ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΕΚΕΦΕ «ΔΗΜΟΚΡΙΤΟΣ»

ΙΝΣΤΙΤΟΥΤΟ ΝΑΝΟΕΠΙΣΤΗΜΗΣ ΚΑΙ ΝΑΝΟΤΕΧΝΟΛΟΓΙΑΣ

ΙΝΣΤΙΤΟΥΤΟ ΠΥΡΗΝΙΚΗΣ ΚΑΙ ΣΩΜΑΤΙΔΙΑΚΗΣ ΦΥΣΙΚΗΣ


Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών

Φυσική και Τεχνολογικές Εφαρμογές

# Μη-Αβελιανές Θεωρίες Βαθμίδες σε Χωροχρονικό Πλέγμα

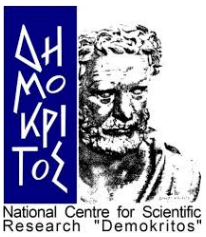
Μεταπτυχιακή Διπλωματική Εργασία του:


Καρύδη Κ. Ευάγγελου

Επιβλέποντες: Γιώργος Σαββίδης, Γιώργος Κουτσούμπας


Αθήνα, Ιούνιος, 2024


National Technical University of Athens
National Centre of Scientific Research "Demokritos"




**Master Degree in Physics**

# Pure Yang-Mills Theory on a Spacetime Lattice

## M A S T E R  T H E S I S

Submitted for the degree of Master by:

**Karydis K. Evangelos**

Under the supervision of:
Prof. George Savvidy
Prof. George Koutsoumbas

Athens, 2024

*Στην Ασπασία, την Ελευθερία και τον Κωνσταντίνο.*

# Acknowledgement

The author would like to express his personal gratitude to the supervising professor George Savvidy, for his continuous support during the preparation of this thesis, his immediate personal response to any difficulty and finally for his continuous effort for deep meaningful knowledge. Working with Professor George Savvidy made me realise that Theoretical research may sometimes be a discouraging procedure but every minute of it remains fascinating!

I would also like to thank Professor George Koutsoumbas, which was an inspiring teacher during the Master degree. He was the first person that introduced me to Quantum Field Theories. Last but not least, I would also like to express my gratitude to all the stuff of the Physics Department and of the Institute of High Energy Physics, for giving me the chance to start my journey in research in the field of Theoretical Physics by writing the present thesis.

# Abstract

The present thesis introduces the formulation of lattice gauge theories as originated by K.G. Wilson[28]. Gauge theories play a crucial role in the modern understanding of elementary particles and their interactions, with a significant application in describing the strong interactions between quarks, mediated by gluons. While Quantum Electrodynamics has achieved remarkable success using perturbation theory via Feynman diagrams, the strong interaction cannot be studied using the same approach across its entire spectrum due to the lack of a consistently small coupling constant. The strong interaction exhibits asymptotic freedom at small quark separations, where perturbation theory holds, but at large separations, quarks appear confined, indicating that the coupling constant becomes too large for perturbative expansion. Lattice gauge theory provides a method for studying gauge theories in their non-perturbative regime and offers a formulation that can be simulated computationally. This thesis presents a review of gauge fields on a lattice and a Python implementation of lattice gauge theory simulations using Metropolis Monte Carlo technique, following the pioneering work of M. Creutz and others, to measure important physical observables [7].

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The present thesis is divided in three main chapters, in order to introduce gauge theories, and especially non-Abelian gauge theories and their formulation on a discrete spacetime lattice. The first chapter constitutes of two distinct parts. The first part is dedicated to introducing the basic notions of differential geometry needed and used in modern theories of theoretical physics. This section was inspired by some lectures on General Relativity organised by Prof. George Savvidy during the time this thesis was still in preparation. Despite the fact that General Relativity and non-Abelian gauge theories do not have a complete analogy, the ideas presented in the first section of Chapter 2 are applicable in Yang-Mills theory after some appropriate modifications. Another reason the framework of General Relativity is introduced is that in that way it is clearer which quantities are Lorentz invariant and can be used in the rest of the thesis, without the urge to explain their invariance.

