process indexed by the loops on an arbitrary surface. We use the notation of the paragraph above.

**Definition 1.3.8** Let $(M, \text{vol}, \emptyset, C)$ be a surface endowed with a measure of area and boundary conditions. Let $x_1, \ldots, x_p$ denote the boundary conditions along $N_1, \ldots, N_p$. Set $x = x_1 \ldots x_p$. Let $W$ be a white noise on $(M, \text{vol})$, that is, an isometry from $L^2(M, \text{vol})$ into a real Gaussian space. Let $T$ be a random variable with values in $\{t \in \mathbb{R} : e^{it} = x\}$, independent of $W$ and with distribution $\forall t \in \mathbb{R}, e^{it} = x, \mathbb{P}(T = t) \propto e^{-\frac{t^2}{2\text{vol}(M)}}$.

Let $U_1, \ldots, U_g$ be uniform independent variables on $U(1)$, independent of $W$ and $T$. For all loop $l \in L(M)$, we set

$$B_l^{(x_1, \ldots, x_p)} = e^{iW(n_0^\perp l)} e^{i\bar{n}^\perp l T} U_1^{\zeta_1} \ldots U_g^{\zeta_g} x_1 \ldots x_p.$$  \hspace{1cm} (1.10)

**Theorem 1.3.9 ([L2, Theorem 3.2])** The process $B^{(x_1, \ldots, x_p)}$ thus defined has the same distribution as the canonical process under the normalised measure $Z_{p,g,\text{vol}(M)}(x_1, \ldots, x_p)^{-1}$ $\mathcal{H}_F_{M,\text{vol},\emptyset,C}$.

In the Abelian case, the structure of the holonomy field associated to the Brownian motion is thus that of a white noise to which a finite-dimensional alea is added. It is for instance possible to reconstruct the white noise starting from the random holonomy field. This indicates, incidentally, that the measure of area is entirely encoded in the holonomy field.

In the next chapter, we will discuss the interest and significance of the definition (1.10) when one substitutes in it to the variable $T$ a deterministic value $t$ such that $e^{it} = x$.

**1.3.4 Link with topological quantum field theories**

We have presented the definition of random holonomy fields (Definition 1.3.3) as a generalisation of the definition of usual Markov processes, indexed by intervals of time. In this section, we are going to emphasise another analogy, between this definition and that of a topological quantum field theory (abbreviated TQFT), due to M. Atiyah.

**Insert 7 – Topological quantum field theories [Ati88]**

Let $n \geq 0$ be an integer. There is a category, denoted by $\text{Cob}_n$, whose objects are the compact oriented smooth manifolds without boundary of dimension $n$ and whose morphisms are given by the cobordisms: if $N_1$ and $N_2$ are two oriented $n$-manifolds, then $\text{Hom}(N_1, N_2)$ is the set of oriented $(n + 1)$-manifolds the
boundary of which is endowed with an orientation-preserving diffeomorphism with the disjoint union of $N_1^*$
and $N_2$, where $N_1^*$ denotes the manifold obtained from $N_1$ by reversing the orientation. On the other
hand, there exists a perhaps much more familiar category, denoted by $\text{Vect}$, whose objects are complex linear spaces
and whose morphisms are linear mappings.

A $(n+1)$-dimensional TQFT is a covariant functor from $\text{Cob}_n$ into $\text{Vect}$. This functor is usually denoted
by $Z$, in reference to the fact that this definition was initially given as a general framework for the partition
functions of certain statistical or quantum field theories. Thus, by definition, $Z$ associates to each $n$-manifold
$N$ a linear space $Z(N)$ and to each $(n+1)$-manifold $M$ endowed with a diffeomorphism between its boundary
and the disjoint union $N_1^* \sqcup N_2$ of two $n$-manifolds, a linear map $Z(M) : Z(N_1) \to Z(N_2)$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cobordisms.png}
\caption{Two examples of cobordisms, in dimensions $n+1 = 1$ and $n+1 = 2$.}
\end{figure}

This functor is required to satisfy a certain number of natural properties.

1. To be multiplicative: for all $n$-manifolds $N_1$ and $N_2$, we insist that $Z(N_1 \sqcup N_2) = Z(N_1) \otimes Z(N_2)$
(the tensor product of linear spaces) and, for any two $(n+1)$-manifolds $M_1$ and $M_2$, that $Z(M_1 \sqcup M_2) =
Z(M_1) \otimes Z(M_2)$ (the tensor product of linear maps).

2. To be involutive: for all $n$-manifold $N$, we insist that $Z(N^*) = Z(N)^*$ (the dual linear space) and, for
all $(n+1)$-manifold, that $Z(M^*) = Z(M)^*$ (the adjoint linear map).

3. To behave well with respect to the composition of cobordisms: if $M_1$ realises a cobordism between $N_1$
and $N_2$, and $M_2$ a cobordism between $N_2$ and $N_3$, and if $M$ is formed by gluing $M_1$ and $M_2$ along $N_2$, then
we insist that $Z(M) = Z(M_2) \circ Z(M_1)$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cobordisms_composition.png}
\caption{Composition of two cobordisms in dimension 2.}
\end{figure}

This last axiom can be reformulated in a more general way by slightly shifting our point of view. If we
identify the space of linear maps from $Z(N_1)$ into $Z(N_2)$ with $Z(N_1)^* \otimes Z(N_2) = Z(N_1^* \sqcup N_2)$, we can say
that the linear map $Z(M)$ is an element of the linear space $Z(\partial M)$. In these terms, the axiom becomes the following.

3'. If the boundary of $M$ can be written as $N_1 \sqcup N_1^* \sqcup N_2$ and if $M'$ is formed by gluing the two copies of $N_1$, then we insist that $Z(M') = \kappa(Z(M))$, where $\kappa : Z(N_1)^* \otimes Z(N_1) \otimes Z(N_2) \to Z(N_2)$ is the natural contraction of the first two factors.

![Figure 1.15: Gluing two components of the boundary of a surface.](image)

4. Finally, we insist that $Z$ behaves well with respect to orientation-preserving diffeomorphisms of manifolds of dimension $n$ and $n + 1$, that is to say that to diffeomorphic $n$-manifolds it associates isomorphic linear spaces and to diffeomorphic cobordisms it associates linear maps conjugated by the linear isomorphisms associated to the diffeomorphisms induced on the boundaries.

If $\emptyset$ designates the empty $n$-manifold, then the multiplicativity condition implies for example that $Z(\emptyset) = Z(\emptyset) \otimes Z(\emptyset)$. Thus, $Z(\emptyset)$ is either the null linear space or $\mathbb{C}$. In order to avoid trivial situations, we will assume that $Z(\emptyset) = \mathbb{C}$. Thus, for all $(n + 1)$-manifold $M$ without boundary, $Z(M)$ is a linear map from $\mathbb{C}$ into itself, that is, a complex number.

If $N$ is an $n$-manifold, then $Z(N \times [0, 1])$ is an endomorphism of $Z(N)$. Since the cylinder $N \times [0, 1]$ is diffeomorphic to two copies of itself glued together along two copies of $N$, one has the identity

$$Z(N \times [0, 1]) = Z(N \times [0, 2]) = Z(N \times [0, 1]) \circ Z(N \times [0, 1]).$$

Thus, $Z(N \times [0, 1])$ is a projection in the linear space $Z(N)$ and, since a surface is not modified up to homeomorphism when one glues a cylinder along one of its boundary components, one loses no generality in assuming that it is the identity. Then, gluing the two ends of this cylinder and using the axiom 3', on finds that $Z(N \times S^1)$, where $S^1$ is the circle, is the natural contraction of the identity of $Z(N)$, that is, the trace of this linear map, which is nothing but the dimension of $Z(N)$.

In the context of Markovian holonomy fields, we should modify slightly the definition of a TQFT in order to incorporate in it the notion of volume, or area when $n + 1 = 2$. We just need to replace $(n + 1)$-manifolds by manifolds endowed with a notion of $(n + 1)$-volume, of which the only invariant under diffeomorphisms is, according to a theorem of Moser, the total volume. In doing this, we modify only the morphisms in the category $\text{Cob}_n$. The axioms 1 to 4 are unchanged. The object thus defined is called an area-dependent topological quantum field theory (ad-TQFT). For all $(n + 1)$-manifold $M$, we will denote by $Z_t(M)$ the element of $Z(\partial M)$ associated to $M$ when it is endowed with a measure of $(n + 1)$-volume of total area $t$.

Let us investigate a 1-dimensional ad-TQFT, which we denote by $Z$. There is only one connected 0-manifold, the point, which we denote by $\text{pt}$. It can have two orientations, denoted by $+$ and $-$. There is only one linear space involved, namely $Z(\text{pt}_+)$, which
we denote by $V$. Then, there are only two oriented connected 1-manifolds, the interval and the circle. It is convenient to represent intervals as intervals of $\mathbb{R}$ endowed with their natural length. Thus, for all $s < t$ reals, $Z([s, t])$ is an endomorphism of $V$, which depends only on the length of $[s, t]$ and which we denote by $P_{t-s}$. Moreover, the axiom 3 implies that the relation $Z([s, u]) = Z([t, u]) \circ Z([s, t])$ holds for all $s < t < u$, so that $(P_t)_{t>0}$ is a semigroup of endomorphisms of $V$. Finally, according to the general argument which we have presented above, we have for all $t > 0$ the relation $Z_t(S^1) = \text{Tr}(P_t)$.

Finally, a 1-dimensional ad-TQFT is the same thing as a semigroup of endomorphisms of a linear space. In contrast with the purely topological case, nothing here prevents this linear space from being infinite-dimensional. For instance, in a Hilbertian framework, it suffices that the semigroup should be of trace class.

A sufficiently regular classical Markov process, indexed by intervals of time, gives rise, through its transition kernels, to a 1-dimensional ad-TQFT. Consider for instance the Brownian motion $X_t$ on a compact Lie group $G$. For all $t > 0$, let $Q_t$ denote the density of the distribution of $X_t$ with respect to the Haar measure on $G$. Define $Z(pt)$ as the space $L^2(G)^G$ of square-integrable functions on $G$ which moreover are invariant by conjugation. Then, for all $s < t$, we can define $Z([s, t])$ as the mapping $(x, y) \mapsto Q_{t-s}(x^{-1}y)$, seen either as the integral kernel of a linear map from $L^2(G)^G$ into itself, or equivalently as an element of $L^2(G)^G \otimes L^2(G)^G$.

Our definition of the 2-dimensional Markovian holonomy fields contains the definition of an ad-TQFT, given by the partition functions of the field. In the 2-dimensional setting, there is also a unique connected 1-manifold, which is the circle $S^1$, so that there is also a unique linear space involved. We set $Z(S^1) = L^2(G)^G$. Note that this space is finite-dimensional when $G$ is finite. Then the partition functions $Z_{p,g,t}^\pm$ of a Markovian holonomy field, which are indeed symmetric, invariant by conjugation and square-integrable with respect to all their arguments, can thus be seen as a symmetric tensor of $(L^2(G)^G)^{\otimes p}$. Let us summarise this whole section in the following result.

**Proposition 1.3.10** Let $G$ be a compact Lie group. Let $(Z_{p,g,t}^\pm)_{p,g \geq 0, t > 0}$ be the partition functions of a Markovian holonomy field, associated to oriented surfaces. Set $Z(S^1) = L^2(G)^G$, the space of square-integrable functions on $G$ invariant by conjugation. For all surface $M$ with genus $g$, total area $t$ and such that $\partial M$ has $p$ connected components, set $Z(M) = Z_{p,g,t}^+$. Then the functor $Z$ is a 2-dimensional ad-TQFT.
Chapter 2

Holonomy fields, coverings and connections

In this second chapter, we are going to describe several geometrical structures which give rise to multiplicative functions on the set of loops on a surface, with values in a group. In two very different situations, we are going to prove that some of the processes constructed in the previous chapter can be related to random versions of these geometrical structures.

2.1 Finite groups and ramified coverings \([L7, \text{Chapter 5}]\)

In this section, we are going to explain, when the group \(G\) is finite, how the Markovian holonomy field associated by Theorem 1.3.7 to a random walk on \(G\) can be realised explicitly by random ramified coverings of the surface on which one is working. Presenting this result will also give us the opportunity to understand in a simple situation how random holonomy fields are related to notions of horizontal lift and parallel transport.

2.1.1 Fundamental group and coverings

We are going to start by looking for the simplest possible geometric structures which give rise to multiplicative functions on the set of loops on a surface, with values in an arbitrary group.

Let \(M\) be a surface, on which we fix a point \(m\). Each loop \(l \in L_m(M)\) has a homotopy class, which is an element of the fundamental group \(\pi_1(M, m)\). The mapping \(l \mapsto [l]\) which to each loop associates its homotopy class is a multiplicative function \(L_m(M) \to \pi_1(M, m)\), which is given by the bare topological structure of \(M\). We are going to reinterpret this mapping in a way which apparently is rather complicated but will prove helpful in understanding many more multiplicative functions.

Let \(\pi : \tilde{M} \to M\) be a universal covering of \(M\). Choose \(\tilde{m} \in \pi^{-1}(m)\). The choice of \(\tilde{m}\) determines an action of the group \(\pi_1(M, m)\) on \(\tilde{M}\), in the following way. Let \([l]\) be a homotopy class of loops based at \(m\), represented by a loop \(l\). Let \(\tilde{l}\) be the lift of \(l\) issued from \(\tilde{m}\) and let \(\tilde{n}\) be the final point of \(\tilde{l}\), which does not depend on the choice of \(l\) in \([l]\). There exists a unique automorphism of the covering \(\pi : \tilde{M} \to M\) (that is, a
homeomorphism \( h \) of \( \tilde{M} \) such that \( \pi \circ h = \pi \) which sends \( \tilde{m} \) on \( \tilde{n} \). This automorphism of covering is the one which we associate to the element \([l]\) of \( \pi_1(M) \).

We have thus a covering of \( M \) (here \( \pi : \tilde{M} \to M \)) and a group (here \( \pi_1(M, m) \)) which acts simply and transitively on the fibres of this covering, which means that for each pair of elements of the same fibre, there exists a unique element of the group which sends the first on the second. We have moreover a basepoint on \( M \) (here \( m \)) and an element of the covering above this basepoint (here, \( \tilde{m} \)). Given a loop \( l \) on \( M \) based at \( m \), we can lift this loop starting from \( \tilde{m} \), determine the final point \( \tilde{n} \) of this lift and associate to the loop \( l \) the unique element of the group which sends \( \tilde{m} \) on \( \tilde{n} \). Here, tautologically, this element is the homotopy class of \( l \).

In order to understand what we have gained with this obscure formulation, let us choose a group \( G \), for example a finite group, and a homomorphism of groups \( \rho : \pi_1(M, m) \to G \). Then we can form the multiplicative function \( l \mapsto \rho([l]) \) from \( \mathbb{L}_m(M) \) into \( G \). This multiplicative function can be interpreted just in the same way as in the last paragraph. After replacing \( G \) by one of its subgroups if necessary, we will assume that \( \rho \) is onto. The theory of coverings guarantees us that to the kernel of \( \rho \), which is a normal subgroup of \( \pi_1(M, m) \), is associated a covering \( E \to M \) whose automorphism group identifies with the quotient of \( \pi_1(M, m) \) by the kernel of \( \rho \), which in turns identifies with \( G \). A way of constructing \( E \) is to realise it as the fibred product \( \tilde{M} \times_{\pi_1(M, m)} G \), which is the quotient of the Cartesian product \( \tilde{M} \times G \) by the equivalence relation which for all \( \tilde{x} \in \tilde{M} \), all \( g \in G \) and all \( \gamma \in \pi_1(M, m) \), identifies \((\tilde{x}, g)\) and \((\tilde{x} \cdot \gamma, \rho(\gamma^{-1})g)\).

This way of constructing \( E \) shows that \( G \) acts on \( E \) on the right, simply and transitively on the fibres: it suffices to set \((\tilde{x}, g) \cdot g' = (\tilde{x}, gg')\). The covering \( E \) is thus a principal \( G \)-bundle.

**Definition 2.1.1** Let \( G \) be a finite group. By a principal \( G \)-bundle over \( M \) we mean a covering \( \pi : E \to M \) endowed with an action on the right of \( G \) on \( E \), which is free and transitive on the fibres.

The construction which we have already done twice is the following one.

**Proposition 2.1.2** Let \( G \) be a finite group. Let \( \pi : E \to M \) be a principal \( G \)-bundle. Let \( m \) be a point of \( M \). Let \( e \) be a point of \( E \) such that \( \pi(e) = m \). Let \( l \) be a loop on \( M \) based at \( m \). Let \( \tilde{l} \) be the horizontal lift of \( l \) issued from \( e \). The unique element of \( G \) which sends \( e \) on the final point of \( \tilde{l} \) is called the monodromy or holonomy of \( l \). The function from \( \mathbb{L}_m(M) \) to \( G \) which to a loop associates its holonomy is multiplicative.

This proposition is the fundamental example, and the simplest one, of the way in which we are going to interpret multiplicative functions in geometric terms. It is however not a very rich source of multiplicative functions. On one hand, the only groups \( G \) for which one can make a non-trivial construction are those which are isomorphic to quotients of the fundamental group of \( M \). On the other hand, the image of a loop by the multiplicative functions which we have constructed so far depends only on its homotopy class. If \( M \) is simply connected, for example a sphere or a disk, we can only construct the multiplicative function which is identically equal to the unit element. If the fundamental group of \( M \) is
Abelian, for example if $M$ is a torus or a projective plane, then $G$ must be Abelian. In the case of the projective plane, it can actually only be $\mathbb{Z}/2\mathbb{Z}$.

2.1.2 Random ramified coverings

In order to enrich this construction, we are going to allow ourselves to puncture $M$, that is, to remove a finite set of points from it. This will solve our problems, since on one hand the fundamental group of a surface of which one has removed a finite non-zero number of points is always a free group, and on the other hand removing points will allow us to distinguish between loops which were originally homotopic in $M$. The drawback is of course that we cannot consider anymore those loops which cross the finite set which we have removed. However, this set is meant to become a random subset of $M$, and the probability that a given loop meets this set will be zero.

**Definition 2.1.3** A **ramified principal $G$-bundle on $M$ with ramification locus $Y$**, where $Y$ is a finite subset of the interior of $M$, is a covering $\pi : R \to M \setminus Y$ endowed with an action of $G$ on the right on $R$, which is free and transitive on the fibres.

The structure of a ramified bundle in the neighbourhood of a ramification point is simple. If $y$ is a point of the ramification locus $Y$, there exists a neighbourhood $V$ of $y$ such that the restriction of $\pi$ to each connected component of $\pi^{-1}(V \setminus \{y\})$ is topologically conjugated to the mapping $z \mapsto z^n$ from $\mathbb{C}^*$ into itself, for a certain integer $n \geq 1$ (see Figure 2.1). In particular, it is possible to add a finite number of points to $R$ and to make it a smooth surface, in such a way that all the new points are sent on $Y$ by $\pi$. The mapping $\pi : R \to M$ thus obtained is called a ramified covering.

![Figure 2.1: The mapping $z \mapsto z^2$ is the simplest ramified covering of $\mathbb{C}$ by itself.](image-url)

The conjugacy class of the monodromy along a small circle around a ramification point $y$ does not depend on the way in which it is computed. It is called the local monodromy.
at \( y \) and we denote it by \( \mathcal{O}(R, y) \). We will use the name \textit{ramification point} only for those points where the local monodromy is not the unit element.

Let \( \pi : R \to M \) be a ramified bundle. Let \( N \) be a connected component of the boundary of \( M \). The restriction of \( \pi \) to \( \pi^{-1}(N) \) is a principal \( G \)-bundle over \( N \), without any singularity since we have assumed that all ramification points are contained in the interior of \( M \). The structure of this \( G \)-bundle is completely determined, up to isomorphism, by the monodromy along \( N \). In fact, again, only the conjugacy class of this monodromy is well-defined, unless one chooses a basepoint on \( N \) and a point in the fibre over this point.

We are going to sort the ramified \( G \)-bundles according to their ramification locus and their structure over the boundary. We consider in fact isomorphism classes of \( G \)-bundles, where an isomorphism between \( \pi : R \to M \) and \( \pi' : R' \to M \) is a diffeomorphism \( \psi : R \to R' \) which intertwines the actions of \( G \) on \( R \) and \( R' \) and which is such that \( \pi' \circ \psi = \pi \).

**Definition 2.1.4** Let \( M \) be a surface. Let \( C : \mathcal{B}(M) \to \text{Conj}(G) \) be a set of \( G \)-constraints on the boundary of \( M \) (see Definition 1.3.2). Let \( Y \) be a finite subset of the interior of \( M \). We denote by \( \mathcal{R}(M, Y, C) \) the set of isomorphism classes of ramified \( G \)-bundles over \( M \) with ramification locus \( Y \) and boundary structure given by \( C \). We denote by \( \mathcal{R}(M, C) \) the union of the sets \( \mathcal{R}(M, Y, C) \) as \( Y \) spans the set of finite subsets of the interior of \( M \).

We will denote by \( O_1, \ldots, O_p \) the \( p \) conjugacy classes of \( G \) associated by \( C \) to the \( p \) connected components of the boundary of \( M \).

We are now going to define a probability measure on \( \mathcal{R}(M, C) \). Let us start by choosing an admissible Lévy process \( X \) on \( G \) (see Definition 1.2.4), characterised by its jump measure \( \Pi \), which is a measure on \( G \setminus \{1\} \). We denote by \( \Pi_1 \) the probability measure \( \Pi/\Pi(G) \). The definition of the measure is done in three steps.

1. Let \( R \) be a ramified \( G \)-bundle with ramification locus \( Y \). We call weight of \( R \) the following non-negative real:

   \[
   \Pi(R) = \prod_{y \in Y} \frac{\Pi_1(\mathcal{O}(R, y))}{\#\mathcal{O}(R, y)}.
   \]

   It is the product over all ramification points of the mass attributed by the probability measure \( \Pi_1 \) to an arbitrary element of the conjugacy class which is the local monodromy at this point.

2. One defines a finite measure \( \mathbb{R}B^X_{M,Y,C} \) on \( \mathcal{R}(M, Y, C) \) by setting

   \[
   \mathbb{R}B^X_{M,Y,C} = \frac{\#G^{1-g}}{\#O_1 \cdots \#O_p} \sum_{R \in \mathcal{R}(M, Y, C)} \frac{\Pi_1(R)}{\#\text{Aut}(R)} \delta_R.
   \]

   It is essentially the uniform measure on the finite set \( \mathcal{R}(M, Y, C) \), with a weighting determined by the monodromy and our choice of \( X \), and the fact that, as often in combinatorics, objects with many symmetries count less than objects with few symmetries.
3. Let $F(M)$ denote the set of finite subsets of $M$, endowed with a natural topology. We consider on $F(M)$ the probability measure $\Xi$ which is the distribution of a Poisson point process with intensity $\text{vol}$. We define a finite measure $\text{RB}^X_{M,\text{vol},C}$ on $\mathcal{R}(M, C)$ by

$$\text{RB}^X_{M,\text{vol},C} = \int_{F(M)} \text{RB}^X_{M,Y,C} \Xi(dY).$$

For some subsets $Y \subset M$, the set $\mathcal{R}(M,Y,C)$ can be empty. It is for example the case if $M$ has no boundary and $G = \mathbb{Z}/2\mathbb{Z}$: if $Y$ has an odd cardinal, then $\mathcal{R}(M,Y,C)$ is empty, whereas if $Y$ has even cardinal, $\mathcal{R}(Y,M,C)$ contains a unique element which admits exactly two automorphisms: the identity and the exchange of the two sheets. In this case, the ramification locus of a covering picked under the measure $\text{RB}^X_{M,\text{vol},C}$ is not exactly a Poisson point process, but rather a Poisson process conditioned to have an even number of points. In all cases, the distribution of the ramification locus of this random covering is absolutely continuous with respect to $\Xi$.

Let $l_1, \ldots, l_n$ be loops on $M$ based at the same point $m$. With $\Xi$-probability 1, the set $Y$ meets the range of none of these loops. Let $f: G^n \to \mathbb{R}$ be a continuous function invariant by the diagonal action of $G$ by conjugation. If we pick a ramified $G$-bundle at random under the measure $\text{RB}^X_{M,\text{vol},C}$ (normalised) and if we choose a reference point in the fibre above $m$, we can compute the monodromies $h(l_1), \ldots, h(l_n)$ of the loops $l_1, \ldots, l_n$. If we change the reference point, each monodromy is conjugated by the same element of $G$. Thus, $f(h(l_1), \ldots, h(l_n))$ is well-defined, independently of the choice of a reference point. Considering the definition of the $\sigma$-field $\mathcal{I}$ (Definition 1.2.8), this makes it plausible that the measure $\text{MF}^X_{M,\text{vol},C}$ induces a finite measure on $(\mathcal{M}(P(M), G), \mathcal{I})$, which we denote by $\text{MF}^X_{M,\text{vol},C}$. The main result is then the following.

**Theorem 2.1.5 ([L7, Theorem 5.4.2])** The finite measures $\text{HF}^X_{M,\text{vol},\varnothing,C}$ and $\text{MF}^X_{M,\text{vol},C}$, defined on the space $(\mathcal{M}(P(M), G), \mathcal{I})$, are equal.

We have thus reached our goal and realised the Markovian holonomy field associated to a random walk on a finite group as the field of random monodromy of a random ramified covering of a surface.

## 2.2 Brownian holonomy field and random connections [L4, LN5]

We are now going to interpret certain holonomy fields with values in continuous groups in terms of analogous structures to those which we have just described. Unfortunately, the geometrical objects are somewhat more complicated and there is no such satisfactory result as Theorem 2.1.5 in this case. We are going to focus on the field associated to the Brownian motion on a compact connected Lie group. For this field, we are going to write a large deviation principle which relates the process which we have constructed to an energy functional of geometric origin, namely the Yang-Mills action.
2.2.1 A random differential 1-form

Let us start by examining the case where the group \( G \) is Abelian. Let \( M \) be a surface. When \( G \) is Abelian, say \( G = U(1) \), there is a natural function to produce multiplicative functions of loops, by choosing a differential 1-form \( \alpha \) on \( M \) and by associating to each loop \( l \) the element \( \exp i \int_{l} \alpha \) of \( U(1) \). We are going to prove that the Brownian motion on \( U(1) \) indexed by loops on a disk, which we have described in Section 1.3.3 and defined by the equation (1.8), can be understood heuristically as the exponential of the integral of a random differential 1-form.

Let us work on the plane or on a disk. Let us consider a loop \( l \). Disregarding any consideration of rigour, we treat the white noise \( W \) as a random function, so that it is for instance possible to write \( W(n_{l}) \) = \( \int_{\mathbb{R}^{2}} n_{l}(x)W(x) \ dx \) (see Sections 1.1.1 and 1.3.3). We want to write this last integral as the integral of a 1-form along \( l \). To start with, let us define, for all \( x = (x_{1}, x_{2}) \in \mathbb{R}^{2} \), the 1-form \( d\theta_{x} \) on \( \mathbb{R}^{2} \setminus \{x\} \) by setting

\[
\forall y = (y_{1}, y_{2}) \in \mathbb{R}^{2}, \quad d\theta_{x}(y) = \frac{(y_{2} - x_{2})dy_{1} - (y_{1} - x_{1})dy_{2}}{(y_{1} - x_{1})^{2} + (y_{2} - x_{2})^{2}}.
\]

The index of \( l \) with respect to \( x \) can be evaluated thanks to the form \( d\theta_{x} \), by the formula

\[
n_{l}(x) = \frac{1}{2\pi} \int_{l} d\theta_{x}.
\]

Let us introduce the function \( G(x, y) = -\frac{1}{2\pi} \log d(x, y) \), where \( d \) is the Euclidian distance. This function is defined outside the diagonal on \( \mathbb{R}^{2} \times \mathbb{R}^{2} \) it is the kernel of the inverse of the Laplace operator. This means that if, for each compactly supported smooth function \( \psi \) we set \( G\psi(x) = \int_{\mathbb{R}^{2}} G(x, y)\psi(y) \ dy \), then we have, for all compactly supported smooth function \( \varphi : \mathbb{R}^{2} \rightarrow \mathbb{R} \), the equality \( G\Delta \varphi = \varphi \).

For all \( x \), let us denote by \( G_{x} \) the function \( G(x, \cdot) \) on \( \mathbb{R}^{2} \setminus x \). The differential of this function \( G_{x} \) resembles closely \( d\theta_{x} \). In fact, we have

\[
\frac{1}{2\pi} d\theta_{x} = *dG_{x},
\]

where \( * \) is the Hodge operator defined by \( *(a \ dy_{1} + b \ dy_{2}) = -b \ dy_{1} + a \ dy_{2} \), so that differential 1-form \( \alpha \) and all vector \( X \), one has \( *(\alpha)(X) = \alpha(J^{-1}X) \), where \( J \) denotes the rotation of angle \( \frac{\pi}{2} \).

Combining formally these observations, we find

\[
W(n_{l}) = \int_{\mathbb{R}^{2}} n_{l}(x)W(x) \ dx = \int_{\mathbb{R}^{2}} \int_{l} *dG_{x}W(x) \ dx = \int_{l} *d_{y} \int_{\mathbb{R}^{2}} G(x, y)W(x) \ dx = \int_{l} *d(GW).
\]
We have thus heuristically defined a random 1-form by taking a white noise, applying to it the inverse of the Laplace operator, taking the total differential of the function thus obtained, and applying the Hodge operator. The exponential of the integral of this 1-form gives the Brownian motion indexed by loops. Since we integrate it along loops only, we have the freedom to modify this 1-form by arbitrarily adding to it an exact 1-form. This corresponds to the invariance of the observable quantities under the action of the gauge group. We are now going to describe the appropriate geometric object which, without the assumption that $G$ is Abelian, allows one to define multiplicative functions of loops.

2.2.2 Connections

Let us start by defining the object which, in the case of a continuous group, plays the role of a ramified covering.

**Definition 2.2.1** Let $M$ be a surface. Let $G$ be a compact Lie group. A principal $G$-bundle on $M$ is a smooth manifold $P$ endowed with a smooth mapping $\pi : P \to M$ and a free action of $G$, denoted by $(p, g) \mapsto pg$, such that $\pi$ induces a diffeomorphism from the quotient $P/G$ onto $M$.

So, $P$ is the disjoint union of the fibres $\pi^{-1}(m)$ where $m$ runs over $M$ and, for each $m$, the group $G$ acts freely and transitively on $\pi^{-1}(m)$. This definition is very similar to the definition which we have given in the finite case. However, in the finite case, the total space of the bundle, which we have denote by $E$ and $R$, was a surface itself, whereas in the present case, it is a manifold whose dimension is the sum of that of $M$ and that of $G$.

**Definition 2.2.2** Let $\pi : P \to M$ and $\pi' : P' \to M$ be two principal $G$-bundles over $M$. An isomorphism of $P$ on $P'$ is a diffeomorphism $\varphi : P \to P'$ such that $\pi' \circ \varphi = \pi$ and which commutes to the actions of $G$ on $P$ and $P'$, that is, such that for all $p \in P$ and all $g \in G$, one has $\varphi(p \cdot g) = \varphi(p) \cdot g$.

The group of automorphisms of a bundle $\pi : P \to M$ is called the gauge group of this bundle. We denote it by $\mathcal{J}(P)$, or simply by $\mathcal{J}$.

The fundamental example of a principal $G$-bundle over $M$ is the Cartesian product $M \times G$, on which $G$ acts according to the formula $(m, h) \cdot g = (m, hg)$. The gauge group of $M \times G$, instead of being a finite group as in the case of ramified bundles, is now a huge group, which identifies naturally with the group of all smooth mappings from $M$ to $G$. If $j : M \to G$ is such a mapping, then it acts on $P$ according to $j \cdot (m, g) = (m, j(m)g)$.

The local structure of a principal bundle is always that of a Cartesian product: the restriction of $\pi : P \to M$ to $\pi^{-1}(U)$, where $U$ is a small open subset of $M$, is always isomorphic to the trivial bundle $U \times G \to U$. Nevertheless, there exist bundles whose global structure is not that of a Cartesian product, as for example the covering of $\mathbb{C}^*$ by itself by the mapping $z \mapsto z^2$, which is a $\mathbb{Z}/2\mathbb{Z}$-principal bundle, but not isomorphic to $\mathbb{C}^* \times \mathbb{Z}/2\mathbb{Z} \to \mathbb{C}^*$.

A fundamental difference between the discrete and continuous cases is that, in a bundle whose structure group is continuous, there is no canonical way of lifting a path in $M$ to
a path in $P$. A choice is necessary, namely that of a connection on $P$. A connection allows one to lift a tangent vector to $M$ to a tangent vector to $P$. By solving differential equations, this allows one to lift piecewise differentiable paths.

More precisely, a connection defines (see Figure 2.2), for all $m \in M$ and all $p \in \pi^{-1}(m)$, an injective linear mapping $r^p_m : T_m M \to T_p P$, which on one hand is a horizontal lift\(^1\), in the sense that, when it is composed with the differential of $\pi$, it gives the identity: $T_p \pi \circ r^p_m = \text{id}_{T_m M}$; and which on the other hand behaves well with respect to the action of $G$: for all $g \in G$, denoting by $R_g$ the action of $g$ on the right on $P$ and $T_p R_g$ its differential at $p$, we have $r^{R_g(p)}_m = T_p R_g \circ r^p_m$.

Figure 2.2: This picture illustrates the compatibility relations between the mappings of horizontal lift $r^p_m$, the projection $\pi$ and the action of $G$ on $P$. The two shaded subspaces of $T_p P$ and $T_{pg} P$ are the horizontal spaces at $p$ and $pg$.

\(^1\)The French word for horizontal lift is *relèvement horizontal*, hence the choice of the letter $r$. 
In order to describe a connection, one generally describes the range of the mapping $r_m^p$ at each point of $P$. It is a subspace of the dimension of $M$ in the tangent space to $P$ at $p$. This subspace admits a canonical supplementary subspace, which is the tangent space to the fibre through $p$. We denote this subspace, called the \textit{vertical} subspace, by $T^V_P \subset T_pP$.

It is canonically identified with the Lie algebra $\mathfrak{g}$ of $G$ by the differential at the unit element of the action of $G$: each vector of $T^V_P$ can be uniquely written as $\frac{d}{dt}|_{t=0}p \cdot e^{tA}$ for some $A$ in $\mathfrak{g}$. One describes a connection by specifying, at each point $p \in P$, the projection on $T^V_P$ along the horizontal subspace. With the identification $T^V_P \cong \mathfrak{g}$, this projection becomes a linear mapping $T_pP \to \mathfrak{g}$, whose kernel is the horizontal subspace.

The technical definition is the following\footnote{For more details on principal bundles and connections, one can consult the first volume of the book by Kobayashi and Nomizu \cite{KN96} and the book by Morita \cite{Mor01}.}

\textbf{Definition 2.2.3} Let $\pi : P \to M$ be a principal $G$-bundle over $M$. A connection on $P$ is a differential 1-form $\omega$ on $P$ with values in $\mathfrak{g}$ which satisfies the following properties.

1. For all $A \in \mathfrak{g}$, all $p \in P$, one has $\omega(\frac{d}{dt}|_{t=0}pe^{tA}) = A$.
2. For all $p \in P$, all $X \in T_pP$, all $g \in G$, one has $\omega(T_pR_g(X)) = \text{ad}(g^{-1})\omega(X) = g^{-1}\omega(X)g$.

We denote by $\mathcal{A}(P)$, or simply $\mathcal{A}$, the space of connections on $P$. The gauge group $\mathcal{J}(P)$ acts on the space $\mathcal{A}(P)$ of connections on $P$, the gauge transformation $j$ sending the connection $\omega$ to the connection $j^*\omega$ defined by $(j^*\omega)_p(X) = \omega_{j(p)}(T_pj(X))$. The space $\mathcal{A}(P)$ is an infinite dimensional affine space and $\mathcal{J}(P)$ acts by...
affine transformations. The quotient space $\mathcal{A}(P)/\mathcal{J}(P)$ is the configuration space of the physical theory which motivates the study of the Yang-Mills field.

Given a connection $\omega \in \mathcal{A}(P)$, a path $c : [0, 1] \to M$ with $c(0) = m$ and a point $p \in \pi^{-1}(m)$, there exists a unique path $\tilde{c} : [0, 1] \to P$ such that $\pi(\tilde{c}(t)) = c(t)$ for all $t \in [0, 1]$ and $\omega(\dot{\tilde{c}}(t)) = 0$ for all $t \in [0, 1]$. The path $\tilde{c}$ is called the horizontal lift of $c$ starting at $p$ (see Figure 2.4). If $c$ is a loop, that is, if $c(0) = c(1)$, the point $\tilde{c}(1)$ belongs to $\pi^{-1}(m)$ and can be written as $p \cdot h$ for a unique $h \in G$. This element $h$ is called the holonomy of $\omega$ along $c$, computed at $p$. We denote it by $\text{hol}_p(\omega, h)$. The invariance properties of the connection with respect to the action of $G$ on $P$ imply that for all $g \in G$, the horizontal lift of $c$ issued from $p \cdot g$ finishes at $\tilde{c}(1) \cdot g = (p \cdot h g) = (p \cdot g) \cdot g^{-1} h g$. Thus, we find the relation

$$\text{hol}_{p g}(\omega, c) = g^{-1} \text{hol}_p(\omega, c) g.$$ 

From this relation follows the fact that if $l_1$ and $l_2$ are two loops based at $m$ and if $l_1 l_2$ denotes their concatenation, then

$$\text{hol}_p(\omega, l_1 l_2) = \text{hol}_p(\omega, l_1) \text{hol}_{\text{hol}_p(\omega, l_1)}(\omega, l_2) = \text{hol}_p(\omega, l_2) \text{hol}_p(\omega, l_1).$$

This justifies the order which we have chosen in the definition of multiplicativity (Definition 1.2.1).

![Figure 2.4](image)

Figure 2.4: Each horizontal lift of a path is an integral curve of a vector field defined in the union of the fibres over the range of the path. This field is obtained by taking all horizontal lifts of the speeds of the path. If the path is a loop, it is possible to compare the finishing point of the horizontal lift with its initial point.

In the case where the group $G$ is finite, $g = \{0\}$, the unique connection on $P$ is the null connection and the notion of horizontal lift which we have just defined coincides with the notion of horizontal lift in a covering.

Given a principal bundle $\pi : P \to M$, a point $m \in M$ and a point $p \in \pi^{-1}(m)$, each connection on $P$ determines a multiplicative function from $L_m(M)$ to $G$, which to each
loop \( l \) associates \( \text{hol}_p(\omega, l) \). Changing the reference point \( p \) conjugates this application by an element of \( G \).

The action of the gauge group also modifies the mapping \( l \mapsto \text{hol}_p(\omega, l) \) by conjugating it by an element of \( G \) which does not depend on \( l \). Finally, there exists a mapping from the quotient \( \mathcal{A}(P)/\mathcal{J}(P) \) into the quotient \( \mathcal{M}(L_m(M), G)/G \).

**Proposition 2.2.4** Let \( \pi : P \rightarrow M \) be a principal \( G \)-bundle over \( M \). The mapping 

\[
\mathcal{A}(P)/\mathcal{J}(P) \longrightarrow \mathcal{M}(L_m(M), G)/G
\]

induced by holonomy is injective.

A random connection should thus determine a random holonomy field. This is precisely the point of view which has initially motivated the study of processes like those which we study here. Indeed, physical theories provide us, at a heuristic level, with a probability measure on the space of connections on a principal bundle, which we are now going to describe.

### 2.3 Yang-Mills action and measure

Let \( \pi : P \rightarrow M \) be a principal \( G \)-bundle over \( M \). Let \( \omega \) be a connection on \( P \). One associates to \( \omega \) a differential 2-form on \( P \), denoted by \( \Omega \) and defined by \( \Omega(X,Y) = d\omega(X,Y) + [\omega(X),\omega(Y)] \), where the bracket is the Lie bracket of \( g \). One proves that this 2-form vanishes as soon as one of its arguments is vertical. It suffices thus to understand it on horizontal vectors. One shows that \( \Omega(X,Y) = \omega([X,Y]^H) \), where \([X,Y]^H\) is the horizontal component of the vector \([X,Y]\). So, the curvature of \( \omega \) is identically zero if and only if the bracket of any two horizontal fields is still horizontal, that is, according to a theorem of Frobenius, if the horizontal distribution is integrable. In the general case, the curvature is a sort of infinitesimal holonomy along small loops. Let us assume that \( X \) and \( Y \) are the horizontal lifts of two fields \( X_M \) and \( Y_M \) on \( M \) such that \([X_M,Y_M] = 0\). Choose \( m \in M \) and \( p \in \pi^{-1}(m) \). If \( l_\varepsilon \) is a small square based at \( m \) with two opposite sides following the flow of \( X_M \) during a time \( \varepsilon \) and the two other opposite sides following the flow of \( Y_M \) during the time \( \varepsilon \), then the holonomy of \( \omega \) along \( l_\varepsilon \) computed at \( p \) is equal to \( \exp(\varepsilon^2\Omega_p(X,Y)) + o(\varepsilon^2) \).

We are now going to define the action of a connection \( \omega \) as a sort of squared \( L^2 \) norm of its curvature. The surface \( M \) is endowed with a surface measure denoted by \( \text{vol} \). For the sake of simplicity, let us assume that \( M \) is orientable so that \( \text{vol} \) can be identified with a volume 2-form on \( M \). Let then \( \omega \) be a connection. The curvature of \( \omega \), like any 2-form on \( P \) which vanishes as soon as one of its arguments is vertical, can be written as \( \Omega = K\pi^*\text{vol} \), where \( K : P \rightarrow g \) is a function. The curvature of a connection inherits from it a property of equivariance which guarantees us that for all \( p \in P \) and \( g \in G \), we have \( K(pg) = \text{ad}(g^{-1})K(p) = g^{-1}K(p)g \).

The Lie algebra of \( G \) is endowed with a scalar product invariant by conjugation, so the last equality implies that \( \|K(p)\| = \|K(pg)\| \). The function \( \|K\| \) is thus constant on
the fibres of $P$, it identifies with a function on $M$. We can then define the Yang-Mills action by setting

$$S(\omega) = \int_M \|K\|^2 \, d\text{vol}. $$

This definition coincides with the more classical definition $S(\omega) = \int_M (\Omega \wedge *\Omega)$, in the case of dimension 2. A fundamental property of $S$ is that it is invariant under the action of the gauge group: any two connections which differ by a gauge transformation have the same Yang-Mills action. In particular, this action induces a functional on the quotient space

$$S : \mathcal{A}(P)/\mathcal{J}(P) \to [0, +\infty). $$

The Yang-Mills measure, defined by physicists, is the following measure, defined on the space of all connections on a given principal bundle.

**Definition 2.3.1 (Heuristics)** Let $M$ be a surface endowed with a measure of area $\text{vol}$. Let $P$ be a principal $G$-bundle over $M$. The Yang-Mills measure on $(M, \text{vol}, P)$ is the probability measure $\text{YM}_{M, \text{vol}, P}$ on $\mathcal{A}(P)$ characterised by the expression

$$\text{YM}_{M, \text{vol}, P}(d\omega) = \frac{1}{Z_{M, \text{vol}, P}} e^{-\frac{1}{2}S(\omega)} \, D\omega. \quad (2.1)$$

The space $\mathcal{A}(P)$ of all connections on a given principal bundle $P$ is an infinite-dimensional affine space and $D\omega$ should be a translation invariant measure on this space. Since the space is not locally compact, such a measure can only be very pathological (it could for instance be the counting measure). The constant $Z_{M, \text{vol}, P}$ is a normalisation constant, which in this formulation should be infinite because of the action of the gauge group, which is not locally compact either and preserves the action $S$.

One could hope to solve these problems by working on the quotient $\mathcal{A}(P)/\mathcal{J}(P)$, of which one could hope that it is small enough, but this space is also non-locally compact. There is no direct way to give a meaning to the definition above.

Nevertheless, if one thinks of the Yang-Mills action as the square of the $L^2$ norm of the curvature, the expression (2.1) is analogous to the heuristic expression of a Gaussian measure in an infinite-dimensional space. So, one could interpret this expression by saying that the Yang-Mills measure is the measure under which the curvature of a connection has the distribution of a white noise on $(M, \text{vol})$ with values in the Lie algebra $\mathfrak{g}$. If one then remembers that this curvature is an infinitesimal version of the holonomy along small loops, it seems natural to take the holonomy field associated by Theorem 1.3.7 to the Brownian motion on $G$ as a candidate for a rigorous definition of the Yang-Mills measure.

### 2.4 A large deviations principle

#### 2.4.1 Sobolev spaces of connections

In order to give a rigorous meaning to the discussion of the preceding section, we are going to give a result which relates the Markovian holonomy field associated with the Brownian motion (which we call the Brownian holonomy field) and the Yang-Mills action. This
2.4. A LARGE DEVIATIONS PRINCIPLE

statement takes the form of a large deviation principle. Before entering into the details of the formulation of this principle, which still requires some preliminary explanations, we recall briefly the theorem which is the prototype of the result that we will state.

Let \( C_0^0([0, 1]) \) be the Banach space of continuous real-valued functions on \([0, 1]\), which vanish at 0, endowed with the uniform norm. The Wiener measure on this space can heuristically be described by the formula

\[
W(df) = \frac{1}{Z} \exp \left[ -\frac{1}{2} \int_0^1 \dot{f}(t)^2 \, dt \right] \, Df.
\]

In this formula a space of functions appears, which is the most natural space of functions for which \( \int_0^1 \dot{f}(t)^2 \, dt < +\infty \). This space is the Sobolev space \( W^{1, 2}([0, 1]) = H_0^1([0, 1]) \) of \( L^2 \) functions on \([0, 1]\) whose derivative in the distributional sense is a \( L^2 \) function and which are equal to 0 at 0. Let us define a functional \( S : C_0^0([0, 1]) \to [0, +\infty) \) by setting \( S(f) = \|f\|_{L^2([0, 1])}^2 \) if \( f \) belongs to \( H^1 \) and \( S(f) = +\infty \) otherwise. Schilder’s theorem is then the following.

**Theorem 2.4.1 (Schilder)** Let \( A \subset C_0^0([0, 1]) \) be a Borel subset. For all \( \lambda \in \mathbb{R} \), set \( \lambda A = \{\lambda f : f \in A\} \). Then

\[
\inf_{A^c} \frac{1}{2} \leq \lim_{T \to 0} \log W \left( \frac{1}{\sqrt{T}} A \right) \leq \lim_{T \to 0} \log W \left( \frac{1}{\sqrt{T}} A \right) \leq -\inf_{A} \frac{1}{2} S.
\]

This theorem expresses the fact that the family of distributions of \( \left((\sqrt{T}B_t)_{t \in [0, 1]}\right)_{T > 0} \), where \( B \) denotes the canonical process under \( W \), satisfy a large deviation principle with speed \( T \) and rate function \( \frac{1}{2} S \) as \( T \) tends to 0.