In the second part of Chapter 2 some basic ideas of Group Theory are introduced in a mathematically rigour way. Groups are used to express transformation of physical systems and therefore their potential symmetries under such transformation. More specifically gauge theories, which is the subject of the present thesis, are in fact theories that remain invariant under a transformation $g : \mathbb{R}^{1,3} \mapsto G$, where $G$ is a Lie Group. In the context of Standard Model, strong interactions are considered non-Abelian gauge theories with $SU(3)$ as their symmetry group. As a result, the purpose of this section is to describe the basic Lie Groups used in Theoretical Physics and their connection to smooth manifolds and Lie Algebras.

Chapter 3 starts with an exploration of gauge invariance in Classical Electrodynamics. Maxwell' s equations are naturally gauge and Lorentz invariant. This fact makes possible to formulate a Lagrangian density in terms of the electromagnetic field strength tensor $F_{\mu\nu}$. With this formulation an analogy between the concepts of differential geometry introduced in Chapter 2 and electromagnetism can be established.

The introduction of the field strength tensor makes it possible to expand the classical theory of Maxwell to a general theory that is invariant after general gauge transformations. These theories are known as Yang-Mills theories [31] and are presented in the second part of Chapter 3. The covariant derivative and the field strength tensor are introduced in these theories using the parallel transporting field $W(x, y)$, which will is the main quantity used to define

lattice gauge theory. This field makes possible to connect the values of a matter field in two different spacetime points, but most importantly can be used to construct the gauge invariant quantity called the "Wilson loop", that can be connected with the electromagnetic tensor and therefore the action of the system. The Wilson loops can be an indicator of confinement in the sense that it must behave following an area law when a meson is confined and a perimeter law when the quarks are free inside a meson.

Chapter 4 is the main work of this thesis. In the first sections pure Yang-Mills theory is formulated on the lattice as introduced by Wilson [28]. More specifically it is proved that the quantity named "Wilson loop ($W$)", introduced in the previous Chapter, is a measure of the term $F_{\mu\nu}F^{\mu\nu}$ and therefore it can be used to construct the action of the theory. As a result, the action on the lattice is expressed only as a sum over all the elementary plaquettes of the discrete spacetime lattice. Observable can be calculated using the usual path integral formulation, where now the integration is performed over $W \in G$. To this end, a section on integrating over Lie group manifold is embedded in Chapter 4. Integration over the group limits the observables that can be measured on the lattice and makes the usual procedure of gauge fixing useless.

After having defined some of the most important observables that can be measured with lattice gauge theory, in Section 4.6 the method of Renromalization and extraction of physical-real results on the lattice is presented. by this section it becomes clear that the lattice is another renormalization scheme and not a tool for computer simulations. To complete lattice gauge theory the Hilbert space of states is defined in Section 4.7. Having the Hilbert space, the transfer matrix and the Hamiltonian of the system are defined as well, without, though, extensive details. More details can be found in [22]. For the Hamiltonian formulation of lattice gauge theory, there is a clear connection between these theories and quantum statistical systems. In Section 4.8 there are some details on the methods that can be used to define and calculate thermodynamic observables of lattice gauge theories.

Simulations appear as side effect and are really important to study non-perturbative effects of Yang-Mills theory. To this end some, of the basic algorithms that are used in the literature are presented. The basic idea of these algorithms is to produce a Markov Chain of configurations by sweeping the links on the lattice and changing their values. It can be shown that such a Markov Chain converges. The two most famous algorithms used to change the value of a link variable is the Metropolis Algorithm [16], according which every link variable is changes at random with a criterion towards the minimization of the action.

Another famous and fast algorithm is the heatbath algorithm, which was formulated by Creutz for $SU(2)$ gauge theories [5]. This algorithm picks a new value from the group manifold. As a result, there is no unacceptable picking. It has though the serious disadvantage that one must know a parametrization of the group measure, in order to apply this method. Tries have been made to extend these algorithm to larger matrix groups [3], but still remains an open question, whose answer will seriously improve lattice simulations. For the purposes of the present thesis a code was developed in python programming language. The main aim of the program is to reproduce Creutz's results from [5, 2] and to explore larger gauge groups as $SU(10)$, $SU(20)$. The code outputs the expectation values of Wilson loops whose size is defined by the user. It uses the Metropolis algorithm for updating the link variable and it

can handles $SU(N)$ groups of maximum size $N = 20$. The results, obtained from this code appear reliable for groups of low dimension when the output is the plaquette variable.