Let us recall the four main steps of the proof of this theorem. One first takes care of finite-dimensional marginals by applying fundamental results of large deviations like Cramér’s theorem. One then gathers the finite-dimensional results into one single large deviation principle on the space \( C_0^0([0, 1]) \) endowed with the topology of pointwise convergence, thanks to the Dawson-Gärtner theorem. At the end of this step, the rate function has the form

\[
\hat{S}(f) = \sup_{0 < t_0 < \ldots < t_n < 1} \frac{1}{2} \sum_{i=0}^{n-1} \frac{(f(t_{i+1}) - f(t_i))^2}{2(t_{i+1} - t_i)}.
\]

The third step consists in identifying the rate function with that which we have defined above. One checks immediately that \( \hat{S}(f) \leq S(f) \). It suffices thus to prove that \( \hat{S}(f) \leq S(f) \) implies \( f \in H_0^1 \) and \( S(f) \leq \hat{S}(f) \). In order to prove this point, one considers a sequence \( (g_n)_{n \geq 0} \) of piecewise affine interpolations of \( f \) corresponding to a sequence of subdivisions of \([0, 1]\) whose mesh tends to 0. Then one knows explicitly the square \( H^1 \) norm of each \( g_n \), it is smaller than \( 2\hat{S}(f) \) which by assumption is finite. The sequence \( (g_n)_{n \geq 0} \) is thus bounded in \( H^1 \) and one can extract from it a weakly convergent subsequence, to some function \( h \in H^1 \), whose square of the \( H^1 \) norm is smaller than \( 2\hat{S}(f) \). Since the injection \( H^1 \to C^0 \) is compact, this subsequence converges uniformly to \( h \), which must then be \( f \). Thus, \( f \) belongs to \( H^1 \) and \( S(f) \leq \hat{S}(f) \). This finishes the proof of the fact that \( \hat{S} = S \). Finally, the fourth and last step consists in going from the product topology on \( C^0 \) to the uniform topology.

We will comment on the step analogous to the third step above in the proof of the large deviation principle for the Brownian holonomy field. Let us simply emphasise that this step involves two main ingredients: the construction (in this case very simple) of piecewise affine approximations of \( f \) and a property of compactness.

In order to state a large deviation principle, we must define a rate function and for this we must identify the space of connections which plays the role of the space \( H^1 \) in
Schilder’s theorem. This space is, as we have said, the space of connections with finite energy. We need to identify a natural space of connections whose Yang-Mills action is finite. For this, we need to understand more concretely the local nature of this action.

In order to work locally in a principal bundle, one chooses an open subset $U$ of $M$, small enough that there exists a local section of $P$ above $U$, that is, a differentiable mapping $\sigma : U \to P$ such that $\pi \circ \sigma$ is the identity of $U$. One can then consider the 1-form $A = \sigma^* \omega$ on $U$ and the 2-form $F = \sigma^* \Omega$, both with values in $\mathfrak{g}$. The form $A$ writes, in local coordinates $(x, y)$, $A = A_x \, dx + A_y \, dy$, where $A_x$ and $A_y$ are mappings from $U$ to $\mathfrak{g}$. The function $[A_x, A_y]$, where the bracket is the Lie bracket of $\mathfrak{g}$, that is, the commutator of matrices, is still a mapping from $U$ to $\mathfrak{g}$ and the form $F$ writes

$$F = (\partial_x A_y - \partial_y A_x + [A_x, A_y]) \, dx \wedge dy.$$  

The contribution of the open subset $U$ to the action $S(\omega)$ is thus

$$\int_U \| \partial_x A_y - \partial_y A_x + [A_x, A_y] \|^2 \, \text{vol}(dx \, dy).$$

Recall that vol has a smooth positive density with respect to the Lebesgue measure in any chart. The presence of the quadratic term $[A_x, A_y]$ in this expression has the consequence that the space of connections with finite action is not defined in such a neat way as in the case of the Brownian motion. Indeed, the expression $[A_x, A_y]$ is meaningless if $A_x$ and $A_y$ are distributions, and there is no canonical space of distributional connections whose curvature is $L^2$. Nevertheless, the action of a connection is finite as soon as its components have $L^2$ derivatives and are themselves in $L^4$. Sobolev embedding theorems guarantee us that the space $H^1(M)$ of $L^2$ functions on $M$ with $L^2$ derivatives is contained in $L^p(M)$ for all $p \in [1, +\infty)$. Incidentally, if $M$ were 3-dimensional (resp. 4-dimensional), we would have $H^1(M) \subset L^6(M)$ (resp. $L^4(M)$), which would be sufficient for the action of a $H^1$ connection to be finite.

We will denote by $H^1A(P)$ the space of $H^1$ connections on $P$. The natural space of gauge transformations which acts on this space is $H^2J(P)$, the space of $H^2$ gauge transformations. We claim that the Yang-Mills actions is well defined and finite on the quotient

$$S : H^1A(P)/H^2J(P) \to \mathbb{R}_+.$$  

However, in contrast with the case of Schilder’s theorem, the functional which we have just defined cannot yet be the rate function of our large deviation principle, for it is not defined on the right space. The Brownian holonomy field is not a random connection, but rather a random holonomy. We need thus to define $S$ on the space of multiplicative functions of loops and for this, we need to make sure that an $H^1$ connection has a holonomy.

### 2.4.2 The rate functions

The problem is that an $H^1$ connection in two dimensions is far less regular than an $H^1$ function on an interval. In particular, an $H^1$ function on $[0, 1]^2$ is not necessarily continuous and cannot be evaluated at a particular point. In order to compute the
holonomy along a loop \( l : [0, 1] \to M \) of a connection locally given by the form \( A \), we need to be able to solve the differential equation
\[
\dot{h}_t = -h_t A(\dot{l}_t), \quad h_0 = 1,
\]
where \( h : [0, 1] \to G \) is the unknown function. The holonomy is then the value of \( h \) at 1.

It suffices thus that the function \( t \mapsto A(\dot{l}_t) \) be \( L^1 \), as one can check for instance by writing the solution as a convergent series of multiple integrals. Fortunately, along a curve which is a smooth submanifold of \( M \), an \( H^1 \) function admits a trace which is in the space \( H^{-\frac{1}{2}} \) of this curve, hence in particular in \( L^2 \), and in \( L^1 \). This remark would still be true in 3 dimensions, where the trace is in \( L^2 \), but not in 4 dimensions, where the trace of a \( H^1 \) connection is only in the space \( H^{-\frac{1}{2}} \).

Let us denote by \( P^\infty(M) \) the space of paths on \( M \) which are concatenations of smooth submanifolds. The following result summarises the discussion above and specifies in which sense an \( H^1 \) connection is characterised by its holonomy.

**Proposition 2.4.2** Let \( \pi : P \to M \) be a principal bundle. The mapping induced by the holonomy
\[
H^1 A(P)/H^2 J(P) \to \mathcal{M}(P^\infty(M), G)/G^M
\]
is injective.

Thanks to this result and (2.2), we can now define a functional on the space of multiplicative functions of loops. In order to avoid working with equivalence classes, we introduce the following notation.

**Definition 2.4.3** Let \( \omega \in H^1 A(P) \) be a connection on \( P \). Let \( h \in \mathcal{M}(P^\infty(M), G) \) be a multiplicative function. We say that \( \omega \) and \( h \) are related, and we write \( \omega \sim h \) if the image of the class of \( \omega \) by the injection determined by the holonomy is the class of \( h \).

The rate function of the large deviation principle is then the following.

**Definition 2.4.4** Let \( M \) be a surface without boundary endowed with a measure of area. Let \( P \) be a principal \( G \)-bundle over \( M \). We define a function \( I_P : \mathcal{M}(P^\infty(M), G) \to [0, +\infty] \) by setting
\[
\forall h \in \mathcal{M}(P^\infty(M), G), \quad I_P(h) = \begin{cases} \frac{1}{2} S(\omega) & \text{if there exists } \omega \in H^1 A(P) \text{ such that } \omega \sim f, \\ +\infty & \text{otherwise}. \end{cases}
\]

In the case where the surface has a boundary, we want to be able to take boundary conditions into account. Suppose \( M \) given and a set \( C \) of boundary conditions in the sense of Definition 1.3.2. We denote by \( H^1 A_C(P) \) the space of \( H^1 \) connections which satisfy these boundary conditions.

**Definition 2.4.5** Let \( M \) be a surface whose boundary is not empty, endowed with a measure of area and a set \( C \) of boundary conditions. Let \( P \) be a principal \( G \)-bundle over \( M \). We define a function \( I_{P,C} : \mathcal{M}(P^\infty(M), G) \to [0, +\infty] \) by setting
\[
\forall h \in \mathcal{M}(P^\infty(M), G), \quad I_{P,C}(h) = \begin{cases} \frac{1}{2} S(\omega) & \text{if there exists } \omega \in H^1 A_C(P) \text{ such that } \omega \sim f, \\ +\infty & \text{otherwise}. \end{cases}
\]
It is clear that the functions $I_P$ and $I_{PC}$ depend on the bundle $P$ only up to isomorphism. We can thus index these functionals by isomorphism classes of principal $G$-bundles. Let us explain how these bundles are classified. Since we are working with the Markovian field associated with the Brownian motion, and since the Brownian motion on a Lie group visits only the connected component of the unit element, we will assume that $G$ is connected.

### Insert 9 – Classification of principal $G$-bundles

Let $M$ be a surface and $\pi : P \to M$ a principal $G$-bundle with $G$ connected. Let us first prove quickly that $P$ is trivial, that is to say, isomorphic to $M \times G$, if and only if there exists a smooth mapping $\sigma : M \to P$ such that $\pi \circ \sigma$ is the identity of $M$. Such a mapping is called a global section of $P$. If $P$ is isomorphic to $M \times G$, let $\varphi : P \to M \times G$ be an isomorphism. Then $\sigma(m) = \varphi^{-1}((m,1))$ is a global section of $P$. Conversely, let $\sigma : M \to P$ be a global section of $P$. For all $p \in P$, the equality $\pi(\sigma(p)) = \pi(p)$ implies that there exists a unique element of $G$, which we denote by $\gamma(p)$, such that $p = \sigma(p) \cdot \gamma(p)$. Then the mapping $\varphi(p) = (\pi(p), \gamma(p))$ is an isomorphism from $P$ onto $M \times G$.

Let us assume first that the boundary of $M$ is not empty. Then the surface $M$ contains a graph, or a bunch of circles, on which it retracts by deformation. For purely topological reasons, any section of $P$ above this graph can be extended to a global section of $P$. On the other hand, since $G$ is connected, the restriction of $P$ over the graph admits a section. Finally, $P$ is necessarily trivial.

Let us now assume that $M$ is orientable and closed, that is, without boundary. In this case, non-trivial bundles over $M$ may exist. Let us consider a curve $C$ on $M$ which splits $M$ into two surfaces $M_1$ and $M_2$, for example the boundary of a small disk, or a more complicated separating curve. The restrictions of $P$ to $M_1$ and $M_2$ admit global sections, because $M_1$ and $M_2$ are surfaces with boundary. Let us choose two such sections, one on each side of $C$. The discrepancy between these two sections can be measured along $C$, with an appropriate convention on orientation, and determines a closed curve in $G$. By altering the sections which we have chosen in a small cylindrical neighbourhood of $C$, we can transform this curve in $G$ into any other curve homotopic to it. It follows easily that if the construction which we have done with $P$ produces with another bundle $P'$ a curve in $G$ homotopic to that obtained with $P$, then $P$ and $P'$ are isomorphic. Thus, there are at most as many equivalence classes of principal $G$-bundles as there are homotopy classes in $G$.

It is slightly less easy to see, but still true, that by changing the sections which we have chosen, we cannot change the homotopy class of the discrepancy between them. In order to justify this, one can say that any two sections over $M_1$ (resp. $M_2$) differ by a smooth mapping from $M_1$ (resp. $M_2$) to $G$. The point is that any such mapping sends the boundary of $M_1$ (resp. $M_2$) to a homotopically trivial curve in $G$. Indeed, the fundamental group of $G$ is Abelian, as that of any topological group. Thus, a mapping from $M_1$ to $G$ being given, the homotopy class of the image of $C$ in $G$ depends only on the image of the homotopy class of $C$ in the quotient $\pi_1(M_1)/[\pi_1(M_1), \pi_1(M_1)]$, which is isomorphic to $H_1(M_1, \mathbb{Z})$. On the other hand, $C$ is the boundary of $M_1$ and represents the null homology class (its homotopy class can in fact easily be written explicitly as a product of commutators).

Finally, the homotopy class in $G$ which we obtain by comparing two sections of the restrictions of $P$ on both sides of a separating curve of $M$ does not depend on the choice of these sections (nor of this curve, by a similar argument) and characterises $P$ up to isomorphism. There are thus exactly as many isomorphism classes of principal $G$-bundles over $M$ as there are elements in $\pi_1(G)$. If $G$ is semi-simple, which is equivalent to saying that its centre is trivial, then this fundamental group is finite. Of course, if $G$ is simply connected, then all $G$-bundles are trivial.

Let us also indicate, although we will not use this fact, that on a closed non-orientable surface, principal $G$-bundles are classify by the quotient of $\pi_1(G)$ by its subgroup formed by elements which are squares.

Let us finish this description by giving some examples of fundamental groups of connected Lie groups. For all $N \geq 1$, the fundamental group of $U(N)$ is isomorphic to $\mathbb{Z}$, the homotopy class of a loop $t \mapsto \gamma_t$ being characterised by the index of the loop $t \mapsto \det \gamma_t$ with respect to 0 in $\mathbb{C}$. On an orientable surface endowed with a Riemannian metric, the unitary tangent bundle is a $U(1)$-principal bundle whose isomorphism class is
2.4. A LARGE DEVIATIONS PRINCIPLE

given by the Euler characteristic of the surface. For all $N \geq 2$, the group SU($N$) is simply connected. All SU($N$)-bundles over surfaces are thus trivial. On the other hand, the centre of SU($N$) is formed by scalar matrices corresponding to $N$-th roots of unity and the fundamental group of the quotient SU($N$)/Z(SU($N$)) is isomorphic to $\mathbb{Z}/N\mathbb{Z}$. Since the fundamental group of $U(1)^d \times SU(n_1)/Z(SU(n_1)) \times \ldots \times SU(n_r)/Z(SU(n_r))$ is isomorphic to $\mathbb{Z}^d \times \mathbb{Z}/n_1\mathbb{Z} \times \ldots \times \mathbb{Z}/n_r\mathbb{Z}$, we see that every finitely generated Abelian group is isomorphic to the fundamental group of a connected Lie group. Finally, for $N \geq 3$, the fundamental group of SO($N$) is $\mathbb{Z}/2\mathbb{Z}$. There are thus exactly two non-isomorphic SO($N$)-principal bundles over a closed connected orientable surface.

Since all bundles over a surface with boundary are trivial (let us emphasise that this is no true for non-trivial finite groups, because we have assumed that the group was connected), the functional $I_{P,C}$ does not depend on $P$. We will simply denote it by $I_C$. In the case of an orientable surface without boundary, there is thus a functional $I_P$ for each element of $\pi_1(G)$. We denote by $I_z$ the functional associated with the element $z \in \pi_1(G)$.

One can moreover compute the homotopy class of $G$ associated to a bundle $P$ in terms of the holonomy determined by any connection on $P$. This means that the subsets of $\mathcal{M}(P^\infty(M), G)$ on which the functionals $I_z$ and $I_{z'}$ are finite are the same if $z = z'$ and disjoint otherwise.

2.4.3 The Brownian holonomy field and its disintegration

In trying to define a rate function, we have been led to defining many of them, corresponding to the various non-isomorphic principal $G$-bundles over a surface. On the other hand, the definition of the Brownian holonomy field, just as that of any other Markovian holonomy field, is independent of the choice of a particular $G$-bundle over $M$. It is thus not likely that the Brownian holonomy field such as we have defined it is going to be the object for which we will be able to state a large deviation principle.

It is in fact possible, and at this point necessary, to disintegrate the Brownian field according to the isomorphism classes of principal bundles. Before explaining this point, let us briefly describe the Brownian motion on $G$, in order to correctly define the field associated to it.

**INSERT 10 – Brownian motion on a compact connected Lie group**

Let $G$ be a connected compact Lie group. Let $\mathfrak{g}$ be its Lie algebra. Each element of $G$ can be identified with a first-order differential operator on $G$: for all $A \in \mathfrak{g}$, one defines the operator $\mathcal{L}_A$ by setting, for all differentiable function $f : G \to \mathbb{R}$ and all $g \in G$,

$$(\mathcal{L}_A f)(g) = \frac{d}{dt}|_{t=0} f(ge^{tA}).$$

Let us now assume that $\mathfrak{g}$ is endowed with a scalar product invariant by conjugation. Consider an orthonormal basis $(A_1, \ldots, A_d)$ of $\mathfrak{g}$. Then we define the Laplace operator on $G$, which is a second-order differential

---

3 This is also true over 3-manifolds. Indeed, $H_2(G, \mathbb{Z}) = 0$ for all connected Lie group, so that, by a classical result of Hurwitz, a simply connected Lie group has also a trivial second homotopy group: $\pi_2(G) = 0$. 
operator, by
\[ \Delta = \sum_{i=1}^{d} (L_{A_i})^2. \]

The Brownian motion on $G$ is the Markov process whose generator is the half of this Laplace operator, which is just the Laplace-Beltrami operator associated with the bi-invariant metric on $G$ determined by the invariant scalar product on $g$.

Let us give another description of this Brownian motion as the solution of a stochastic differential equation. We will assume that $G$ is a subgroup of the unitary group $U(N)$ and $g$, accordingly, is a linear subspace of $u(N)$, the space of skew-Hermitian matrices.

Let again $(A_1, \ldots, A_d)$ be an orthonormal basis of $g$. Let $(B_t)_{t \geq 0}$ be the Brownian motion in $g$, that is, the process whose components on $A_1, \ldots, A_d$ are independent standard real Brownian motions. Let us define $C = A_1^2 + \ldots + A_d^2$. This matrix, called the Casimir matrix, commutes to all elements of $G$. For example, if $G$ is $U(N)$ itself and if $u(N)$ is endowed with the scalar product $\langle A, B \rangle = \text{Tr}(A^* B)$, then $C = -NI_N$. We consider the following stochastic differential equation in $\mathbb{M}_N(C)$, written in Itô notation:

\[ dX_t = X_t \, dB_t + \frac{1}{2} CX_t \, dt, \quad X_0 = I_N. \]

One checks that the solution $(X_t)_{t \geq 0}$ of this equation stays in $G$ and it is called the Brownian motion in $G$.

Let us emphasise that this Brownian motion is not the exponential of the Brownian motion $B$ in $g$. It is a non-commutative exponential, also called $P$-exponential (for path-ordered). One can also say that $X$ is obtained by rolling $G$ on $g$ along $B$. Finally, for all $t \geq 0$, one has

\[ B_t = \lim_{n \to \infty} \exp X_{\frac{t}{n}} \exp(X_{\frac{2t}{n}} - X_{\frac{2t}{n}}) \ldots \exp(X_{\frac{(n-1)t}{n}} - X_{\frac{(n-1)t}{n}}). \]

Let $G$ be a connected compact Lie group endowed with a bi-invariant metric. Let $\text{YM}$ denote the holonomy field associated with the Brownian motion on $G$.

**Theorem 2.4.6** Let $(M, \text{vol})$ be a surface without boundary. There exists a random variable

\[ \phi : \mathcal{M}(\mathcal{P}(M), G) \to \pi_1(G), \]

defined under the measure $\text{YM}_{M, \text{vol}}$, which measures in a natural sense the isomorphism class of the fibre bundle underlying a given multiplicative function.

The disintegration of the measure $\text{YM}_{M, \text{vol}}$ with respect to $\phi$ determines, for each $z \in \pi_1(G)$, a finite measure $\text{YM}_{M, \text{vol}}^z$ on $\mathcal{M}(\mathcal{P}(M), G)$, in such a way that

1. $\text{YM}_{M, \text{vol}} = \sum_{z \in \pi_1(G)} \text{YM}_{M, \text{vol}}^z$,
2. for all $z \in \pi_1(G)$, on a $\text{YM}_{M, \text{vol}}^z(\phi \neq z) = 0$.

It is not difficult to describe the masses of the finite measures $\text{YM}_{M, \text{vol}}^z$, at least in the case where $M$ is the sphere $S^2$. In this case, the mass of $\text{YM}_{M, \text{vol}}$ is $Q_{\text{vol}}(1)$, that is, the density at the unit element of $G$ of the distribution at time $\text{vol}(M)$ of the Brownian motion on $G$ issued from the unit element. It is thus the natural mass of the Brownian bridge of length $\text{vol}(M)$ on $G$. If we lift this Brownian bridge to a universal covering of $G$, which we denote by $\tilde{G}$, we find a Brownian motion issued from the unit element and
conditioned to finish at time \( \text{vol}(M) \) in a point of the fibre over the unit element of \( G \). This fibre identifies canonically with \( \pi_1(G) \) and if we denote by \( \tilde{Q}_{\text{vol}(M)} \) the density at time \( \text{vol}(M) \) of the Brownian motion on \( \tilde{G} \), we can say that the mass of \( \text{YM}_{M,\text{vol}}^z \) is equal to \( \tilde{Q}_{\text{vol}(M)}(z) \).

The construction of the measures \( \text{YM}_{M,\text{vol}}^z \) is the object of the paper [?]. We use a procedure of discrete approximation analogous to that use in order to define Markovian holonomy fields. However, it is necessary to modify the definition of the discrete theory in order to take into account the fact that the restriction of a principal bundle to a graph on a surface is always trivial, so that this restriction ignores the topology of \( P \). One checks at the end of this constructions that there is a slightly stronger property of mutual singularity.

**Proposition 2.4.7** For all \( T, T' > 0 \) and all \( z, z' \in \pi_1(G) \), the measures \( \text{YM}_{M,T\text{vol}}^z \) and \( \text{YM}_{M,T'\text{vol}}^{z'} \) are mutually singular, unless \( (T, z) = (T', z') \).

The discussion of the universal covering of \( G \) above is very similar to the discussion in Section 1.3.3. Indeed, in the case without boundary, the random variable \( T \) which appeared there was indeed the same variable which we have called \( \alpha \). Thus, the disintegration given by Theorem 2.4.6 can be explicitly obtained, in the Abelian case, by replacing the variable \( T \) by the deterministic value corresponding to the topological type of the chosen bundle. In this Abelian case, where \( W \), the white noise, played the role of the curvature, one can in fact understand \( T \) as the total curvature of the bundle under consideration, which, according to Gauss-Bonnet’s theorem, is a topological invariant of the bundle.

### 2.4.4 The large deviation principle

We can now state a rigorous result which links the measures \( \text{YM}_{M,\text{vol}}^z \) and the Yang-Mills action. The following two theorems are proved in [LN5].

**Theorem 2.4.8** ([LN5, Theorem 4]) Let \( (M, \text{vol}) \) be a closed surface endowed with a measure of area. Let \( G \) be a connected compact Lie group endowed with a bi-invariant Riemanian metric. Let \( z \) be an element of \( \pi_1(G) \). Then the family of probability measures associated to the finite measures \( \text{YM}_{M,T\text{vol}}^z \) on the space \( (\mathcal{M}(P^\infty(M), G), \mathcal{C}) \) satisfies a large deviation principle with speed \( T \) and rate function \( I_z \).

**Theorem 2.4.9** ([LN5, Theorem 3]) Let \( (M, \text{vol}) \) be a surface whose boundary is non-empty, endowed with a measure of area and a set \( C \) of boundary conditions. Let \( G \) be a connected compact Lie group endowed with a bi-invariant Riemanian metric. Then the family of probability measures associated to the finite measures \( \text{YM}_{M,T\text{vol},z,C}^z \) on the space \( (\mathcal{M}(P^\infty(M), G), \mathcal{C}) \) satisfies a large deviation principle with speed \( T \) and rate function \( I_C \).

The first theorem means, by definition of a large deviation principle, that for all measurable subset \( A \subset \mathcal{M}(P^\infty(M, G)) \), one has the inequalities

\[
- \inf_{f \in A^c} I_z(f) \leq \lim_{T \to 0} T \log \text{YM}_{M,T\text{vol}}^z(A) \leq \lim_{T \to 0} T \log \text{YM}_{M,T\text{vol}}^z(A) \leq -\inf_{f \in A} I_z(f),
\]  

(2.3)
where \( A^\circ \) and \( \overline{A} \) denote respectively the interior and closure of \( A \) for the topology on \( \mathcal{M}(\mathbb{P}^\infty(M,G)) \) which is the trace of the product topology on \( G^\mathbb{P}^\infty(M) \). Besides, saying that \( I_z \) is a rate function implies that it is lower semi-continuous for the same topology. This topology is compact. The level sets of \( I_z \) are thus automatically compact: it is a good rate function, in the technical sense of large deviations. However, it is so by the coarseness of the ambient topology and this precision brings little extra information.