# Chapter 2

# Mathematical Background

## 2.1 Aspects of Differential Geometry

This section presents some general results of differential geometry, in a language that the physicist is used to. There is no intention of mathematical rigour. The whole section is presented in terms of the General Theory of Relativity, where the manifold is spacetime, which is somehow easier to imagine and understand, because all vectors take values on the tangent bundle. In Chapter 2 of the present thesis these results are converted to appropriate quantities of Yang-Mills theory where the geometry is much more complicated, because of the existence of the g-bundle, which is somekind of an internal space. It is emphasized that the analogy between general relativity and Yang-Mills theory is presented naively and in an intuitive manner.

### 2.1.1 Maps

A reason that set theory became quickly very important in mathematics is because one can define a rule that takes elements from an input set $M$ (often called domain) and match this element to a value from another set $N$ (often called target). If this rule satisfies the two conditions given in the Definition 2.1.1 bellow then it is called a function or a map from the domain set $M$ to the target set $N$ .

**Definition 2.1.1** (Function/Map)**.** A rule that assigns to *each* element of a domain set $M$ *one and only one* element of a target set $N$ is called a function/map.
The notation:

$$f : M \longrightarrow N$$

means that $f$ is a map that takes the elements of the domain set $M$ and assigns to them one value from the target set $N$.
The notation:

$$f : m \longmapsto n$$

means that the element $m \in M$ is mapped to the element $n \in N$.

According to Definition 2.1.1 it is possible for two different elements of the domain set to be mapped to the same element of the target set. It is also possible that some elements of the target remain unmatched, as there is no restriction for the elements of the target set. The subset of the target set $N$ that contains *all the matched elements* of $N$ is often called the *range of the function* and is denoted with $f(M)$, where $M$ denotes, as usual, the domanin set of the function $f$.

A consequence of the above is that some maps could match every different element of the domain set $M$ to a different element of the target set $N$. These special cases of maps are called injective or one-to-one $(1-1)$ maps.

**Definition 2.1.2** (Injective/One-to-one map)**.** A map $f : M \longrightarrow N$ is called injective or one-to-one $(1-1)$ if:

$$\forall\, m_1, m_2 \in M \text{ with } m_1 \neq m_2 :$$

$$f(m_1) \neq f(m_2)$$

The Definition 2.1.2 states that every element from the range of an injective map $f(M)$ is matched to exactly one element of the domain set $M$. This means that a map $f^{-1} : f(M) \longrightarrow M$ is well defined in the context of the Definition 2.1.1. This map is called the inverse[1] map of $f$ and, of course, it can only be defined if $f$ is an injective map. An obvious identity between the map $f$ and its inverse $f^{-1}$ is[2]:

$$\forall\, m \in M : \left(f^{-1} \circ f\right)(m) = m$$

$$\forall\, n \in N : \left(f \circ f^{-1}\right)(n) = n$$

A very usual proposition, which is practically the easiest and the most common way to prove that a map is injective, follows directly from the Definition 2.1.2 when it is formulated inversely. It is also commonly refered to as a Corollary of the Definition 2.1.2.

**Corollary 2.1.1.** *A map $f : M \longrightarrow N$ is called injective or one-to-one $(1-1)$ if:*

$$\forall\, m_1, m_2 \in M \text{ with } f(m_1) = f(m_2) :$$

$$m_1 = m_2$$

As stated above, the Definition 2.1.1 does not contain any restrictions for the target set $N$, as long as every element of the domain $M$ is mapped to exactly one element of the target. The

[1] That explains the use of the notation $f^{-1}$

[2] The operator $\circ$ is used, as usual, to denote map composition: $(f \circ g)(m) \equiv f(g(m)), \{m \in D_g | g(m) \in D_f\}$

result of this remark is that there could be elements of the target $N$ that are not linked with any elements of the domain $M$ through a map $f$, that is why the range $f(M)$ was defined and used above in the definition of the inverse. The special case of a map $f$, that links all elements of the target $N$ is the case of the *surjective* or *onto* map. An equivalent formulation for the surjective map could be the condition $f(M) = N$.