Note that we have stated (2.3) with finite measure rather than probability measures, but the effect of normalisation is harmless at the scale in which we are working, since \( \text{YM}^\text{vol}_M(1) = O(\frac{1}{\sqrt{T}}) \).

The proof of this theorem follows the main steps of the proof of Schilder’s theorem (see Insert 8 above), except that the fourth step is useless, since our results use the product topology on the space of multiplicative functions. The first step is achieved by using a classical result of S. Varadhan [Var67a, Var67b] which determines the large deviation functional for the heat kernel on \( G \). The second step uses the Dawson-Gärtner theorem [DZ93], with a complication due to the fact that all finite-dimensional marginals of holonomy fields are not explicitly known. It is thus necessary to proceed to approximations and to check that they are good enough in the exponential scale in which we work. The third step is much more difficult than in Schilder’s theorem. The first expression that one obtains for the rate function is the following:

\[
\hat{I}(h) = \sup_G \sum_{F \in \mathcal{F}} \frac{d_G(1, h(\partial F))^2}{2\text{vol}(F)},
\]

where the supremum is taken over the set of all graphs on \( M \). The fact that \( \hat{I}(h) \) is smaller than \( I(h) \) when \( h \) is, up to gauge transformation, the holonomy of an \( H^1 \) connection, follows from an energy inequality initially due to A. Sengupta [Sen98]. One must then consider an \( h \) such that \( \hat{I}(h) < +\infty \), prove that it is associated to an \( H^1 \) connection and that \( I(h) \leq \hat{S}(h) \). For this, we take graphs on the surface and look for connections with minimal energy which have, along the edges of this graph, the same value as \( h \). The existence of such connections is established by the following result.

Let \( \tilde{G} \) denote a universal covering of \( G \), endowed with a Riemannian metric which makes the covering map \( \pi : \tilde{G} \rightarrow G \) a local isometry. Recall that the fundamental group of \( G \) identifies with the fibre \( \pi^{-1}(1) \).

**Proposition 2.4.10 ([LN5, Proposition 22])** Let \( M \) be a closed surface. Let \( \mathcal{G} \) be a graph on \( M \) whose edges are smooth submanifolds of \( M \). Let \( g \) be an element of \( \mathcal{M}(\mathbb{E},G) \). Let \( \tilde{g} \) be an element of \( \mathcal{M}(\mathbb{E},\tilde{G}) \) such that for all \( e \in \mathbb{E} \) one has \( \pi(\tilde{g}(e)) = g(e) \). Let \( z \) be an element of \( \pi_1(G) \). Let \( \tilde{z} = (\tilde{z}_F)_{F \in \mathcal{F}} \) be an element of \( \pi_1(G)^\mathcal{F} \) such that \( \prod_{F \in \mathcal{F}} \tilde{z}_F = z \).

There exists a principal \( G \)-bundle \( P \) over \( M \) and a connection \( \omega \) on \( P \) such that the following properties are satisfied.

1. The isomorphism class of \( P \) is represented by \( z \).
2. The connection \( \omega \) is Lipschitz continuous and its restriction to the interior of each face is smooth.
3. The element of \( \mathcal{M}(\mathbb{E},G) \) determined by \( \omega \) is equal, up to gauge transformation, to
4. The contribution of each face $F$ to the Yang-Mills action of $\omega$ is

$$S_F(\omega) = \frac{d\tilde{g}(\partial F) z_F^2}{\text{vol}(F)}.$$ 

The connection $\omega$ whose existence is granted by this proposition is thus the analogue of the piecewise affine approximation of a continuous function in the proof of Schilder’s theorem. This connection has the minimal energy allowed by the energy inequality mentioned above, given the constraints imposed on its holonomy and on the isomorphism class of the bundle.

By doing this construction for our element $h$ and for a family of graphs which are finer and finer, we get a family of connection whose action is bounded by $\hat{I}(h)$. We need now to use a compactness result in order to extract a convergent subsequence. But while it is true and easy to check that the $H^1$ energy of a connection dominates its action, in that there exists a constant $K$ such that, at least locally, one has $S(\omega) \leq K\|\omega\|_{H^1}$, the reversed inequality is not true. The fact that a set of connections has bounded action does not imply that it has bounded $H^1$ norm. Fortunately, a difficult theorem of Uhlenbeck guarantees that one can apply an appropriate gauge transformation to each element of the set in such a way that the resulting set has bounded $H^1$ norm. This theorem allows us to conclude that the two expressions of our rate function coincide.

**Theorem 2.4.11 ([Uhl82])** Let $P$ be a principal bundle over $M$. Let $(\omega_n)_{n \geq 1}$ be a sequence of $H^1$ connections on $P$, such that the sequence $(S(\omega_n))_{n \geq 0}$ is bounded. Then there exists a $H^1$ connection $\omega$ such that, after extracting a subsequence from $(\omega_n)_{n \geq 1}$ and letting a gauge transformation act on each of its terms (the gauge transformation can be different for each term of the sequence), one has the weak convergence

$$\omega_n \xrightarrow{H^1} \omega.$$
Chapter 3

Unitary Brownian motion

One of the directions of study of the Yang-Mills field which look most appealing to me is the so-called large $N$ limit, that is, the asymptotic behaviour of this field taking its values in the unitary group $U(N)$ as $N$ tends to infinity. This point of view relates the Yang-Mills field to the huge field of research constituted by the study of random matrices and their limits. In particular, according to the seminal paper of I. Singer [Sin95], the theory of free probability should be a natural framework for expressing certain asymptotic properties of the Yang-Mills field.

The first step in the asymptotic study of the Yang-Mills field consists in studying the asymptotics of the Brownian motion on the unitary group. In this chapter, we are going to present three distinct works related to this Brownian motion, which correspond to three distinct aspects of the same object. We will start by recalling what this is about and by recalling some results which were known before our own work.

This chapter is written so that it might be read independently of the preceding two chapters.

3.1 Combinatorial aspects of the unitary Brownian motion [L6]

3.1.1 Definition of the unitary Brownian motion

Let $N \geq 1$ be an integer, which in all the sequel represents the size of the matrices which we consider. We consider the unitary group $U(N) = \{ U \in \text{GL}_N(\mathbb{C}) : UU^* = U^*U = I_N \}$. It is a connected compact subgroup of $\text{GL}_N(\mathbb{C})$. It is also a smooth real submanifold of dimension $N^2$ of $\text{M}_N(\mathbb{C})$. Its Lie algebra is the real linear subspace $\text{u}(N)$ of $\text{M}_N(\mathbb{C})$ formed by skew-Hermitian matrices : $\text{u}(N) = \{ A \in \text{M}_N(\mathbb{C}) : A^* = -A \}$. It is also the set of matrices $A$ such that $e^{tA}$ is unitary for all real $t$. For all $U \in U(N)$, the space $\text{u}(N)$ is left stable by the conjugation by $U$, that is, by the transformation $A \mapsto UAU^{-1}$. The linear action of $U(N)$ on $\text{u}(N)$ is called the adjoint action, or simply the action by conjugation. We endow $\text{u}(N)$ with the following scalar product:

$$\forall A, B \in \text{u}(N), \langle A, B \rangle = N \text{Tr}(A^*B).$$

Let us clarify our notation for the trace: we denote by $\text{Tr}(\cdot)$ the usual trace and by $\text{tr}(\cdot)$ the normalised trace, so that $\text{Tr}(I_N) = N$ and $\text{tr}(I_N) = 1$. The scalar product $\langle \cdot, \cdot \rangle$ is
invariant under the action of $U(N)$ by conjugation. It is however not the only one, even up to a multiplicative constant and the reasons why we have chosen this particular one will appear soon.

On the Euclidean space $(\mathfrak{u}(N), \langle \cdot, \cdot \rangle)$, there is a natural Brownian motion, which is the unique centred Gaussian process $(K_N(t))_{t \geq 0}$ indexed by $\mathbb{R}_+$ and with values in $\mathfrak{u}(N)$, whose covariance is the following:

$$\forall s, t \in \mathbb{R}_+, \forall A, B \in \mathfrak{u}(N), \quad \mathbb{E}[\langle A, K_N(t) \rangle \langle B, K_N(s) \rangle] = \min(s, t) \langle A, B \rangle.$$

One can think of $K_N$ as a stochastic process with values in a linear subspace of $\mathbb{M}_N(\mathbb{C})$ or, if one prefers, as a skew-Hermitian matrix of complex Brownian motions. From the latter point of view, the entries of $K_N$ are complex Brownian motions as independent as the condition of skew-Hermiticity allows them to be. More precisely, each diagonal coefficient has the distribution of $iB/\sqrt{N}$, where $B$ is a standard real Brownian motion, and each off-diagonal entry has the distribution of $(B + iB')/\sqrt{2N}$, where $B$ and $B'$ are two independent standard real Brownian motions. Finally, the family of entries located on the diagonal or above form an independent family. One checks that the matrix of quadratic covariations $\langle dK_N, dK_N \rangle_t$ is $-I_N \, dt$.

One defines the Brownian motion on the group $U(N)$ as the solution of the following stochastic differential equation:

$$dU_N(t) = U_N(t)dK_N(t) - \frac{1}{2}U_N(t) \, dt. \quad (3.1)$$

This equation is written in the Itô sense and its solution is a process with values in $\mathbb{M}_N(\mathbb{C})$. One can deduce an equation satisfied by $U_N^*$ and, given the quadratic covariation $\langle K_N, K_N \rangle$, a simple instance of Itô’s formula implies that $d(U_NU_N^*) = 0$, so that for all unitary initial condition $U_N(0)$, the process $U_N$ stays in the unitary group. Most of the time, we will take $U_N(0) = I_N$.

We are now going to describe the generator of $U_N$ and write the Itô formula under a form which will be useful. For this, observe that each element of $\mathfrak{u}(N)$ determines a left-invariant differential operator on $U(N)$. So, if $A$ belongs to $\mathfrak{u}(N)$, we define a differential operator $\mathcal{L}_A$ by setting, for all differentiable function $F : U(N) \to \mathbb{R}$ and all $U \in U(N)$,

$$(\mathcal{L}_A F)(U) = \frac{d}{dt}_{|t=0} F(Ue^{tA}).$$

Let us emphasise that in this formula, the function $F$ is evaluated only at points located in the unitary group. Let now $X_1, \ldots, X_{N^2}$ be an orthonormal basis of $\mathfrak{u}(N)$. We define the Laplace operator on $U(N)$ as the second-order differential operator

$$\Delta = \sum_{k=1}^{N^2} (\mathcal{L}_{X_k})^2.$$

One checks that this operator does not depend on the choice of the orthonormal basis. The Itô formula for $U_N$ writes as follows. Observe that, by definition of $K_N$, the processes $(\langle X_k, K_N \rangle : k = 1, \ldots, N^2)$ are independent standard real Brownian motions.
3.1. COMBINATORIAL ASPECTS OF THE UNITARY BROWNIAN MOTION

Proposition 3.1.1 (Itô formula) Let $F : \mathbb{R}_+ \times U(N) \to \mathbb{R}$ be a function of class $C^2$. For all $t \geq 0$, one has the equality

$$F(t, U_N(t)) = F(0, U_N(0)) + \sum_{k=1}^{N^2} \int_0^t \left( \mathcal{L}_{X_k} F(s, U_N(s)) \right) d\langle X_k, K_N \rangle_s + \int_0^t \left( \frac{1}{2} \Delta F + \partial_t F \right)(s, U_N(s)) \, ds.$$ 

Using the stochastic differential equation (3.1) or Itô’s formula, one checks that the Brownian motion is invariant by conjugation and inversion. This means that for all $V \in U(N)$, the processes $U_N, VU_NV^{-1}$ and $U_N^{-1}$ have the same distribution.

3.1.2 Distribution of the eigenvalues and Newton sums

Let $U$ be an element of $U(N)$. It admits $N$ eigenvalues $z_1, \ldots, z_N$ which are complex numbers of modulus 1 and we associate to $U$ its empirical spectral measure which is the following probability measure on the unit circle $U = \{ z \in \mathbb{C} : |z| = 1 \}$:

$$\hat{\mu}_U = \frac{1}{N} \sum_{i=1}^N \delta_{z_i}. \quad (3.2)$$

For all integer $N \geq 1$ and all real $t \geq 0$, the measure $\hat{\mu}_{U_N(t)}$ is a random probability measure on $U$ and the first result which we are going to present describes partially its distribution.

Since $U$ is a compact subset of $\mathbb{C}$, a probability measure $\nu$ on $U$ is characterised by its moments which are the numbers $\int_U z^n \nu(dz), n \in \mathbb{Z}$. These moments are complex numbers of modulus less than 1. If $\nu$ is a random measure, it is characterised by the distribution of the sequence of its moments, which in turn, since these moments are uniformly bounded, is characterised by all the expectations of the form

$$E \left[ \int_U z^{m_1} \nu(dz) \ldots \int_U z^{m_r} \nu(dz) \right], \ r \geq 1, m_1, \ldots, m_r \in \mathbb{Z}.$$ 

We are going to give a formula for these expectations for the measure $\hat{\mu}_{U_N(t)}$ when all the integers $m_1, \ldots, m_r$ have the same sign. Since the Brownian motion has the same distribution as its inverse, the measure $\hat{\mu}_{U_N(t)}$ has the same distribution as its image by complex conjugation and it suffices to consider the case where $m_1, \ldots, m_r$ are non-negative.

Observe that the moments of the measure $\hat{\mu}_{U_N(t)}$ can be written in a simple way in terms of $U_N(t)$. Indeed, for all $n \in \mathbb{Z}$, denoting by $z_1, \ldots, z_N$ the eigenvalues of $U_N(t)$, one has

$$\int_U z^n \hat{\mu}_{U_N(t)}(dz) = \frac{1}{N} \sum_{i=1}^N z_i^n = \text{tr}(U_N(t)^n).$$

We are thus going to consider quantities of the form $\text{tr}(U_N(t)^{m_1}) \ldots \text{tr}(U_N(t)^{m_r})$ where $m_1, \ldots, m_r$ are positive integers. Such a quantity, as a function of the eigenvalues of
$U_N(t)$ is a symmetric function, namely a product of Newton sums. The classical notation for such a product is $p_\lambda$, where $\lambda$ is the partition of the integer $n = m_1 + \ldots + m_r$ formed by $m_1, \ldots, m_r$ sorted in non-increasing order. However, instead of partitions, we are going to use permutations to index these functions. The fact that this indexation is very redundant will be compensated by other advantages.

In what follows, we will denote permutations as products of disjoint cycles, omitting cycles of length 1. Thus, the element $(124)$ of $S_4$ is the permutation which sends 1 on 2, 2 on 4, 3 on 3, and 4 on 1.

**Definition 3.1.2** Let $n \geq 1$ be an integer. Let $\sigma$ be an element of $S_n$. Let $m_1, \ldots, m_r$ be the lengths of the cycles of $\sigma$. Let $M$ be an element of $\mathbb{M}_N(C)$. We set

$$p_\sigma(M) = \operatorname{tr}(M^{m_1}) \ldots \operatorname{tr}(M^{m_r})$$

and

$$P_\sigma(M) = \operatorname{Tr}(M^{m_1}) \ldots \operatorname{Tr}(M^{m_r}) = N^r p_\sigma(M).$$

More generally, let $M_1, \ldots, M_n$ be elements of $\mathbb{M}_N(C)$. We set

$$p_\sigma(M_1, \ldots, M_n) = \prod_{\text{cycle of } \sigma} \operatorname{tr}(M_{i_1} \ldots M_{i_s})$$

and

$$P_\sigma(M_1, \ldots, M_n) = \prod_{\text{cycle of } \sigma} \operatorname{Tr}(M_{i_1} \ldots M_{i_s}) = N^r p_\sigma(M_1, \ldots, M_n).$$

Let us emphasise that the last two definitions make sense thanks to the invariance of the trace under cyclic permutation. Thus, for example, if $n = 4$ and $\sigma = (124)$, one has

$$p_\sigma(M_1, M_2, M_3, M_4) = \operatorname{tr}(M_1 M_2 M_4) \operatorname{tr}(M_3) = \operatorname{tr}(M_2 M_4 M_1) \operatorname{tr}(M_3).$$

### 3.1.3 Combinatorial expression of the moments of the spectral measure

In order to state the main result of this section, we need to introduce a few notions related to a discrete geometry of the symmetric group. Let $n \geq 1$ be an integer, $\mathfrak{S}_n$ the group of all permutations of $\{1, \ldots, n\}$, and $T_n \subset \mathfrak{S}_n$ the set of all transpositions of $\mathfrak{S}_n$. The set $T_n$ generates $S_n$ and one forms the corresponding Cayley graph by taking the elements of $\mathfrak{S}_n$ as vertices and by joining two elements $\sigma_1$ and $\sigma_2$ of $\mathfrak{S}_n$ by an edge if $\sigma_1 \sigma_2^{-1}$ belongs to $T_n$.

For each permutation $\sigma$, we call $|\sigma|$ the distance between $\sigma$ and the identity, that is, the smallest length of a decomposition of $\sigma$ as a product of transpositions. If $\ell(\sigma)$ denotes the number of cycles of $\sigma$, including the cycles of length 1, then $|\sigma| = n - \ell(\sigma)$. In order to understand this equality, one can for instance observe what happens when one multiplies a permutation $\sigma$ by a transposition $(kl)$. If $k$ and $l$ belong to the same cycle of $\sigma$, then this cycle is split into two cycles and $(kl)\sigma$ has one more cycle than $\sigma$. On the contrary, if $k$ and $l$ belong to two distinct cycles of $\sigma$, then these two cycles coalesce and $(kl)\sigma$ has one cycle fewer than $\sigma$. 

3.1. COMBINATORIAL ASPECTS OF THE UNITARY BROWNIAN MOTION [L6]

This remark implies also the following consequence: when one follows a path \((\sigma_0, \ldots, \sigma_k)\) in the Cayley graph, the number of cycles varies at each step by 1 or \(-1\): for each \(i = 0, \ldots, k - 1\), one has \(|\sigma_{i+1}| = |\sigma_i| \pm 1\).

**Definition 3.1.3** Let \((\sigma_0, \ldots, \sigma_k)\) be a path in the Cayley graph of \(S_n\). By the length of this path we mean the integer \(k\) and by its defect the integer

\[d = \#\{i \in \{0, \ldots, k - 1\} : |\sigma_{i+1}| = |\sigma_i| + 1\}.\]

For all \(\sigma, \sigma'\) in \(S_n\), all \(k, d \geq 0\), we denote by \(\Pi_k(\sigma)\) the set of paths of length \(k\) issued from \(\sigma\), by \(\Pi_k(\sigma \to \sigma')\) the set of paths of length \(k\) issued from \(\sigma\) and finishing at \(\sigma'\), and by \(S(\sigma, k, d)\) the number of paths issued from \(\sigma\), with length \(k\) and defect \(d\).

The defect of a path is thus the number of steps which take it further from the identity. If a path of length \(k\) and defect \(d\) joins \(\sigma\) to \(\sigma'\), then we have \(|\sigma'| = |\sigma| - (k - 2d)|. In particular, \(|k - 2d| \leq n - 1\).

The integer \(S(\sigma, k, d)\) is thus 0 as soon as \(|k - 2d| \geq n\). On the other hand, this integer depends only on the conjugacy class of \(\sigma\), since the Cayley graph is unchanged if we rename all vertices by conjugating them by a fixed permutation.

The main result is the following.

**Theorem 3.1.4 ([L6, Theorem 3.3])** Let \(N, n \geq 1\) be two integers. Let \(U_N\) be the Brownian motion on \(U(N)\) associated to the scalar product \(\langle X, Y \rangle = N \text{Tr}(X^*Y)\) on \(\mathfrak{u}(N)\). Let \(M_1, \ldots, M_n\) be elements of \(\mathcal{M}_N(\mathbb{C})\). Let \(\sigma\) be an element of \(S_n\). Then, for all \(t \geq 0\), we have the following expansion:

\[
\mathbb{E}[p_\sigma(M_1 U_N(t), \ldots, M_N U_N(t))] = e^{-\frac{nt^2}{2}} \sum_{k,d \geq 0} \frac{(-t)^k}{k!N^{2d}} \sum_{|\sigma'| = |\sigma| - (k - 2d)} \# \Pi_k(\sigma \to \sigma') p_{\sigma'}(M_1, \ldots, M_N).
\]

In particular, denoting by \(m_1, \ldots, m_r\) the lengths of the cycles of \(\sigma\), we have

\[
\mathbb{E}[\text{tr}(U_N(t)^{m_1}) \ldots \text{tr}(U_N(t)^{m_1})] = e^{-\frac{nt^2}{2}} \sum_{k,d \geq 0} \frac{(-t)^k}{k!N^{2d}} S(\sigma, k, d).
\]
For all $T \geq 0$, these series expansion converge normally for $(N, t) \in \mathbb{N}^* \times [0, T]$.

The sum over the permutation $\sigma'$ which occurs in the first expansion can be rewritten as a sum over all paths issued from $\sigma$ with length $k$ and defect $d$, taking for $\sigma'$ the finishing point of the path. This immediately implies the second expansion when all matrices $M_1, \ldots, M_N$ are taken equal to the identity matrix.

### 3.1.4 Enumeration of paths in the symmetric group

Let us make a few comments about the statement of Theorem 3.1.4. First of all, we have already observed that the numbers $S(\sigma, k, d)$ vanish when $|k - 2d| > n - 1$. Thus, for all $k \geq 0$, the contribution of order $t^k$ is a polynomial in $\frac{1}{N}$ and, for all $d \geq 0$, the contribution of order $\frac{1}{N^2}$ is polynomial in $t$.

Then, the numbers $S(\sigma, k, d)$ are extremely difficult to compute in general. As an amusement, one can check, using Figure 3.1, that $S((12)(34), 3, 1) = 104$. There is no formula for most of these numbers. The $S(\sigma, k, 0)$, which count paths which go straight to the identity, are a notable exception. This can be understood by observing that along these paths, two elements which do not belong to the same cycle will never be reunited: along these paths, cycles do only split. This indicates that it is enough to know $S(\sigma, k, 0)$ when $\sigma$ has only one non-trivial cycle and that in this case, the number depends only on the length of this cycle, not on the number of fixed points of $\sigma$ outside this cycle. Thus, it suffices to know the numbers $S((1\ldots n), k, 0)$.

**Proposition 3.1.5** ([L6, Proposition 6.6]) For all $n \geq 1$ and $k \geq 0$, one has

$$S((1\ldots n), k, 0) = \binom{n}{k+1} n^{k-1}.$$  

We recover for example the following classical result: the number $S((1\ldots n), n-1, 0)$ of ways of writing the $n$-cycle $(1\ldots n)$ as a product of $n-1$ transpositions is $n^{n-2}$.

Fortunately, the coefficients corresponding to $d = 0$ are exactly those which matter when $N$ tends to infinity. Since the series converges uniformly with respect to $N$, we can take the limit under the summation sign and recover the following result, which was proved in another way by P. Biane [Bia97].

**Proposition 3.1.6** For all $n \geq 1$ and all $t \geq 0$, one has

$$\lim_{N \to \infty} \mathbb{E}[\text{tr}(U_N(t)^n)] = e^{-\frac{n^2}{2}} \sum_{k=0}^{n-1} \frac{(-t)^k}{k!} \binom{n}{k+1} n^{k-1}.$$  

We will come back in greater detail to this result. For the time being, let us mention a formula which expresses the numbers $S((1\ldots n), k, d)$ for all $d \geq 0$. Let us recall the definition of the Stirling numbers of the second kind, $s(n, k)$, also denoted by $[n]_k$. They are defined by the following identity in $\mathbb{C}[x]$:

$$x(x-1)\ldots(x-n+1) = \sum_{k=0}^{n} \binom{n}{k} x^k.$$
One checks that $|\binom{n}{k}|$ is the number of permutations of $\mathfrak{S}_n$ at distance $n - k$ from the identity. In particular, $\binom{n}{0} = 0$. Let us take the convention $\binom{n}{k} = 0$ if $k < 0$. We have the following result.

**Proposition 3.1.7** For all $n, k, d \geq 0$,

$$S((1 \ldots n), k, d) = \frac{1}{n} \sum_{r,s,l,m \geq 0 \atop l + m = n - 1 - k + 2d} (-1)^{l+r} \frac{(n/2)^{s+r}}{r!s!} \left[ \begin{array}{c} s + 1 \\ r + 1 \end{array} \right] \left[ \begin{array}{c} r + 1 \\ s + 1 - l \end{array} \right].$$

The sum above is a finite sum. It is not obvious that the expression given here when $d = 0$ coincides with the previous expression - actually I don’t know how to prove it directly. Moreover, when $d > 0$, this formula does not allow one to compute $S(\sigma, k, d)$ for more general permutations. Indeed, when one counts paths with non-zero defect, one cannot work cycle by cycle as for paths with defect zero. Nevertheless, this formula allows one to prove the following amusing combinatorial result.

**Proposition 3.1.8** Let $n \geq 1$ be an integer. For all integer $p \geq 0$, let $c_{n,p}$ be the number of ways of writing the cycle $(1 \ldots n) \in \mathfrak{S}_n$ as a product of $p$ transpositions. The number $c_{n,p}$ is non-zero only if $p = n - 1 + 2d$ for a certain $d \geq 0$. In this case,

$$c_{n,p} = S((1 \ldots n), n - 1 + 2d, d) = \frac{n^p}{n!} \sum_{r=0}^{n-1} (-1)^r \binom{n-1}{r} \left( \frac{n-1}{2} - r \right)^p.$$

For all $n \geq 1$, one has the equality

$$\sum_{p \geq 0} c_{n,p} \frac{x^p}{p!} = \frac{1}{n!} e^{\frac{n(n-1)x}{2}} \left( 1 - e^{-nx} \right)^{n-1}.$$

In particular, $c_{n,n-1} = n^{n-2}$, $c_{n,n+1} = \frac{1}{24} (n^2 - 1)n^{n+1}$ and

$$c_{n,n+3} = \frac{1}{5760} (5n - 7)(n + 3)(n + 2)(n^2 - 1)n^{n+3}.$$

### 3.1.5 Unitary Brownian motion and symmetric random walk

In this paragraph, we are going to explain how Schur-Weyl duality makes Theorem 3.1.4 plausible, and state a result which relates the Brownian motion on $U(N)$ and the most natural random walk on the symmetric $\mathfrak{S}_n$.

**INSERT 11 – Schur-Weyl duality**

Let $n, N \geq 1$ be two integers. On the vector space $(\mathbb{C}^N)^{\otimes n}$ of $n$-linear forms on $(\mathbb{C}^N)^*$, there are two natural actions $\rho$ and $\pi$, respectively of the unitary group $U(N)$ and the symmetric group $\mathfrak{S}_n$, given by

$$\rho(U)(x_1 \otimes \ldots \otimes x_N) = Ux_1 \otimes \ldots \otimes Ux_n$$

$$\pi(\sigma)(x_1 \otimes \ldots \otimes x_N) = x_{\sigma^{-1}(1)} \otimes \ldots \otimes x_{\sigma^{-1}(n)}.$$
These actions commute to each other: for all $U \in U(N)$ and all $\sigma \in S_n$, one has $\rho(U) \circ \pi(\sigma) = \pi(\sigma) \circ \rho(U)$. Let $\mathcal{R}$ denote the subalgebra of $\text{End}((C^N)^{\otimes n})$ generated by the $(\rho(U) : U \in U(N))$ and $\mathcal{P}$ that generated by the $(\pi(\sigma) : \sigma \in S_n)$. Then each element of $\mathcal{R}$ commutes to each element of $\mathcal{P}$. Schur-Weyl’s theorem states that the algebras $\mathcal{R}$ and $\mathcal{P}$ are each other’s commutant.

**Theorem 3.1.9 (Schur-Weyl duality)** Each endomorphism of $(C^N)^{\otimes n}$ which commutes to $\mathcal{R}$ belongs to $\mathcal{P}$ and each endomorphism of $(C^N)^{\otimes n}$ which commutes to $\mathcal{P}$ belongs to $\mathcal{R}$.

By the Jacobson density theorem, which implies that a subalgebra of the algebra of endomorphisms of a finite-dimensional vector space is equal to the commutant of its commutant, the two assertions of the theorem are equivalent.

The algebra $\mathcal{P}$ is simply the algebra of the linear combinations of the $\pi(\sigma)$ where $\sigma$ runs over $S_n$. In order to describe it, it is convenient to introduce the group algebra $C[S_n]$ of $S_n$, which is the vector space of formal linear combinations $\sum_\sigma \lambda_\sigma \sigma$ of elements of $\sigma$ endowed with the multiplication

$$\sum_\sigma a_\sigma \sigma \sum_\sigma' b_{\sigma'} \sigma' = \sum_{\sigma, \sigma'} a_\sigma b_{\sigma'} \sigma \sigma'.$$

This algebra identifies naturally with the convolution algebra of functions on $S_n$. The action $\pi : S_n \to \text{End}((C^N)^{\otimes n})$ extends by linearity to $C[S_n]$ to give a morphism of algebras whose range is precisely $\mathcal{P}$.

We are going to apply Schur-Weyl duality to the image by $\rho$ of the Laplace operator on $U(N)$. In order to give a meaning to this image, we start by extending $\rho$ to $\mathfrak{u}(N)$, by differentiation: for all $A \in \mathfrak{u}(N)$, one can define $\rho(A) = \frac{d}{dt}|_{t=0} \rho(e^{tA}) = (L_A \rho)(I_N)$. Concretely, one has

$$\rho(A)(x_1 \otimes \ldots \otimes x_N) = \sum_{k=1}^N x_1 \otimes \ldots \otimes x_{k-1} \otimes Ax_k \otimes x_{k+1} \otimes \ldots \otimes x_N.$$

One can then define

$$\rho(\Delta) = \sum_{k=1}^{N^2} \rho(X_k)^2,$$

which is none other than $(\Delta \rho)(I_N)$ if we see $\rho$ as a smooth mapping from $U(N)$ into the vector space $\text{End}((C^N)^{\otimes n})$.

The operator has the property of being invariant under translations. This can be checked infinitesimally by checking that it commutes to all operators $L_A$, where $A \in \mathfrak{u}(N)$. This implies that the endomorphism $\rho(\Delta)$ of $(C^N)^{\otimes n}$ commutes to $\mathcal{R}$. According to Schur-Weyl’s theorem, it belongs to $\mathcal{P}$ and can thus be written as a linear combination of the $\pi(\sigma)$ for $\sigma \in S_n$.

Let us call $L$ the generator of the Markov chain on the Cayley graph of the symmetric group which jumps at rate $\frac{1}{N}$ from its present state to each of its $n(n-1)/2$ neighbours. The operator $L$ acts on functions from $S_n$ into $C$ according to the formula

$$Lf(\sigma) = -\frac{n(n-1)}{2N}f(\sigma) + \frac{1}{N} \sum_{\tau \in T_n} f(\sigma \tau).$$
If we identify the space of functions from $S_n$ into $C$ with the algebra $C[S_n]$, the action of $L$ is simply the multiplication by the element

$$L = -\frac{n(n-1)}{2N} + \frac{1}{N} \sum_{\tau \in T_n} \tau.$$ 

The result, implicitly present in the work of D. Gross and W. Taylor [GT93a], is then the following.

**Lemma 3.1.10** One has the equality

$$\frac{1}{2} \rho(\Delta) + \pi(L) = -\frac{n}{2} - \frac{n(n-1)}{2N}.$$ 

This result is remarkable because it relates two objects which have a probabilistic interpretation. To the expense of a small algebraic manipulation, it allows us to prove the following result.

**Proposition 3.1.11** Let $(\sigma_t)_{t \geq 0}$ be the Markov chain on $S_n$ with generator $L$. Then the process

$$\left( e^{\frac{ nt}{2} + \frac{n(n-1)t}{2N} P_{\sigma_t}(U_N(t))} \right)_{t \geq 0}$$

is a martingale. In particular, for all $t \geq 0$, one has the equality

$$\mathbb{E}[P_{\sigma_t}(U_N(t))] = e^{-\frac{ nt}{2} - \frac{n(n-1)t}{2N}} \mathbb{E}[N^{\ell(\sigma_0)}],$$

where $\ell(\sigma_0)$ denotes the number of cycles of $\sigma_0$.

This result, apart from the fact that it relates the Brownian motion on the unitary group to a random walk on the symmetric group, is the key to the proof of Theorem 3.1.4. Indeed, by letting the Markov chain $(\sigma_t)_{t \geq 0}$ start from the various elements of $S_n$, one gets enough information to find $\mathbb{E}[P_{\sigma}(U_N(t))]$ for all $\sigma$. It is then the change from $P_{\sigma}$ to $p_{\sigma}$ which forces one to consider the number of cycles of the permutations and to let the defect of paths into the picture.

### 3.1.6 Random ramified coverings

In the two papers [GT93a, GT93b], D. Gross and W. Taylor state many formulas which relate on one hand partition functions of the Yang-Mills measure on surfaces and with values in the unitary group and on the other hand average values of geometric quantities associated to ramified coverings over these surfaces. In some cases, singularities other than ramification points are authorised for these coverings. However, in the simplest case, where the surface is a disk, the partition function is simply the heat kernel on the unitary group and the coverings considered by Gross and Taylor are good ramified coverings. Theorem 3.1.4 allows one to prove rigorously their formula.

Let $D$ be the unit disk of the plane $\mathbb{R}^2$. We need to describe the probability measure on the set of ramified coverings which occur in this formula. It is not exactly of the same form as those which we have constructed in Section 2.1.2 but it looks very much like it.
The finite group which we consider here is the symmetric group $\mathfrak{S}_n$. The first difference with Section 2.1.2 is that instead of considering ramified $\mathfrak{S}_n$-principal bundles, we consider ramified coverings of degree $n$, without action of any group. There is however a bijective correspondence between these two types of coverings, analogous to the correspondence between the frame bundle and the tangent bundle of a smooth manifold.

Let $\pi : R \to D$ be a ramified covering of degree $n$. Let $y \in D$ be a ramification point. Let $l$ be a loop in $D$ which goes around $y$. Let us denote by $x$ the base point of $l$. If we label from 1 to $n$ the points of the fibre $\pi^{-1}(x) = \{\tilde{x}_1, \ldots, \tilde{x}_n\}$, then the mapping $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ which to each $i$ associates the label of the endpoint of the lift of $l$ issued from $x_i$ is a bijection. Changing the labelling of $\pi^{-1}(x)$ modifies this bijection by conjugating it by another bijection. The conjugacy class of $\sigma$ is in fact independent of all the choices which we have made, in particular that of $l$, and characterises the type of ramification at $y$. It is called the monodromy at $y$. We say that a ramification point is generic if the monodromy at this point is the conjugacy class of a transposition. This is equivalent to the equality $\#\pi^{-1}(y) = n - 1$.

The restriction of $\pi : R \to D$ to the boundary of $D$ is a covering of the unit circle, which is characterised by a conjugacy class of $\mathfrak{S}_n$, or equivalently a partition of the integer $n$, but we will stick to permutations.

**Definition 3.1.12** Let $D$ be the unit disk of $\mathbb{R}^2$. Let $n \geq 1$ be an integer. Let $\sigma$ be a permutation of $\mathfrak{S}_n$. Let $X$ be a finite subset of $D \setminus \{0\}$. We denote by $\mathcal{R}_{n,X,\sigma}$ the set of ramified coverings $\pi : R \to D$ of degree $n$ such that the following properties are satisfied.

1. For all $x \in X$, the covering $R$ admits at $x$ a generic ramification, that is, $\#\pi^{-1}(x) = n - 1$.
2. The covering $R$ is ramified at no point of $D \setminus (X \cup \{0\})$.
3. The monodromy of $R$ the boundary of $D$ is the conjugacy class of $\sigma$.

We denote by $\mathcal{R}_{n,\sigma}$ the union of the $\mathcal{R}_{n,X,\lambda}$ where $X$ runs over the set of finite subsets of $D \setminus \{0\}$.

There is no restriction on the ramification at 0: it can be a regular point or a point with an arbitrary ramification. This is what Gross and Taylor call an $\Omega$-point.

For all finite subset $X$ of $D \setminus \{0\}$, the set $\mathcal{R}_{n,X,\sigma}$ is finite. We endow it with the uniform measure

$$\rho_{n,X,\sigma} = \sum_{R \in \mathcal{R}_{n,X,\sigma}} \frac{n!}{\text{Aut}(R)} \delta_R.$$ 

Let us now consider a measure of area $\text{vol}$ on $D$, with total area equal to 1, for example $\frac{1}{\pi}$ times the Lebesgue measure. For all $t$, let $\Xi_t$ be a Poisson point process on $D$ with intensity $t \text{vol}$, which we see as a probability measure on the set $\mathcal{X}$ of finite subsets of $D \setminus \{0\}$. We then set

$$\rho_{n,\sigma}^t = \int_{\mathcal{X}} \rho_{n,X,\sigma} \Xi_t(dX).$$

The measure $\rho_{n,\sigma}^t$ is a finite measure, with total mass $e^{t(n^2 - 1)}$. We denote by $\mu_{n,\sigma}^t$ the corresponding probability measure on $\mathcal{R}_{n,\sigma}$.
Finally, for all \( \pi : R \rightarrow M \) in \( \mathcal{R}_{n,\sigma} \), we denote by \( k(R) \) the number of ramification points of \( R \) other than 0, and by \( \chi(R) \) the Euler characteristic of \( R \). The formula of Gross and Taylor is now the following.

**Theorem 3.1.13 ([L6, Theorem 8.3])** Let \( N, n \geq 1 \) be integers. Let \( \sigma \) be an elements of \( S_n \). Let \( t \geq 0 \) be a real. Then one has the equality

\[
\mathbb{E}[P_\sigma(U_N(t))] = e^{-\frac{nt}{2}} \int_{\mathcal{R}_{n,\sigma}} (-1)^{k(R)} N^{\chi(R)} \mu_{n,\sigma}^t(\mu).
\]

This theorem gives a geometric explanation for the fact that only non-negative even exponents of \( \frac{1}{N} \) appear in Theorem 3.1.4. Indeed, if we divide both sides of the equality by \( N^{\ell(\sigma)} \), where \( \ell(\sigma) \) is the number of cycles of \( \sigma \), we see that \( \mathbb{E}[p_\sigma(U_N(t))] \) expands in powers of \( N \) of the form \( N^{\chi(R) - \ell(\sigma)} \). Observe that the boundary of a covering \( R \in \mathcal{R}_{n,\sigma} \) has exactly \( \ell(\sigma) \) connected components. So, on one hand, \( \chi(R) - \ell(\sigma) \) has the same parity as \( \chi(R) + \ell(\sigma) \) which is the Euler characteristic of the closed surface obtained by gluing a disk along each connected component of its boundary, and is thus an even integer. On the other hand, let \( c(R) \) denote the number of connected components of \( R \), which is also the number of connected components of the closed surface \( R' \) described in the last sentence. Then \( \ell(\sigma) \geq c(R) = c(R') \), because from each regular point of \( R \) one can reach the restriction of \( R \) over the boundary of the unit circle, for example by following the horizontal lift of a well-chosen path. So, \( \chi(R) - \ell(\sigma) = \chi(R) + \ell(\sigma) - 2\ell(\sigma) = \chi(R') - 2\ell(\sigma) \leq \chi(R') - 2c(R') \leq 0 \). The powers of \( N \) which appear are thus non-negative even integers.

When \( N \) tends to infinity, only ramified coverings with maximal characteristic contribute: these are the coverings by a finite union of disks.

Let us conclude this section by emphasising that the link which this theorem reveals between the Yang-Mills field and random ramified coverings is not of the same nature as the link which we have described in Section 2.1, which was a link of analogy. The link revealed by the formula of Gross and Taylor is a link of duality, which can be understood as a consequence of Schur-Weyl duality.

### 3.2 Asymptotic behaviour of the eigenvalues [LM9]

#### 3.2.1 Distribution of the eigenvalues: asymptotic results

Proposition 3.1.6 indicates that the intensity of the random measure \( \hat{\mu}_{U_N(t)} \) has, for all \( t \geq 0 \), a limit as \( N \) tends to infinity. Let us start by describing this limit. We set \( U = \{ z \in \mathbb{C} : |z| = 1 \} \).

**Proposition 3.2.1** For all \( t \geq 0 \), there exists a unique probability measure \( \nu_t \) on \( U \) such that for all \( n \geq 0 \) one has

\[
\int_U \xi^n \nu_t(d\xi) = \int_U \xi^{-n} \nu_t(d\xi) = e^{-\frac{nt}{2}} \sum_{k=0}^{n-1} \frac{(-t)^k}{k!} \left( \begin{array}{c} n \\ k+1 \end{array} \right) n^{k-1}.
\]
This measure is also characterised by the fact that, for all complex number \( z \) in a neighbourhood of \( 0 \), one has the equality
\[
\int_{\mathbb{U}} \frac{1}{1 + z \exp \left( e^{\frac{t}{2}} \xi \right)} \nu_t(d\xi) = 1 + z.
\]

For all \( t > 0 \), the measure \( \nu_t \) admits a density with respect to the uniform measure on \( \mathbb{U} \), which is analytic inside its support. This support is an interval which contains 1, symmetric with respect to the real axis, which grows with \( t \) and covers the whole circle \( \mathbb{U} \) precisely at time \( t = 4 \) (see [Bia97]).

It is however impossible to express the density of \( \nu_t \) with usual functions only, unless one includes in such functions the Lambert \( W \) function, which is the reciprocal function of \( z \mapsto ze^z \).

Figure 3.2: Density of \( \nu_t \) at \( e^{i\theta} \) in function of \( \theta \in [-\pi, \pi] \) and \( t \). The support of the measure is known explicitly at each time. This support is the entire circle \( \mathbb{U} \) if and only if \( t \geq 4 \). The regularity of the density of \( \nu_t \) changes when \( t \) crosses this value, from Hölder continuous with exponent \( \frac{1}{2} \) for \( t < 4 \) to real analytic for \( t > 4 \). This is indeed a phase transition. For \( t > 4 \), one sees on the picture that the density converges rapidly to 1. This convergence is in fact exponential, uniformly in \( \theta \).

A convenient way of expressing limit theorems for the distribution of eigenvalues of unitary matrices whose size tends to infinity is to consider a function \( f : \mathbb{U} \to \mathbb{R} \) which is regular enough and to associate to it the sequence of real random variables \( (\text{tr}(f(U_N(t))))_{N \geq 1} \). Proposition 3.1.6 implies easily the following result.

**Proposition 3.2.2** Let \( f : \mathbb{U} \to \mathbb{R} \) be a continuous function. Then for all \( t \geq 0 \) one has the convergence
\[
\mathbb{E}[\text{tr}(f(U_N(t)))] \xrightarrow{N \to \infty} \int_{\mathbb{U}} f \, d\nu_t.
\]
In the case where the matrix is picked under the Haar measure, the invariance under translation of this measure implies immediately the following result.

**Lemma 3.2.3** Let $f : \mathbb{U} \to \mathbb{R}$ be a continuous function. For all $N \geq 1$, let $V_N$ be a random unitary matrix distributed according to the Haar measure on $U(N)$. Then for all $N \geq 1$, one has

$$
\mathbb{E}[\text{tr} f(V_N)] = \int_{\mathbb{U}} f(\xi) \, d\xi.
$$

This lemma can be seen as a limit of the previous one when $t$ tends to $+\infty$. Let us observe that, since the average repartition of the eigenvalues of $V_N$ is uniform on the circle, the random variable is defined almost surely as soon as $f$ is defined almost everywhere.

P. Diaconis and S. Evans have studied the fluctuations of $\text{tr} f(V_N)$ and they have established a central limit theorem [DE01]. The Sobolev space $H^{1\frac{1}{2}}$, the definition of which is recalled below, plays an essential role in this theorem. For all integrable function $f$ on $\mathbb{U}$ and all $n \in \mathbb{Z}$, we denote by $\hat{f}(n)$ the $n$-th Fourier coefficient of $f$.

**Definition 3.2.4** The space $H^{1\frac{1}{2}}(\mathbb{U})$ is the subspace of $L^2(\mathbb{U})$ defined by

$$
H^{1\frac{1}{2}}(\mathbb{U}) = \{ f \in L^2(\mathbb{U}) : \sum_{n \in \mathbb{Z}} |n||\hat{f}(n)|^2 < +\infty \},
$$

endowed with the Hilbert scalar product

$$
\langle f, g \rangle_{H^{1\frac{1}{2}}} = \hat{f}(0)\overline{\hat{g}(0)} + \sum_{n \in \mathbb{Z}^*} |n|\hat{f}(n)\overline{\hat{g}(n)}.
$$

The elements $H^{1\frac{1}{2}}$ are not necessarily continuous functions. The result of Diaconis and Evans is the following.

**Theorem 3.2.5** ([DE01]) Let $f_1, \ldots, f_n : \mathbb{U} \to \mathbb{R}$ be functions of $H^{1\frac{1}{2}}$ such that $\int_{\mathbb{U}} f_i(\xi) \, d\xi = 0$ for all $i \in \{1, \ldots, n\}$. Let $\Sigma$ be the $n \times n$ non-negative symmetric matrix defined by $\Sigma(f_1, \ldots, f_n) = (\langle f_i, f_j \rangle_{H^{1\frac{1}{2}}})_{i,j=1,\ldots,n}$. Then, with the notation of Lemma 3.2.3, one has, as $N$ tends to $+\infty$, the following convergence in distribution:

$$
N(\text{tr}(f_1(V_N)), \ldots, \text{tr}(f_n(V_n))) \xrightarrow{\text{law}} N(0, \Sigma(f_1, \ldots, f_n)).
$$

With Mylène Maïda, we have established an analogous theorem for random variables of the form $\text{tr} f(U_N(t))$. The expression of the covariance in this case is however more complicated. It requires the introduction of some notions of free probability.

**Insert 12 – Free probability and free unitary Brownian motion** [VDN92, ?, Bia97]

Free probability is a non-commutative analogue of the theory of probability, in which the notion of independence is replaced by the notion of freeness.
Definition 3.2.6 A non-commutative probability space is a pair \((A, \tau)\) where \(A\) is a complex involutive unital algebra and \(\tau\) is a linear form on \(A\) which is

1. **positive**: \(\tau(a^*a) \geq 0\) for all \(a \in A\),
2. **normalised**: \(\tau(1) = 1\),
3. **tracial**: \(\tau(ab) = \tau(ba)\) for all \(a, b \in A\).

The form \(\tau\) is also called a state.

The two fundamental examples of non-commutative probability spaces are on one hand \((L^\infty(\Omega, \mathcal{P}), \mathbb{E})\), where \((\Omega, \mathcal{P})\) is a classical probability space, and on the other \((\mathcal{M}_N(\mathbb{C}), \text{tr})\). Their tensor product is a new non-commutative probability space, the space of \(N \times N\) random matrices with bounded entries endowed with the state \(\mathbb{E}[\text{tr}(\cdot)]\). The main definition of free probability theory is the following.

Definition 3.2.7 Let \((A, \tau)\) be a non-commutative probability space. Let \(A_1, \ldots, A_r\) be involutive subalgebras of \(A\). The family \(\{A_1, \ldots, A_r\}\) is said to be free with respect to \(\tau\), or simply free, if the following property is satisfied.

\(\text{(F)}\) For all integer \(n \geq 1\), for all sequence of integers \(i_1, \ldots, i_n\) which belong to \(\{1, \ldots, r\}\) and such that \(i_1 \neq i_2, i_2 \neq i_3, \ldots, i_{n-1} \neq i_n\), for all \(a_1, \ldots, a_n\) such that \(a_1 \in A_{i_1}, \ldots, a_n \in A_{i_n}\) and \(\tau(a_1) = \cdots = \tau(a_n) = 0\), one has

\[
\tau(a_1 \cdots a_n) = 0.
\]

Elements \(b_1, \ldots, b_r\) of \(A\) are said to form a free family if the involutive subalgebras which they generate form a free family.

The examples of non-commutative probability spaces which we have already given are not rich enough to provide us with non-trivial examples of free families. For example, two elements of \(\mathcal{M}_N(\mathbb{C})\) are free. Let us start by assuming that \(A\) is a non-scalar projector and \(B\) is Hermitian and traceless. The formulas (3.3) imply the relations

\[
\text{tr}(AB^2) = 0, \quad \text{tr}(ABAB) = \text{tr}(A^2)\text{tr}(B^2) \quad \text{and} \quad \text{tr}((IN - A)B(IN - A)B) = (1 - \text{tr}(A))^2\text{tr}(B^2)
\]

from which it follows, since \(0 < \text{tr}(A) < 1\), that \(\text{tr}(B^2) = 0\) and \(B = 0\). It follows immediately that if \(A\) is a non-scalar projector and \(B\) is Hermitian or skew-Hermitian, then \(B\) is scalar. Since the assumption that \(A\) and \(B\) are free implies that \(A\) is free with any polynomial in \(B\) and \(B^*\), we find that a matrix which is free with a non-scalar projector is scalar. Finally, if we suppose only that \(A\) is not scalar, then one of the two diagonalisable matrices \(A + A^*\) and \(A - A^*\) admits two distinct eigenvalues, so that there exists a polynomial in \(A\) and \(A^*\) which is a non-scalar projector. The matrix \(B\) is free with this projector, hence scalar.

The simplest example of two free non-scalar elements is the following, which in fact has motivated the definition of the notion of freeness by D. Voiculescu. Let \(F_2\) be the free group on the set \(\{x, y\}\). The elements of \(F_2\) are thus reduced words in the letters \(x, y\) and their inverses, and one multiplies them by concatenating them and reducing them. We denote by \(1\) the empty word, which is the unit element. We consider the algebra \(A = \mathbb{C}[F_2]\) of formal finite linear combinations of elements of \(F_2\) with complex coefficients, endowed with the involution \((\sum_w c_w w)^* = \sum_w \bar{c}_w w^{-1}\). We endow \(A\) with the linear form

\[
\tau \left( \sum_w c_w w \right) = c_1.
\]
3.2. ASYMPTOTIC BEHAVIOUR OF THE EIGENVALUES \[ LM9 \]

One checks easily that \( \tau \) is a state. We have thus constructed a new example of a non-commutative probability space. In this space, the elements \( x \) and \( y \) are free, by definition of the free group \( F_2 \).

Let \( (A, \tau) \) be a non-commutative probability space. We say that an element \( a \in A \) is self-adjoint if \( a = a^* \) and an element \( u \) is unitary if \( uu^* = u^*u = 1 \). The positivity condition which is in the definition of a state implies that if \( a \) is self-adjoint, then the sequence \( (\tau(a^n))_{n \geq 0} \) is the sequence of the moments of a probability measure on \( \mathbb{R} \) which, if it is unique, is called the distribution of \( a \). A condition which suffices to ensure the unicity of this measure is that \( A \) is a Banach algebra and \( \tau \) is continuous. Indeed, in this case, \( \tau(a^n) = \mathcal{O}([|a|^n]) \), which implies that the distribution of \( a \) has compact support. Similarly, if \( u \) is unitary, the sequence \( (\tau(u^n))_{n \in \mathbb{Z}} \) is the sequence of the Fourier coefficients of a probability measure on \( U \), also called the distribution of \( u \). One checks successively, in the three examples that we have given, that the non-commutative distribution of a real random variable coincides with its usual distribution, that the distribution of a Hermitian or unitary matrix is its empirical spectral measure (see \( \text{(3.2)} \)) and that the elements \( x \) and \( y \) of \( \mathbb{C}[F_2] \), which are unitary, have the uniform measure on \( U \) as their distribution.

The notion of distribution in free probability identifies with the notion of moments. By definition, the distribution of a unitary element \( u \), for example, is the linear form on \( \mathbb{C}[t, t^{-1}] \) which to a polynomial \( P \) associates \( \tau(P(u, u^{-1})) \). In order to describe the joint distribution of several elements, one uses polynomials with several non-commuting indeterminates. We denote for example by \( \mathbb{C}(t_1, \ldots, t_n, t_1^{-1}, \ldots, t_n^{-1}) \) the algebra of polynomials with \( n \) indeterminates and their inverses, which can be identified naturally with \( \mathbb{C}[F_n] \), the group algebra of the free group on \( n \) letters.

**Definition 3.2.8** The joint non-commutative distribution of \( n \) unitary elements \( u_1, \ldots, u_n \) is the linear form on \( \mathbb{C}(t_1, \ldots, t_n, t_1^{-1}, \ldots, t_n^{-1}) \) which to a polynomial \( P \) associates \( \tau(P(u_1, \ldots, u_n, u_1^{-1}, \ldots, u_n^{-1})) \).

The convergence in distribution is also defined as the convergence of all moments.

**Definition 3.2.9** Let \( (A, \tau) \) and \( (A', \tau') \), \( r \geq 1 \) be non-commutative probability spaces. Let \( (u_i)_{i \in I} \) and \( (u_{i, r})_{i \in I, r \geq 1} \) be families of elements of these spaces indexed by the same set \( I \). We say that the sequence of families \( (u_{i, r})_{i \in I, r \geq 1} \) converges in non-commutative distribution to \( (u_i)_{i \in I} \) if for all \( n \geq 1 \), all \( t_1, \ldots, t_n \in I \) and all \( P \in \mathbb{C}(t_1, \ldots, t_n, t_1^{-1}, \ldots, t_n^{-1}) \) one has the convergence

\[
\tau_r(P(u_{i_1, r}, \ldots, u_{i_n, r}, u_{i_1}^{-1}, \ldots, u_{i_n}^{-1})) \underset{r \to \infty}{\longrightarrow} \tau(P(u_1, \ldots, u_n, u_1^{-1}, \ldots, u_n^{-1})).
\]

We can now define the free unitary Brownian motion, which is a particular process with stationary free increments.

**Definition 3.2.10** Let \( (A, \tau) \) be a non-commutative probability space. A free multiplicative Brownian motion is a family \( (u_t)_{t \geq 0} \) of unitary elements of \( A \) which satisfies the following properties.

1. For all \( 0 \leq t_1 \leq \ldots \leq t_n \), the elements \( u_{t_1}, u_{t_2} u_{t_1}, \ldots, u_{t_n} u_{t_1}^{-1} \) are free.
2. For all \( 0 \leq s \leq t \), the distribution of \( u_t u_s^* \) is equal to the distribution of \( u_{t-s} \).
3. For all \( t \geq 0 \), the distribution of \( u_t \) is the probability measure \( \nu_t \) on \( U \).

This free process is the limit in distribution in the sense of Definition \[3.2.9\] of the unitary Brownian motion.

**Proposition 3.2.11** The sequence of families of random variables \( \{(U_N(t))_{t \geq 0}\}_{N \geq 0} \) where each \( U_N(t) \) is seen as an element of the non-commutative probability space \( (L^\infty(\Omega, \mathcal{F}), \mathbb{M}_N(C), \mathbb{E} \otimes \text{tr}) \) for an appropriate \( (\Omega, \mathcal{F}) \), converges in non-commutative distribution to a free unitary Brownian motion.

It is not obvious, but true, that there exists a free unitary Brownian motion. One can even construct an arbitrary finite number of them in such a way that the algebras which they generate are mutually free.

The covariance of the fluctuations of the random variables of the form \( \text{tr} f(U_N(t)) \) is given by the following bilinear form.
Definition 3.2.12 Let \((A, \tau)\) be a \(C^*\)-non-commutative probability space in which there exists three free multiplicative Brownian motions \(u, v, w\) which are mutually free. Let \(T \geq 0\) be a real. Let \(f, g : U \to \mathbb{R}\) be two functions of class \(C^1\). We define
\[
\sigma_T(f, g) = \int_0^T \tau(f'(u_s v_{T-s}) g'(u_s w_{T-s})) \, ds.
\]

The assumption that \((A, \tau)\) is a \(C^*\)-non-commutative probability space means that \(A\) is a \(C^*\)-algebra, which allows us, for all unitary element \(u \in A\) and all continuous function \(f : U \to \mathbb{R}\), to define, by functional calculus, a new element \(f(u)\) of \(A\). The central limit theorem, analogous to the theorem 3.2.5 of Diaconis and Evans, is the following.

Theorem 3.2.13 ([LM9, Theorem 2.6]) Let \(U_N\) be the Brownian motion on \(U(N)\) associated to the scalar product on \(u(N)\) given by \(\langle X, Y \rangle = N \text{Tr}(X^* Y)\). Let \(T \geq 0\) be real. Let \(n \geq 1\) be an integer. Let \(f_1, \ldots, f_n : U \to \mathbb{R}\) be functions of class \(C^1(U)\) whose derivatives are Lipschitz continuous. The \(n \times n\) real symmetric matrix defined by \(\Sigma_T(f_1, \ldots, f_n) = (\sigma_T(f_i, f_j))_{i,j \in \{1, \ldots, n\}}\) is non-negative. Moreover, as \(N\) tends to infinity, the following convergence in distribution of random vectors of \(\mathbb{R}^n\) holds:
\[
N \left( \text{tr} f_i(U_N(T)) - \mathbb{E} \left[ \text{tr} f_i(U_N(T)) \right] \right) \xrightarrow{\text{law}} \mathcal{N}(0, \Sigma_T(f_1, \ldots, f_n)). \tag{3.4}
\]

The regularity of the functions that we consider is not there to ensure the positivity of the matrix \(\Sigma_T\), which is true already for \(C^1\) functions, but rather to ensure that our proof works, since it relies on arguments of stochastic calculus. However, the result of Diaconis and Evans and the convergence of the Brownian motion to its invariant measure suggest that it should be possible to improve this result by enlarging the class of functions for which the result holds. We have at least established the following result for the covariance.

Theorem 3.2.14 ([LM9, Theorem 8.3]) For all pair \((f, g)\) of functions of class \(C^1\) on \(U\) and whose derivatives are Lipschitz continuous, one has, with the notation of Theorems 3.2.5 and 3.2.13, the following convergence:
\[
\sigma_T(f, g) \xrightarrow{T \to \infty} \langle f, g \rangle_{H^1_2}.
\]

We have in fact obtained a better result by showing that, for \(T\) large enough, one can extend the definition of \(\sigma_T(f, g)\) to all pairs of function \((f, g)\) which belong to the space \(H^{\frac{1}{2}}\) and that with this extended definition, the previous result still holds. More precisely, denoting, for all \(j, k \in \mathbb{Z}\) and all \(T \geq 0\),
\[
\tau_{j,k}(T) = \int_0^T \tau \left( u_s v_{T-s} j (u_s w_{T-s})^k \right) \, ds,
\]
we have been able to prove a bound which guarantees us that, for \(T\) large enough, if \(f\) and \(g\) are two functions in \(H^{\frac{1}{2}}\), then the series
\[
\sigma_T(f, g) = - \sum_{j, k \in \mathbb{Z}} j k \hat{f}(j) \hat{g}(k) \tau_{j,k}(T)
\]
converges. Moreover, on one hand, if $f$ and $g$ have Lipschitz continuous derivatives, then the two definitions of $\sigma_T(f, g)$ which we have given coincide, and on the other hand, the convergence stated by Theorem 3.2.14 still holds with this new definition of $\sigma_T$.

Let us emphasise that the extended definition of $\sigma_T$ is valid only for $T$ large enough (say $T \geq 32$). Generally speaking, it is very difficult to control moments or covariances in small time. This could be compared with the fact that the measure $\nu_t$, the distribution of the free unitary Brownian motion, has its support strictly contained in $\mathbb{U}$ for $t < 4$ and an analytic density for $t > 4$. Figure 3.3 illustrates the fairly mysterious behaviour in small time of some of the functions which we have considered.

### 3.3. BETWEEN INDEPENDENCE AND FREENESS [LBG8]

With Florent Benaych-Georges, we have used the unitary Brownian motion to propose an interpolation between the notions of independence and freeness for elements of a non-commutative probability space. We are going to introduce this question from the point of view of the convolution of measures.
3.3.1 Convolutions

Let $\mu$ and $\nu$ be two probability measures on $\mathbb{R}$. The classical convolution of $\mu$ and $\nu$, denoted by $\mu * \nu$, is another probability measure on $\mathbb{R}$ which can be described as the distribution of the sum of two independent random variables of respective distributions $\mu$ and $\nu$. Another way of describing this measure is the following. Let $(A_N)_{N \geq 1}$ and $(B_N)_{N \geq 1}$ be two sequences of diagonal matrices such that for all $N \geq 1$, $A_N$ and $B_N$ have size $N \times N$ and, as $N$ tends to infinity, the empirical spectral measures $\hat{\mu}_{A_N}$ and $\hat{\mu}_{B_N}$ converge respectively to $\mu$ and $\nu$. For all $N \geq 1$, let $S_N$ be a random matrix distributed uniformly on the set of the $N!$ permutation matrices of size $N$. Then the following convergence in distribution holds:

$$\hat{\mu}_{A_N + S_N B_N S_N^{-1}} \xrightarrow{\text{law}} \mu * \nu.$$ 

For all $N \geq 1$, let us now consider a matrix $V_N$ uniformly distributed on $U(N)$. Then we have the following convergence:

$$\hat{\mu}_{A_N + V_N B_N V_N^{-1}} \xrightarrow{\text{law}} \mu \boxplus \nu,$$

where $\mu \boxplus \nu$ is the measure on $\mathbb{R}$ which can be described as the distribution of the sum of two free self-adjoint elements of a non-commutative probability space whose respective distributions are $\mu$ and $\nu$. This measure depends only on $\mu$ and $\nu$ since, just as for independence in the classical case, the joint distribution of two free elements is completely determined by their individual distributions.

For all $N \geq 1$, there is a natural interpolation between the distribution of $S_N$ and that of $V_N$: it is the family of the distributions of $U_N(t)S_N$ where $t$ runs over $\mathbb{R}_+$. For $t = 0$, we have $U_N(0)S_N = S_N$ and, when $t$ tends to $+\infty$, $U_N(t)S_N$ converges in distribution to a uniform unitary matrix.

**Theorem 3.3.1 ([LBG8, Corollary 2.10])** Let $\mu$ and $\nu$ be probability measures on $\mathbb{R}$. Let $(A_N)_{N \geq 1}$ and $(B_N)_{N \geq 1}$ be two sequences of $N \times N$ diagonal matrices as above. Let $U_N$ be a Brownian motion on $U(N)$. Let $t \geq 0$ be a real. Then, as $N$ tends to infinity, the empirical spectral measure of the matrix $A_N + U_N(t)S_N B_N S_N^{-1}U_N(t)^{-1}$ converges weakly in probability to a deterministic measure on $\mathbb{R}$ which depends only on $\mu$, $\nu$ and $t$, which we denote by $\mu *_t \nu$ and call $t$-free convolution of $\mu$ and $\nu$.

A classical example where one can compute explicitly the free convolution of two measures is that where $\mu = \nu = \frac{1}{2}(\delta_1 + \delta_{-1})$. In this case, $\mu * \nu = \frac{1}{4}\delta_2 + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_2$ and $\mu \boxplus \nu = \mathbb{1}_{[-2,2]}(x) \frac{dx}{\pi\sqrt{4 - x^2}}$, a dilation of the arcsine law. One can show that for all $t > 0$,

$$\frac{\delta_1 + \delta_{-1}}{2} *_t \frac{\delta_1 + \delta_{-1}}{2} = \mathbb{1}_{[-2,2]}(x) \frac{\rho_t(e^{i\arccos x/2})}{\pi\sqrt{4 - x^2}} \, dx,$$

where, for all $t > 0$ and all $\theta \in \mathbb{R}$, $\rho_t(e^{i\theta})$ is the density at $e^{i\theta}$ of the measure $\nu_t$ with respect to the uniform measure on $U$. 


3.3. BETWEEN INDEPENDENCE AND FREENESS [LBG8]

Figure 3.4: Density of $\frac{d_1 + d_{-1}}{2}$, $\frac{d_1 - d_{-1}}{2}t$ at $x \in [-2, 2]$ in function of $x$ and $t$. One can describe completely the support of the measure at each time. The first time at which this support is the whole interval $[-2, 2]$ is $t = 1$ and the last two points to enter this support are $-\sqrt{2}$ and $\sqrt{2}$.

3.3.2 Dependence structures and $t$-freeness

Just as in the case of independence and freeness, the existence of the $t$-free convolution of two measures is a by-product of the existence of a structure of dependence between subalgebras of a non-commutative probability space. By a structure of dependence, we mean the following: in a non-commutative probability space $(\mathcal{A}, \tau)$ where two subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are given, a way of reconstructing the restriction of $\tau$ to the subalgebra generated by $\mathcal{A}_1 \cup \mathcal{A}_2$ from the restrictions of $\tau$ to $\mathcal{A}_1$ and $\mathcal{A}_2$. In commutative terms, we would say that such a structure allows one to infer the joint distribution of a family of random variables from the knowledge of their individual distributions.

INSERT 13 – Dependence structures and universal models

A way of describing a structure of dependence between two subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ of a non-commutative probability space is giving a rule which allows one to compute $\tau(P(a_1, a_2))$ for all $a_1 \in \mathcal{A}_1$, $a_2 \in \mathcal{A}_2$ and all polynomial $P \in \mathbb{C}[t_1, t_2]$ given the moments of $a_1$ and $a_2$. It is exactly in this way that we have defined freeness\footnote{It would be more honest to say that it can be shown that the definition that we have given allows one, in principle, to perform this computation.} (see Definition 3.2.7). The rule $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ which characterises independence of random variables is also of this nature. Such a description does however not ensure that there exist non-trivial examples of the structure of dependence which one is defining. The example of two free elements which we have given in Insert 12, albeit natural, was not contained in the definition of freeness.

In classical probability theory, the construction which guarantees the existence of independent variables with arbitrary distributions is the Cartesian product of measurable spaces endowed with the tensor product.
of two probability measures. From the point of view of the algebras of random variables, this construction corresponds to the tensor product of algebras. Let us consider for example a probability space \((\Omega, P)\) and let us denote by \((A, \tau)\) the (commutative!) non-commutative probability space \((L^\infty(\Omega, P), \mathbb{E})\). Let us take two bounded random variables \(X\) and \(Y\) on \(\Omega\). Let us denote by \((A_X, \tau_X)\) the subalgebra of \(A\) generated by \(X\), that is, the set of polynomials in \(X\), endowed with the linear form induced by the expectation. Let us define \((A_{X,Y}, \tau_{X,Y})\) similarly from \(Y\).

The subalgebra of \(A\) generated by \(X\) and \(Y\), which is nothing but the space of polynomials in \(X\) and \(Y\), is not necessarily isomorphic to \(C[\mathbb{C}, X, Y]\) since there may be algebraic relations between \(X\) and \(Y\). Nevertheless, this subalgebra is the image of \(C[\mathbb{C}, X, Y]\) by the natural morphism of algebras \(P \mapsto P(X, Y)\). The algebra \(C[\mathbb{C}, X, Y]\) serves as a concrete model for the tensor product \(A_X \otimes A_Y\) which, in a vague way, is the largest possible algebra which is generated by \(A_X\) and \(A_Y\) and in which each element of \(A_X\) commutes to each element of \(A_Y\). There is in particular a unique morphism of algebras \(m : A_X \otimes A_Y \to A\) such that \(m(X \otimes 1) = X\) and \(m(1 \otimes Y) = Y\), and the range of this morphism is the subalgebra generated by \(X\) and \(Y\).

One can define a state \(\tau_X \otimes \tau_Y\) on \(A_X \otimes A_Y\) by setting \((\tau_X \otimes \tau_Y)(a \otimes b) = \tau_X(a) \tau_Y(b)\). This step corresponds to the construction of the tensor product of two measures. One can then, for all element \(c\) in \(A_X \otimes A_Y\), compare its expectation in \(A_X \otimes A_Y\) and in \(A\), or more precisely compare \((\tau_X \otimes \tau_Y)(c)\) and \(\tau(m(c))\). The following proposition is the purely algebraic reformulation of independence that we were looking for: the variables \(X\) and \(Y\) are independent if and only if the morphism of algebras \(m : A_X \otimes A_Y \to A\) preserves the expectations, that is, if \(\tau_X \otimes \tau_Y = \tau \circ m\). The general definition is the following.

**Definition 3.3.2** Let \((A, \tau)\) be a non-commutative probability space. Let \(A_1\) and \(A_2\) be two subalgebras of \(A\). Let \(\tau_1\) and \(\tau_2\) be the restrictions of \(\tau\) to \(A_1\) and \(A_2\) respectively. We say that \(A_1\) and \(A_2\) are independent if every element of \(A_1\) commutes to every element of \(A_2\) and if the natural morphism of algebras \(m : A_1 \otimes A_2 \to A\) satisfies the equality \(\tau \circ m = \tau_1 \otimes \tau_2\).

We have gained two things with respect to the commutative case: on one hand the ambient space \(A\) needs not be commutative and on the other, none of the algebras \(A_1\) and \(A_2\) need to be commutative. We only insist that they commute to each other. This is for example the case of the two subalgebras \(R\) and \(P\) of \(Eul((\mathbb{C}^N)^{\otimes n})\) defined in the context of Schur- Weyl duality (see Theorem 3.1.9). However, these two subalgebras are not independent.

The algebraic construction which corresponds to freeness is the free product of two algebras. If \(A_1\) and \(A_2\) are two unital algebras, their free product \(A_1 \ast A_2\) is the largest possible unital algebra which is generated by \(A_1\) and \(A_2\), with the only condition that the unit of \(A_1\) coincides with the unit of \(A_2\). More precisely, a free product of \(A_1\) and \(A_2\) is an algebra \(A\) endowed with two morphisms of algebras \(j_1 : A_1 \to A\) and \(j_2 : A_2 \to A\) such that, an arbitrary algebra \(B\) is given, as well as two morphisms of algebras \(m_1 : A_1 \to B\) and \(m_2 : A_2 \to B\), there exists a unique morphism \(m : A \to B\) such that \(m \circ j_1 = m_1\) and \(m \circ j_2 = m_2\). This description characterises the triple \((\tau, j_1, j_2)\) up to isomorphism.

A way of proving that such a free product exists consists in choosing in each of the unital algebras \(A_1\) and \(A_2\) a supplementary subspace of \(A\): we write \(A_1 = \mathbb{C} \oplus A_1^0\) and \(A_2 = \mathbb{C} \oplus A_2^0\). Then the algebra

\[
\mathbb{C} \oplus \left( A_1^0 \oplus A_2^0 \oplus (A_1^0 \otimes A_2^0) \oplus (A_1^0 \otimes A_2^0) \oplus \bigoplus_{n\geq 4} (A_1^0 \otimes A_2^0 \otimes A_1^0 \otimes A_2^0 \otimes \ldots) \right)_{n \text{ factors}}
\]

endowed with the most natural product (which is not absolutely straightforward to define precisely since the product of two elements of \(A_1^0\) for instance does not necessarily belong to \(A_1^0\)) and the morphisms \(j_1\) and \(j_2\) which identify respectively \(A_1\) and \(A_2\) with \(\mathbb{C} \oplus A_1^0\) and \(\mathbb{C} \oplus A_2^0\), is a free product of \(A_1\) and \(A_2\).

Given states \(\tau_1\) and \(\tau_2\) on \(A_1\) and \(A_2\), one can choose \(A_1^0\) and \(A_2^0\) as the kernels of \(\tau_1\) and \(\tau_2\). A natural state appears then on the free product, whose kernel is the hyperplane between the brackets in \((3.6)\). We denote this state by \(\tau_1 \ast \tau_2\) and it can be checked that \(A_1\) and \(A_2\) are free with respect to this state. We now have a characterisation of freeness analogous to the characterisation of independence given above.
3.3. BETWEEN INDEPENDENCE AND FREENESS [LBG8]

**Definition 3.3.3** Let \((\mathcal{A}, \tau)\) be a non-commutative probability space. Let \(\mathcal{A}_1\) and \(\mathcal{A}_2\) be two subalgebras of \(\mathcal{A}\). Let \(\tau_1\) and \(\tau_2\) be the restrictions of \(\tau\) to \(\mathcal{A}_1\) and \(\mathcal{A}_2\) respectively. We say that \(\mathcal{A}_1\) and \(\mathcal{A}_2\) are free if the natural morphism of algebras \(m : \mathcal{A}_1 * \mathcal{A}_2 \to \mathcal{A}\) satisfies the equality \(\tau \circ m = \tau_1 * \tau_2\).

We are going to define a universal model for \(t\)-freeness by defining a state on the free product of the algebras underlying two arbitrary non-commutative probability spaces.

**Definition 3.3.4** Let \((\mathcal{A}_1, \tau_1)\) and \((\mathcal{A}_2, \tau_2)\) be two non-commutative probability spaces. Let \(t\) be a positive real. Let \((\mathcal{U}, \nu)\) be a non-commutative probability space generated by a unitary element \(u_i\) whose distribution is the measure \(\nu_i\) on \(\mathcal{U}\). Let \(m\) be the unique morphism of algebras

\[
m : \mathcal{A}_1 * \mathcal{A}_2 \to (\mathcal{A}_1 \otimes \mathcal{A}_2) * \mathcal{U}
\]

such that \(m(a_1) = a_1 \otimes 1\) for all \(a_1 \in \mathcal{A}_1\) and \(m(a_2) = u(1 \otimes a_2)u^{-1}\) for all \(a_2 \in \mathcal{A}_2\). We call \(t\)-free product of \(\tau_1\) and \(\tau_2\) the state \(\tau_1 * \tau_2\) on \(\mathcal{A}_1 * \mathcal{A}_2\) defined by

\[
\tau_1 * \tau_2 = [(\tau_1 \otimes \tau_2) * \nu] \circ m.
\]

The fact that the measure \(\nu_i\) is invariant by complex conjugation implies that the pair \((u, u^{-1})\) has the same distribution as the pair \((u^{-1}, u)\), so that replacing \(m\) by the morphism \(m'\) such that \(m'(a_1) = u(a_1 \otimes 1)u^{-1}\) and \(m'(a_2) = 1 \otimes a_2\) would lead to the same definition of \(\tau_1 * \tau_2\).

For \(t = 0\), we recover the definition of the tensor product of two states, transported from the tensor product of the algebras to their free product by the natural morphism \(\mathcal{A}_1 * \mathcal{A}_2 \to \mathcal{A}_1 \otimes \mathcal{A}_2\). On the other hand, if \(t > 0\), the element \(u_i\) is not scalar in \(\mathcal{U}\) and this allows one to prove that \(m\) is injective. So, the subalgebra of \((\mathcal{A}_1 \otimes \mathcal{A}_2) * \mathcal{U}\) generated by \(\mathcal{A}_1 \otimes 1\) and \(u(1 \otimes \mathcal{A}_2)u^{-1}\) is a realisation of the free product of \(\mathcal{A}_1\) and \(\mathcal{A}_2\).

Once a universal model is defined, we can define \(t\)-freeness.

**Definition 3.3.5 ([LBG8, Definition 2.5])** Let \((\mathcal{A}, \tau)\) be a non-commutative probability space. Let \(\mathcal{A}_1\) and \(\mathcal{A}_2\) be two subalgebras of \(\mathcal{A}\). Let \(\tau_1\) and \(\tau_2\) denote the restrictions of \(\tau\) to \(\mathcal{A}_1\) and \(\mathcal{A}_2\) respectively. We say that \(\mathcal{A}_1\) and \(\mathcal{A}_2\) are \(t\)-free if the natural morphism of algebras \(m : \mathcal{A}_1 * \mathcal{A}_2 \to \mathcal{A}\) satisfies the equality \(\tau \circ m = \tau_1 * \tau_2\). We say that two subsets of \(\mathcal{A}\) are \(t\)-free if the involutive subalgebras which they generate are.

The observation made after Definition 3.3.4 ensures that this definition is symmetric in \(\mathcal{A}_1\) and \(\mathcal{A}_2\).

We now check that \(t\)-convolution corresponds to \(t\)-freeness.

**Proposition 3.3.6** Let \((\mathcal{A}, \tau)\) be a non-commutative probability space. Let \(t \geq 0\) be a real and let \(a, b \in \mathcal{A}\) be two self-adjoint elements which are \(t\)-free, with respective distributions \(\mu\) and \(\nu\). Then the distribution of \(a + b\) is \(\mu *_{t} \nu\).

We can define other \(t\)-free convolutions: if, with the notation of the proposition above, \(\mathcal{A}\) is a \(C^*\)-algebra and \(a, b\) are non-negative, we can define \(\mu \odot_{t} \nu\) as the distribution of \(\sqrt{\nu}a\sqrt{\mu}\). Also, if \(a\) and \(b\) are unitary, in which case \(\mu\) and \(\nu\) are measures on \(\mathcal{U}\), we can define \(\mu \odot_{t} \nu\) as the distribution of \(ab\).
CHAPTER 3. UNITARY BROWNIAN MOTION

3.3.3 Differential systems

When one tries to perform computations with pairs of $t$-free variables, one is led to compute expressions of the form

$$\tau(a_1 u_t b_1 u_t^* \ldots a_n u_t b_n u_t^*), \quad (3.7)$$

where the family $\{a_1, \ldots, a_n\}$ is independent of $\{b_1, \ldots, b_n\}$ and $u_t$ is free with the union of these two families.

The best grip that one has on such expressions is to consider them as functions of $t$ and to establish as small a differential system as possible which they satisfy. In order to differentiate with respect to $t$ an expression like (3.7), one uses a free stochastic differential equation satisfied by the free unitary Brownian motion, which is the analogue (and in a vague sense the limit as $N$ tends to infinity) of the stochastic differential equation (3.1) which defines the unitary Brownian motion [Bia97].

This derivative involves products of expressions of the form (3.7). By considering all possible products of expressions of this form where $a_1, \ldots, a_n, u_t b_1 u_t^*, \ldots, u_t b_n u_t^*$ appear each exactly once, one obtains a finite set of functions of $t$ which satisfies a closed differential system. By solving such systems, one obtains the following result.

**Proposition 3.3.7 ([LBG8, Proposition 3.5])** Let us define a function $G(t, z)$ in a neighbourhood of $(0, 0)$ in $\mathbb{R}_+ \times \mathbb{C}$.

Let $(\mathcal{A}, \tau)$ be a non-commutative probability space. Let $a$ and $b$ be two normal elements whose distributions have compact support and are symmetric (by this we mean that $a$ and $-a$ on one hand, $b$ and $-b$ on the other, have the same distribution). Choose $t \geq 0$. Assume that $a$ and $b$ are $t$-free. We set

$$G(t, z) = \sum_{n \geq 1} \tau((ab)^{2n}) e^{2nt} z^n.$$ 

Then $G$ is the unique solution, in a neighbourhood of $(0, 0)$ in $\mathbb{R}_+ \times \mathbb{C}$, of the non-linear equation

$$\partial_t G + 2z \partial_z (G^2) = 0,$$

$$G(0, z) = \sum_{n \geq 1} \tau(a^{2n}) \tau(b^{2n}) z^n.$$ 

This proposition allows one for example to prove (3.5). We can also use it to compute the distribution of the product of two $t$-free Bernoulli variables.

**Proposition 3.3.8** For all $t > 0$, one has the following equality of measures on $\mathbb{U}$:

$$\frac{\delta_{-1} + \delta_1}{2} \circ_t \frac{\delta_{-1} + \delta_1}{2} = \rho_t(\xi^2) d\xi,$$

where $\rho_t$ is the density of the measure $\nu_t$ with respect to the uniform measure.
3.3. BETWEEN INDEPENDENCE AND FREENESS

Figure 3.5: Density of $\delta_{t-1}^{+\frac{1}{2}} \odot t^{1+i} \delta_{t-1}^{+\frac{1}{2}}$ at $e^{i\theta}$ in function of $\theta$ and $t$. The support of the measure fills the entire circle for the first time at $t = 1$. The last points to enter the support are $i$ and $-i$.

3.3.4 Non-existence of $t$-free cumulants

By analogy with the case of freeness, we have wondered if it was possible to find $t$-free cumulants, that is, multilinear forms defined on every non-commutative probability space and which would vanish as soon as they are evaluated on a set of arguments which can be split into two non-empty families which are $t$-free.

Let $\mu$ be a compactly supported probability measure on $\mathbb{R}$. The classical cumulants of $\mu$ are the complex numbers $(c_n(\mu))_{n \geq 1}$ defined by the equality

$$\log \int_{\mathbb{R}} e^{zt} \mu(dt) = \sum_{n \geq 1} c_n(\mu) \frac{z^n}{n!}.$$  

Cumulants linearise the convolution: if $\mu$ and $\nu$ are compactly supported, then $c_n(\mu * \nu) = c_n(\mu) + c_n(\nu)$ for all $n \geq 1$.

Cumulants are related in a combinatorial way to moments. For all $n \geq 1$, let us denote by $m_n(\mu) = \int_{\mathbb{R}} t^n \mu(dt)$ the $n$-th moment of $\mu$. Let us consider the set $\mathcal{P}_n$ of all partitions of $\{1, \ldots, n\}$. For all $\pi \in \mathcal{P}_n$, let us define $m_\pi(\mu)$ as the product of the moments of $\mu$ whose orders are the sizes of the blocks of $\pi$. For example, $m_{\{1,3\},\{2,5\},\{4\}}(\mu) = m_1(\mu)m_2(\mu)^2$. Let us define in a similar way that cumulants $c_\pi(\mu)$. Then the relation which links moments and cumulants is the following: for all $n \geq 1$,

$$m_n(\mu) = \sum_{\pi \in \mathcal{P}_n} c_\pi(\mu).$$

This relation can be inverted in order to express each cumulant $c_n$ in terms of the moments $m_\pi$, by using the Möbius function of the lattice $\mathcal{P}_n$ endowed with the partial order for which $\pi_1 \leq \pi_2$ if each block of $\pi_1$ is contained in a block of $\pi_2$. We denote by $1_n = \{\{1, \ldots, n\}\}$ the largest element of $\mathcal{P}_n$. Then

$$c_n(\mu) = \sum_{\pi \in \mathcal{P}_n} \text{Moeb}(1_n, \pi)m_\pi(\mu).$$
Let \((\mathcal{A}, \tau)\) be a non-commutative probability space. By analogy with what precedes, for all partition \(\pi \in \mathcal{P}_n\) and all \(a \in \mathcal{A}\), we can define \(m_\pi(a)\) as the product of the moments \(m(a^k)\) of \(a\) corresponding to the sizes of the blocks of \(\pi\). In fact, if we choose a cyclic order on \([1, \ldots, n]\), for example the most natural one, we can extend this definition by setting
\[
\forall a_1, \ldots, a_n \in \mathcal{A}, \ m_\pi(a_1, \ldots, a_n) = \prod_{B \text{ block of } \pi} \tau(a_{i_1} \cdots a_{i_r}).
\]

The assumption of traciality that we have made on \(\tau\) implies that this definition does indeed only depend on the cyclic order of \([1, \ldots, n]\).

In order to define free cumulants, we need to introduce the lattice of non-crossing partitions.

**Definition 3.3.9** Let \(\pi\) be a partition of the set \([1, \ldots, n]\) endowed with the natural cyclic order. To each block \(B = \{i_1, \ldots, i_r\}\) of \(\pi\) one associates the polygon \(P_B\) which is the convex hull of the points \(e^{ik_i} \pi, \ldots, e^{ik_1} \pi\). We say that \(\pi\) is non-crossing if for all \(B\) and \(B'\) distinct blocks of \(\pi\), one has \(P_B \cap P_{B'} = \emptyset\).

![Figure 3.6: The partition \(\{\{1, 2, 3, 7\}, \{4, 6\}, \{5\}, \{8, 9, 10\}\}\) is non-crossing but \(\{\{1, 3, 7\}, \{2, 8, 9, 10\}, \{4, 6\}, \{5\}\}\) is crossing.](image)

The set of non-crossing partitions of \([1, \ldots, n]\) is denoted by \(\mathcal{NC}_n\). The partial order on \(\mathcal{P}_n\) induces a partial order on \(\mathcal{NC}_n\) which makes it a lattice. Let us denote by \(\text{moeb}\) the Möbius function of this lattice. Then the free cumulants are defined in the following way.

**Definition 3.3.10** Let \((\mathcal{A}, \tau)\) be a non-commutative probability space. Let \(a_1, \ldots, a_n\) be elements of \(\mathcal{A}\). One defines the free cumulant of \(a_1, \ldots, a_n\) by
\[
k_n(a_1, \ldots, a_n) = \sum_{\pi \in \mathcal{NC}_n} \text{moeb}(1_n, \pi)m_\pi(a_1, \ldots, a_n).
\]

Finally, for all \(a \in \mathcal{A}\) and all \(n \geq 1\), one sets \(k_n(a) = k_n(a, \ldots, a)\).

One has the following properties.

**Proposition 3.3.11** Let \((\mathcal{A}, \tau)\) be a probability space. Let \(\mathcal{A}_1\) and \(\mathcal{A}_2\) be two subalgebras of \(\mathcal{A}\). Then the following two properties are equivalent.

1. For all \(n \geq 1\) and all \(a_1, \ldots, a_n\) elements of \(\mathcal{A}\) which each belong either to \(\mathcal{A}_1\) or to \(\mathcal{A}_2\), but neither all to \(\mathcal{A}_1\) nor all to \(\mathcal{A}_2\), that is,
\[
\{a_1, \ldots, a_n\} \subset \mathcal{A}_1 \cup \mathcal{A}_2, \{a_1, \ldots, a_n\} \not\subset \mathcal{A}_1, \{a_1, \ldots, a_n\} \not\subset \mathcal{A}_2,
\]
3.3. BETWEEN INDEPENDENCE AND FREENESS \cite{LBGS}

one has $k_n(a_1, \ldots, a_n) = 0$.

2. $\mathcal{A}_1$ and $\mathcal{A}_2$ are free with respect to $\tau$.

In particular, if $a$ and $b$ are two free elements of $\mathcal{A}$, then one has for all $n \geq 1$ the equality $k_n(a + b) = k_n(a) + k_n(b)$.

Thus, free cumulants linearise the free convolution and characterise freeness.

Unfortunately, in the $t$-free case for $t > 0$, we have shown that there exists nothing as simple and powerful as the free cumulants. We have looked for candidates to the role of $t$-free cumulants in the following set of multilinear forms.

Definition 3.3.12 Let $(\mathcal{A}, \tau)$ be a probability space. Let $n \geq 1$ be an integer. Let $\sigma \in \mathfrak{S}_n$ be a permutation. One defines an $n$-linear form on $\mathcal{A}$ by setting, for all $a_1, \ldots, a_n \in \mathcal{A}$,

$$
\tau_\sigma(a_1, \ldots, a_n) = \prod_{(i_1 \ldots i_r) \text{ cycle of } \sigma} \tau(a_{i_1} \ldots a_{i_r}).
$$

This definition makes sense thanks to the traciality of $\tau$. We now look for a linear combination of the $\tau_\sigma, \sigma \in \mathfrak{S}_n$ with the property that it vanishes as soon as it is evaluated on arguments which can be split in two non-empty subsets which form two $t$-free families.

Definition 3.3.13 Let $n \geq 2$ be an integer. Let $t \geq 0$ be a real. A $t$-free cumulant of order $n$ is a collection $(c(\sigma))_{\sigma \in \mathfrak{S}_n}$ of complex numbers such that

$$
\sum_{\sigma \in \mathfrak{S}_n} c(\sigma) \neq 0
$$

and such that the following property is satisfied in every non-commutative probability space $(\mathcal{A}, \tau)$:

for all pair $(\mathcal{A}_1, \mathcal{A}_2)$ of subalgebras of $\mathcal{A}$ which are $t$-free with respect to $\tau$, and for all $a_1, \ldots, a_n$ elements of $\mathcal{A}$ which belong each to either to $\mathcal{A}_1$ or to $\mathcal{A}_2$, but neither all to $\mathcal{A}_1$ nor all to $\mathcal{A}_2$, one has

$$
\sum_{\sigma \in \mathfrak{S}_n} c(\sigma) \tau_\sigma(m_1, \ldots, m_n) = 0.
$$

Our result is the following. By $t$-free for $t = +\infty$, we mean free.

Theorem 3.3.14 (\cite{LBGS, Theorems 4.3 and 4.4}) 1. For all $t \in [0, +\infty]$ and all $n \in \{2, 3, 4, 5, 6\}$, there exists a $t$-free cumulant of order $n$. If moreover one insists that this cumulant is invariant by conjugation, that is such that $c(\sigma_1 \sigma_2 \sigma_1^{-1}) = c(\sigma_2)$ for all $\sigma_1, \sigma_2 \in \mathfrak{S}_n$, then it is unique up to multiplication by a constant.

2. There exists a $t$-free cumulant of order 7 if and only if $t = 0$ or $t = +\infty$. 

Perspectives

As a conclusion to these notes, I will indicate a few developments which I think arise naturally in the continuation of the work which I have presented here, and I will do so by taking approximately the themes in the order in which they have been treated.

After the construction of Markovian holonomy fields such as it was exposed here, it seems natural to try to complete the partial result of classification that we obtained (Theorem 1.3.6). We showed that to each regular Markovian holonomy field is associated an admissible Lévy process. Is it true that this Lévy process characterises completely the Markovian holonomy field? In other words, is a Markovian holonomy field determined by its partition functions? I believe that the answer is affirmative if the structure group is Abelian, but I don’t know the answer in the general case. If it turned out to be negative, which is not absolutely impossible, the best thing to do would be, in my view, to try to add an assumption to the definition of Markovian holonomy fields which would discard this ambiguity.

Still at the general level of the construction, I think that it would be interesting to pursue the point of view according to which a random holonomy field is a random homomorphism from the group of rectifiable loops on a surface and a Lie group. Ben Hambly and Terry Lyons have given in [HL06] a precise definition of the group of reduced rectifiable loops in a vector space or a manifold, and provided tools to handle them. It would probably be very difficult to understand the random variable associated to a loop by a random holonomy field as an explicit function of its signature (see [LCL] for more details on the signature of a path), but achieving this even in simple cases would probably illuminate the exact regularity of Markovian holonomy fields. A perhaps less fundamental question would be to study the kernel of the random homomorphism constituted by a random holonomy field. Here and in many other questions, the case of finite structure groups may be much simpler than the general case and yet inspiring.

A possible and natural extension of Markovian holonomy fields could be the definition of a random holonomy along random paths, maybe Brownian paths. I think that the techniques used in [L7] could hardly suffice to define a random holonomy along paths as irregular as typical Brownian paths, but it could be conceivable that subtle cancellations arise due to the particular structure of the Brownian motion, allowing one to define a non-trivial object. To the best of my knowledge, only one attempt has been made in this direction, by Sergio Albeverio and Shigeo Kusuoka [AK95].

However, a direction of research which to my eyes is much more important is that which would lead one to couple Markovian holonomy fields with other random objects.
These random objects could be paths, in the spirit of the last paragraph, but the physical origin of the objects which we consider indicates that the most interesting couplings would probably be those with objects which could be interpreted as matter fields. This question of coupling has a sense as much in a discrete setting, on graphs, where one would try to couple gauge fields with various models of statistical physics, as in a continuous setting, where one would perhaps consider random sections of a bundle associated with the underlying principal bundle. It is also possible that the meaningful objects are of non-commutative nature, thus incorporating at a more fundamental level the fermionic nature of matter.

The principle of large deviations which we have proved with James Norris concerns the Yang-Mills field, which is associated with the Brownian motion on the structure group. One may ask if there are similar results for other Markovian holonomy fields, associated with other Lévy processes. In order to establish such results, one would probably need to understand these Markovian fields as reflections of random connections, or mixtures of random connections and ramified coverings. This idea is close to that exposed by Gross and Taylor in [GT93a, GT93b] where, by mixing the Yang-Mills theory and more or less mysterious models of random coverings, they obtain remarkable formulas on the Brownian motion in the unitary group.

One aspect of the two-dimensional Yang-Mills theory which has raised the interest of physicists is the existence, when the structure group is the unitary group $U(N)$, of a phase transition in the limit where $N$ tends to infinity. This phase transition takes several forms depending on the point of view from which it is looked at, but I believe that it must be related to this fundamental phase transition: the limiting distribution of the eigenvalues of the Brownian motion on the unitary group, which for each $t \geq 0$ is a probability measure on the unit circle, is at first a Dirac mass at 1 for $t = 0$, then has its support growing progressively up to time $t = 4$, which is the first time at which this support covers the whole circle, and then becomes analytic for $t > 4$. For $t \in (0, 4)$, the moments of this limiting distribution decay like a power of their order, whereas for $t > 4$ they decay exponentially (see Section 3.2.1). The existence of a phase transition for the Yang-Mills measure on a sphere in function of the total area of the sphere [DK94] can be made plausible in the light of these facts, but there is still much to do in order to produce a rigorous statement. The computations on the Brownian bridge on the unitary group, which seem to be the natural way to attack such questions, become very quickly intractable, even more quickly than those related to the Brownian motion. A characteristic feature of these computations is the fact that one is constantly expressing small quantities (typically numbers of modulus less than 1) as alternated sums of huge numbers, which of course is not very convenient for estimating those quantities.

The graphs presented in Figure 3.3 also illustrate the difference of behaviour of the Brownian motion on $U(N)$ in small time and large time as $N$ tends to infinity. In the study of the fluctuations of random variables of the form $\text{tr} f(U_N)$, there is still much to do. We have stated a central limit theorem (Theorem 3.2.13) for functions with Lipschitz continuous derivative, but, with Mylène Maida, we suspect that this theorem is true for much less regular functions, at least when $T$ is large enough. A more precise conjecture would be that the class of functions for which the theorem is true is the space $H^2$ for
$T > 4$, and a class which depends on $T$ for $T \leq 4$. It is for instance possible that special conditions should be imposed on the boundary of the spectrum of $\nu_T$ when $T \leq 4$. However, we have no precise guess to offer about exactly what this class of functions could be.

With Florent Benaych-Georges, we are following our investigation of the notion of $t$-freeness, or of closely related notions, in particular in trying to prove theorems which mimic the classical theorems about sums of classical independent random variables. We hope in particular to obtain an interpolation between the Bercovici-Pata bijection and the identical transformation of the set of probability measures on $\mathbb{R}$.

A problem which mixes many of the questions which I have just mentioned, and which I find most intriguing, is this *master field* which I. Singer describes informally in his paper [Sin95] and to which many physicists have devoted some effort. In the case where the space-time is a disk, or the whole plane, all the tools required to describe this limiting field are available, in the framework of free probability, and to prove convergence results of the fields associated to the groups $U(N)$ towards this limiting field as $N$ tends to infinity. I believe that a consistent formulation of this master field could involve the group of reduced rectifiable loops which is a byproduct of the theory of *rough paths* developed by Terry Lyons [LCL], and which I have mentioned above. This group, which T. Lyons has studied with Ben Hambly, is a kind of free group with an uncountable number of generators (although it is not a free group in the algebraic sense) with which it is possible to do analysis. In the spirit of the description of I. Singer, I believe that the master field could be rigorously defined as a trace on this group and I am currently working in this direction.

Among the many aspects of the *large $N$ limit*, I find the Makeenko-Migdal equations [MM79] particularly appealing and they appear to me as a means to understand better the nature of the master field. I am trying to establish them rigorously and to understand them as a graphical translation, in the free group of the reduced loops in a graph, of the computation rules derived from the Itô formula applied to the Brownian motion in the unitary group. The proof of these equations which physicists give, based on the Schwinger-Dyson equations, is both beautiful and mysterious, and does not seem to let itself be easily translated into rigorous mathematical language.
Appendix

From the Maxwell equations to the Yang-Mills measure

In this appendix, I will try to sketch a path which relates the classical description of electromagnetism, the one that the reader is likely to have already studied and manipulated, to a small piece of quantum field theory which gives some insight into the physical interest of the Yang-Mills measure.

A.1 Fields an potentials

The classical theory of electromagnetism, as it was formulated at the end of the nineteenth century, consists of the four Maxwell equations and the Lorentz law. The Maxwell equations determine two time-dependent vector fields in $\mathbb{R}^3$, the electric field $\vec{E}$ and the magnetic field $\vec{B}$, in terms of the density of charge $\rho$ and the density of current $\vec{j}$. They read as follows:

\[
\begin{align*}
\text{div } \vec{E} &= \rho, \quad \text{(Gauss)} \\
\overrightarrow{\text{rot}} \vec{E} &= -\partial_t \vec{B}, \quad \text{(Faraday)} \\
\text{div } \vec{B} &= 0, \quad \text{(Thomson)} \\
\overrightarrow{\text{rot}} \vec{B} &= \vec{j} + \partial_t \vec{E}. \quad \text{(Ampère)}
\end{align*}
\]

The Gauss equation is the fundamental law of electrostatics, the Faraday equation allows one for example to compute the current induced in a conducting spiral by a varying magnetic field, the Thomson law is equivalent to the non-existence of magnetic charges (or magnetic monopoles) and the Ampère law allows one for example to compute the magnetic field produced by a conducting wire traversed by an electric current.

The six components of $\vec{E}$ and $\vec{B}$ are constrained, independently of the distribution and motion of charges in space, by two homogeneous equations, namely the Faraday equation and the Thomson law. This suggests that it should be possible to express their six components in terms of four scalar quantities. The Thomson law implies the existence, at least locally, of a vector field $\vec{A}$ such that $\vec{B} = \overrightarrow{\text{rot}} \vec{A}$. Such a field is called a vector potential. Once a vector potential has been chosen, the Faraday equation implies the existence, at least locally, of a scalar field $V$, called the electric potential, such that
\( \vec{E} = -\vec{\nabla} V - \partial_t \vec{A} \). When one is working in \( \mathbb{R}^3 \), the potentials can be defined globally and it is possible to define \( \vec{E} \) and \( \vec{B} \) in terms of the four components of \( \vec{A} \) and \( V \).

However, the potentials are never unique. One checks easily that if \( \vec{A} \) and \( V \) are suitable potentials in a given situation, then, for all time-dependent function \( \varphi : \mathbb{R}^3 \to \mathbb{R} \), the potentials \( \vec{A} + \vec{\nabla} \varphi \) and \( V - \partial_t \varphi \) are just as suitable. By varying \( \varphi \) one obtains all the pairs \((\vec{A}, V)\) which describe a given situation.

### A.2 The Schrödinger equation for a charged particle

The Lorentz law determines the force \( \vec{F} \) which is exerted on a particle in function of its charge \( q \), its speed \( \vec{v} \) and the electric and magnetic fields \( \vec{E} \) and \( \vec{B} \):

\[
\vec{F} = q\vec{E} + q\vec{v} \wedge \vec{B}.
\]

(A.1)

The first step of our progression consists in writing the Schrödinger equation for a charged particle exposed to a given electric and magnetic field.

In order to pass from a classical to a quantum theory, the first thing to do is to formulate the classical theory in a Lagrangian form, then in a Hamiltonian form. In our case, we need to find a function \( \mathcal{L}(\vec{r}, \vec{v}) \) of two vectors of \( \mathbb{R}^3 \) such that the Euler-Lagrange equations

\[
\frac{d}{dt} \left( \frac{\partial \mathcal{L}(\vec{r}, \vec{v})}{\partial \vec{v}_i} \right) = \frac{\partial \mathcal{L}(\vec{r}, \vec{v})}{\partial \vec{r}_i}, \quad i \in \{1, 2, 3\},
\]

are equivalent to the Lorentz law (A.1). One checks directly that if a pair of potentials \((\vec{A}, V)\) has been chosen, then the function

\[
\mathcal{L}(\vec{r}, \vec{v}) = \frac{1}{2} m \| \vec{v} \|^2 - qV(\vec{r}) + q\vec{v} \cdot \vec{A}(\vec{r})
\]

does this for us. In order to go from a Lagrangian formulation to a Hamiltonian one, the rule consists in defining an impulsion conjugated to \( \vec{r} \), denoted by \( \vec{p} \), defined by \( p_i = \frac{\partial \mathcal{L}}{\partial \vec{v}_i}, \quad i \in \{1, 2, 3\} \), and then in expressing the function \( \mathcal{H}(\vec{r}, \vec{p}) = \vec{p} \cdot \vec{v} - \mathcal{L}(\vec{r}, \vec{v}) \) in function of \( \vec{r} \) and \( \vec{p} \).

In our case, we find \( \vec{p} = m\vec{v} + q\vec{A} \) (in contrast with the most common case where \( \vec{p} = m\vec{v} \)), then \( \mathcal{H}(\vec{r}, \vec{p}) = \frac{1}{2} m \| \vec{v} \|^2 + qV(\vec{r}) \). In function of \( \vec{r} \) and \( \vec{p} \), we find

\[
\mathcal{H}(\vec{r}, \vec{p}) = \frac{1}{2m} \left( \vec{p} - q\vec{A} \right)^2 + qV(\vec{r}).
\]

With this definition of \( \mathcal{H} \), the Lorentz law is thus a particular case of the Hamilton equations:

\[
\frac{d\vec{r}_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial \vec{r}_i}, \quad i \in \{1, 2, 3\}.
\]

For us, the main interest of this formulation is that it allows us to write a Schrödinger equation in an automatic way. Indeed, we are going to define a linear operator on the Hilbert space \( L^2(\mathbb{R}^3) \) by replacing, in the expression of \( \mathcal{H} \), each scalar function of \( \vec{r} \) by
A.3. THE EFFECT OF A CHANGE OF GAUGE ON THE WAVE FUNCTION

the operator of multiplication by this function, and each component of \( \mathbf{p} \) by \(-i\) times the partial derivation in the corresponding direction. In this way, we find the Hamiltonian operator

\[
H = -\frac{1}{2m} \left( \nabla - iqA \right)^2 + qV.
\]

The Schrödinger equation writes then simply

\[
\partial_t \Psi = -iH \Psi,
\]

where \( \Psi \) is the wave function of a particle with charge \( q \).

A.3 The effect of a change of gauge on the wave function

It is a pleasant fact that the Hamiltonian which we have just written depends on the physical data \( \mathbf{E} \) and \( \mathbf{B} \) through the potentials \( V \) and \( A \). Indeed, these potential encode the same information contained in \( \mathbf{E} \) and \( \mathbf{B} \) in a less redundant way. However, we have already mentioned that these potentials are not unique. Choosing a particular pair \( (A, V) \) is called choosing a gauge and we are going to see that this is indeed the same gauge as the one mentioned in gauge theories, of which the Yang-Mills theory is an example.

Let us thus pick a gauge, as we did earlier, by choosing \( V \) and \( A \). For all time-dependent function \( \varphi : \mathbb{R}^3 \to \mathbb{R} \), let \( H_\varphi \) denote the Hamiltonian obtained by replacing \( A \) by \( A + \nabla \varphi \) and \( V \) by \( V - \partial_t \varphi \). For all \( \varphi \), the Hamiltonians \( H \) and \( H_\varphi \) describe the same physical situation, we could have obtained one as well as the other by the procedure which we have followed. As one could expect, there is a simple transformation which from a wave function satisfying the Schrödinger equation for \( H \) makes a wave function satisfying the Schrödinger equation for \( H_\varphi \). In order to determine this transformation, we observe that

\[
e^{iq\varphi} (\nabla - iqA)^2 e^{-iq\varphi} = \left( \nabla - iq(A + \nabla \varphi) \right)^2,
\]

from which we extract the relation

\[
H_\varphi = e^{iq\varphi} He^{-iq\varphi} - q \partial_t V,
\]

and the equivalence

\[
\partial_t \Psi = -iH \Psi \iff \partial_t (e^{iq\varphi} \Psi) = -iH_\varphi (e^{iq\varphi} \Psi).
\]

We can thus describe the same physical situation by two wave functions which differ by a phase which depends on the time and the point where one stands. Of course, the physics is the same, since these two descriptions are associated with two distinct Hamiltonians. Moreover, if we add a second particle to our system, its wave function will be affected by the same transformation as that of the first particle when passing from one Hamiltonian to the other. So, possible interferences between the wave functions which would have observable consequences will be identical in the two descriptions.
A.4 The geometrical nature of the vector potential

For the sake of simplicity, let us work in a situation where the electric field is zero and the magnetic field is constant. We choose \( V = 0 \). Even under these restrictive hypothese, we have just seen that a huge group acts on the set of possible ways of describing the physical situation which we consider: the group of all functions on \( \mathbb{R}^3 \) with values in the group \( U(1) \) of complex numbers of modulus 1. This group is called the gauge group – we now know why – and we denote it by \( \mathcal{J} \). The product in \( \mathcal{J} \) of two elements \( e^{i\varphi} \) and \( e^{i\varphi'} \) is simply \( e^{i(\varphi+\varphi')} \). An element \( e^{i\varphi} \) of \( \mathcal{J} \) acts on the wave function \( \Psi \) of a particle of charge \( q \) according to the formula \( (e^{i\varphi} \cdot \Psi)(\vec{r}) = e^{iq\varphi(\vec{r})}\Psi(\vec{r}) \). It also transforms the vector potential \( \vec{A} \) into \( \vec{A} + \vec{\nabla}\varphi \), and the differential operator \( \nabla - iq\vec{A} \) into \( \nabla - iq(\vec{A} + \vec{\nabla}\varphi) \).

In order to go from the multitude of descriptions to which we are confronted to an intrinsic description, which requires no particular choice of a gauge, we need to ground our description on an object whose symmetry group is the gauge group, that is, a principal bundle. In the present situation, we need to consider the space \( \mathbb{R}^3 \times U(1) \) endowed with the action, denoted on the right, of \( U(1) \) given by \( (\vec{r}, e^{i\theta}) \cdot e^{i\alpha} = (\vec{r}, e^{i(\theta+\alpha)}) \). This space is fibred by the projection on the first component \( \mathbb{R}^3 \times U(1) \to \mathbb{R}^3 \). An automorphism of this principal bundle is by definition a diffeomorphism of \( \mathbb{R}^3 \times U(1) \) which commutes to the action of \( U(1) \) and preserves the fibres, that is, which sends all pair \((\vec{r}, e^{i\theta})\) on a pair of the form \((\vec{r}, e^{i\theta'})\). It is not difficult to see that the group of automorphisms of the principal bundle \( \mathbb{R}^3 \times U(1) \) identifies with \( \mathcal{J} \): an element \( e^{i\varphi} \) of \( \mathcal{J} \) acts on a pair \((\vec{r}, e^{i\theta})\) according to \( e^{i\varphi} \cdot (\vec{r}, e^{i\theta}) = (\vec{r}, e^{i(\varphi+\theta)}) \).

We claim now that the wave function of a particle of charge \( q \) is in fact a function \( \psi : \mathbb{R}^3 \times U(1) \to \mathbb{C} \) which is \( q \)-equivariant in the following sense: for all \((\vec{r}, e^{i\theta}) \in \mathbb{R}^3 \times U(1) \) and all \( e^{i\alpha} \in U(1) \), one has

\[
\psi((\vec{r}, e^{i\theta}) \cdot e^{i\alpha}) = \psi((\vec{r}, e^{i(\theta+\alpha)}) = e^{iq\alpha}\psi(\vec{r}, e^{i\theta}).
\]

Each usual wave function, one of those which we have denote by \( \Psi \), is just the function \( \psi \) read through a particular section of the bundle \( \mathbb{R}^3 \times U(1) \) : for each function \( \sigma : \mathbb{R}^3 \to U(1) \), called section of the bundle, we can define

\[
\Psi_\sigma(\vec{r}) = \psi(\vec{r}, \sigma(\vec{r})).
\]

When the section \( \sigma \) is replaced by a new section \( \sigma' \), the latter can be written under the form \( e^{i\varphi}\sigma \) for a certain \( e^{i\varphi} \in \mathcal{J} \) and one finds the following relation between \( \Psi_\sigma \) and \( \Psi_{\sigma'} \):

\[
\Psi_{\sigma'}(\vec{r}) = \psi(\vec{r}, e^{i\varphi(\vec{r})}\sigma(\vec{r})) = e^{iq\varphi(\vec{r})}\Psi_\sigma(\vec{r}).
\]

Changing the section through which one reads the wave function produces thus the same effect on the wave function as changing the vector potential.

We now need to discuss the nature of \( \vec{A} \) in this new geometrical context. In the Schrödinger equation, the field \( \vec{A} \) appears as the 0 order term in the differential operator \( \nabla - iq\vec{A} \). In order to understand this term, we are going to see how the derivatives of \( \psi \) can be expressed in terms of those of \( \Psi_\sigma \).

Let thus \( \psi \) be our wave function and let \( \vec{r} \) be a point of \( \mathbb{R}^3 \). Let us choose a section \( \sigma : \mathbb{R}^3 \to U(1) \) and set \( \sigma(\vec{r}) = e^{i\theta} \). A tangent vector to \( \mathbb{R}^3 \times U(1) \) at \((\vec{r}, e^{i\theta})\) is a pair \((\vec{x}, i\xi)\)
A.5. A LAGRANGIAN FOR THE YANG-MILLS EQUATIONS

belonging to $\mathbb{R}^3 \oplus i\mathbb{R}$. Let us choose such a vector $(\vec{x}, i\xi)$ and compute the differential of $\psi$ at $(\vec{r}, e^{i\theta})$ in the direction $(\vec{x}, i\xi)$ in terms of $\Psi_\sigma$. Writing, for all $\vec{s} \in \mathbb{R}^3$, $\sigma(\vec{s}) = e^{i(\theta + \varphi(\vec{s}))}$, with $\varphi(\vec{r}) = 0$, we find

$$d_{(\vec{r}, e^{i\theta})} \psi(\vec{x}, i\xi) = \frac{d}{dt}\big|_{t=0} \psi((\vec{r} + t\vec{x}, e^{i(\theta + t\xi)})) = \vec{\nabla} \Psi_\sigma(\vec{r}) \cdot \vec{x} - iq(-\xi + \vec{\nabla} \varphi(\vec{r}) \cdot \vec{x}) \Psi_\sigma(\vec{r}).$$

We recognise a familiar expression. Indeed, if we constrain $\vec{x}$ and $\xi$ by the relation $\xi = -\vec{A} \cdot \vec{x}$ in the last expression, we find

$$(\vec{\nabla} - iq(\vec{A} + \vec{\nabla} \varphi)) \Psi_\sigma(\vec{r}) \cdot \vec{x} = d_{(\vec{r}, e^{i\theta})} \psi(\vec{x}, i\vec{A} \cdot \vec{x}).$$

In what precedes, we have implicitly used the constant section equal to $e^{i\theta}$ as reference. Any other section writes $e^{i\varphi} e^{i\theta}$ for some $e^{i\varphi} \in J$. In fact, the choice of $\theta$ is harmless, we could have taken $\theta = 0$, and we have thus obtained the relation

$$(\vec{\nabla} - iq(\vec{A} + \vec{\nabla} \varphi)) \Psi_{e^{i\theta}}(\vec{r}) \cdot \vec{x} = d_{(\vec{r}, e^{i\varphi})} \psi(\vec{x}, i\vec{A} \cdot \vec{x}).$$

Thus, applying the differential operator $\vec{\nabla} - iq(\vec{A} + \vec{\nabla} \varphi)$ to $\Psi_{e^{i\theta}}$ – both depending on the gauge – amounts to differentiating $\psi$ in a certain direction which does not depend on the gauge, but only on $\vec{A}$. In fact, $\vec{A}$ determines, at each point $\vec{r}$ of $\mathbb{R}^3$ and for all $e^{i\theta}$ of $U(1)$, a linear mapping from the tangent space to $\mathbb{R}^3$ at $\vec{r}$ into the tangent space to $\mathbb{R}^3 \times U(1)$ at $(\vec{r}, e^{i\theta})$:

$$\vec{x} \mapsto (\vec{x}, i\vec{A} \cdot \vec{x}) \in \mathbb{R}^3 \oplus i\mathbb{R}.$$
way, but also the electromagnetic field itself. In a caricatural way, one can say that, just
as a particle, classically represented at each time by a point $\vec{r}$ in $\mathbb{R}^3$, is represented in
a quantum theory by a wave function depending on time on $\mathbb{R}^3$, the electromagnetic
field, classically represented by a function $(\vec{E}, \vec{B})$ from $\mathbb{R}^3$ to $\mathbb{R}^6$, is replaced by a “wave
functional” on the space of these functions. And just like for particles, this quantification
requires that one writes the Maxwell equations in a Lagrangian form.

In order to do so, we need first to rewrite the Maxwell equations under a form which
emphasises better their geometric structure. Without even referring to the discussion
above where we have analysed the structure of $\vec{A}$, simple experiments where a physical
installation is compared with its image in a mirror or by a homothety indicate that
neither $\vec{B}$ nor $\vec{E}$ transform like vectors. They transform in fact like differential forms.
More precisely, it will be soon apparent that it is very convenient to define a 1-form $E$
and a 2-form $B$
by setting

$$E = E_x dx + E_y dy + E_z dz \quad \text{and} \quad B = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy.$$ 

One then combines these two forms into one single 2-form $F$ called electromagnetic field
by setting

$$F = B - dt \wedge E.$$ 

Finally, one encodes the density of charge $\rho$ and the density of current $\vec{j}$ by a differential
form, setting

$$J = -\rho dt + j_x dx + j_y dy + j_z dz.$$ 

If we endow $\mathbb{R}^4$ with the metric $-dt^2 + dx^2 + dy^2 + dz^2$, we define a Hodge operator on
differential forms, denoted by $\ast$. It differs from the operator associated with the Euclidean
metric only with an extra minus sign when it is applied to a form which involves $dt$.

The Maxwell equations can then be rewritten in very few symbols:

$$\begin{cases}
  dF = 0, & \text{(Gauss, Thomson)} \\
  d(\ast F) = \ast J. & \text{(Faraday, Ampère)}
\end{cases}$$ 

The homogeneous equation implies the existence of a 1-form $A$ such that $F = dA$.
With our previous notation, one can take $A = -V dt + A_x dx + A_y dy + A_z dz$.

This formulation of the Maxwell equations has the huge advantage of making it
transparent that they are invariant under the transformations of $\mathbb{R}^4$ which preserve the
Minkowski metric. Thus, the theory of electromagnetism as Maxwell had formulated it
was already a relativistic theory and it is indeed some of its consequences which have led
Einstein to formulate the theory of special relativity.

For us on the other hand who would like to quantify the electromagnetic field, a
formulation such as this one which does not anymore distinguish clearly between space
and time is slightly uncomfortable. We are simply going to check that the Lagrangian

$$\mathcal{L}(A) = \frac{1}{2} F \wedge \ast F + A \wedge \ast J$$ 

is a Lagrangian for the electromagnetic field in presence of the charges and currents
described by $J$. 

First of all, this Lagrangian comes out as a function of the 1-form $A$ rather than of $F$, and the relation $F = dA$ is implicit, which immediately implies $dF = 0$. Disregarding the issue of boundary conditions, we need now to determine the 1-forms $A$ for which the integral over $\mathbb{R}^4$ of the Lagrangian is stationary. If we vary $A$ by $\delta A$, then $F$ varies by $d\delta A$. Using the fact that, for arbitrary 2-forms $\alpha$ and $\beta$, one has $\alpha \wedge (\ast \beta) = (\ast \alpha) \wedge \beta$, we find, at the first order,

$$\mathcal{L}(A + \delta A) = \mathcal{L}(A) + \int_{\mathbb{R}^4} (d \ast F - \ast J) \wedge \delta A.$$  

Since the pairing $(\alpha, \beta) \mapsto \int_{\mathbb{R}^4} \alpha \wedge \beta$ is non-degenerate, it follows that the integral of the Lagrangian is stationary if and only if $d \ast F = \ast J$.

### A.6 Path integrals

With a Lagrangian formulation of the theory of electromagnetism in empty space, we can apply Feynman’s method of quantification by path integrals. In this method, one assigns to each configuration $A$ the integral of its Lagrangian $S(A) = \int_{\mathbb{R}^4} \mathcal{L}(A)$ and then a probability amplitude $e^{iS(A)}$. In order to determine the probability that a certain phenomenon occurs, one adds up all the amplitudes of the configurations of $A$ which could produce it and one computes the square of the modulus of the result. One thus arrives at integrals of the form

$$\int_{\mathcal{B}} e^{iS(A)} \, dA,$$

where $\mathcal{B}$ is the space of configurations corresponding to the phenomenon of which we want to compute the probability.

We have arrived at a point where integrals have appeared on sets of connections, with an exponential weight involving the Lagrangian of electromagnetism, that is, almost the point at which the introduction of these notes started. Replacing the Minkowski metric on $\mathbb{R}^4$ by the Euclidean metric, the integral of the Lagrangian becomes a positive function on the space of connections. Replacing $e^{iS(A)}$ by $e^{-\frac{1}{2}S(A)}$ and working in empty space, where $J = 0$, one finds the integral

$$\int_{\mathcal{B}} e^{-\frac{1}{2}S(A)} \, dA,$$

which is a variant of the heuristic expressions which we have given for the Yang-Mills measure (see (2) and (2.1)).
List of works


Seminars and proceedings


Lecture notes

References


REFERENCES


[GM95] David J. Gross and Andrei Matytsin. Some properties of large-


REFERENCES


## Contents

**Introduction**

- Presentation of the results ................................................................. 3
  - The mathematical difficulties of gauge theories ................................. 4
  - The construction of the Yang-Mills measure .................................... 7
  - Large unitary matrices ........................................................................ 9
  - Markovian holonomy fields .................................................................. 10
  - Localisation of the main statements .................................................... 11

1 Markovian holonomy fields [L2, L3, L7]

1.1 Introduction ............................................................................................ 13
  - The Poisson process modulo \( n \) indexed by loops ............................... 13
  - A first definition of Markovian holonomy fields .................................. 16
  - Topology of surfaces [Mas77, Moi77] ................................................... 16
  - Transition kernels and surgery of surfaces .......................................... 17
  - Multiplicative processes indexed by paths .......................................... 21
    - The measurable space ......................................................................... 21
    - Graphs ............................................................................................. 23
    - Lévy processes and marginals associated with graphs ....................... 24
      - Lévy processes on a compact Lie group [Lia04] .............................. 24
    - Gauge symmetry ................................................................................ 26
  - A random representation of the group of loops .................................... 28
    - Group of loops in a graph .................................................................. 28
    - Wilson loops [L3] .............................................................................. 30
    - Invariance by subdivision .................................................................... 31
    - Extension to rectifiable loops ............................................................. 34
    - Topology on the space of paths .......................................................... 34
    - The Banchoff-Pohl inequality [BP72, Vog81] ..................................... 35
  - Markovian holonomy fields .................................................................... 36
    - Definition ............................................................................................ 36
    - Existence and partial classification ...................................................... 38
    - The Yang-Mills field ......................................................................... 40
    - Link with topological quantum field theories ...................................... 42
      - Topological quantum field theories [Ati88] ...................................... 42

2 Holonomy fields, coverings and connections ........................................ 47

2.1 Finite groups and ramified coverings [L7, Chapter 5] ............................ 47
  - Fundamental group and coverings .......................................................... 47
  - Random ramified coverings .................................................................... 49
  - Brownian holonomy field and random connections [L4, LNS] ............... 51