**Definition 2.1.3** (Surjective/onto map)**.** A map $f : M \longrightarrow N$ is called surjective or onto if:

$$\forall\, n \in N\ \exists m \in M :\ f(m) = n$$

**Definition 2.1.4** (Bijective map)**.** A map $f$ is called bijective if it is injective and surjective simultaneously.

In this section there was no mention on the elements of the target set or the domain set. The target and the domain can be chosen to be abstract and contain any possible mathematical object that serves the desired application. In that way, maps are defined in a general way preserving their applicability in many applications.

### 2.1.2 Vectors, Tensors and Invariants

The basic postulate of relativity is that physics should not be aware of the coordinate system chosen to be described on. In other words, all laws of physics should hold in every coordinate system. To formulate this postulate mathematically, let' s state at first that relativity give no distinction between space and time. Spacetime is described by 4 points in a 4 dimensional continuum forming a smooth manifold $\mathcal{M}$ of dimension $\dim\{\mathcal{M}\} = 4$. The following discussion and results are adapted from [24].

Every open subset of the spacetime manifold is equipped with a coordinate system, which is a homeomorphism $x :\ \mathcal{M} \longrightarrow \mathbb{R}^4$. Laws of physics must remain the same by changing the coordinate system in use. Equations that describe physical laws will change when passing from a coordinate system $x$ to a different coordinate system $x'$, but they will keep their structure. This property is known as covariance.

Formally, the change between the coordinate systems $x$ and $x'$ is a diffeomorphism expressed as:

$$\begin{aligned} x'^0 &= x'^0(x^0, x^1, x^2, x^3) \\ x'^1 &= x'^1(x^0, x^1, x^2, x^3) \\ x'^2 &= x'^2(x^0, x^1, x^2, x^3) \\ x'^3 &= x'^3(x^0, x^1, x^2, x^3) \end{aligned}$$

or in a more compact way:

$$x'^\mu = x'^\mu(x^\mu),\ \mu \in \{0, 1, 2, 3\} \tag{2.1}$$

A common convention is that greek indices ($\mu, \nu, \rho, \sigma$, etc.) take values in the set $\{1, 2, 3, 4\}$, where 1 denotes the time dimension and $\{2, 3, 4\}$ denote the three spatial dimensions. Latin indices ($i, j, k$, etc.) run only in the three spatial dimensions.

Assume a theory that contains a scalar field $\phi(x^\mu)$. The exterior derivative of this field in a new coordinate frame will be:

$$d\phi'(x'^\nu) = \frac{\partial \phi'(x'^\nu)}{\partial x'^\mu} dx'^\mu$$

where Einstein' s summation convention is implied. To avoid confusion, we can drop the arguments of the scalar function $\phi$ and let $\phi$ denote the scalar field in the old frame, while $\phi'$ is the scalar field in the new frame.

$$\begin{aligned}
d\phi' &= \frac{\partial \phi'}{\partial x'^\mu} dx'^\mu \\
d\phi' &= \frac{\partial \phi'}{\partial x'^\mu} \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu \\
d\phi' &= \frac{\partial \phi}{\partial x^\sigma} \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu \\
d\phi' &= \frac{\partial \phi}{\partial x^\sigma} \frac{\partial x^\sigma}{\partial x^\nu} dx^\nu \\
d\phi' &= \frac{\partial \phi}{\partial x^\sigma} \delta^\sigma_\nu dx^\nu \\
d\phi' &= \frac{\partial \phi}{\partial x^\nu} dx^\nu \\
d\phi' &= d\phi
\end{aligned} \tag{2.2}$$

Equation 2.2 shows that the 1-form $d\phi$ remains the same under the transformation of the coordinates. The quantities with this property are called **invariants**.

In the derivation of Equation 2.2 it was obvious by the chain rule that the scalar field transformed as:

$$\frac{\partial \phi'}{\partial x'^\mu} = \frac{\partial x^\nu}{\partial x'^\mu} \frac{\partial \phi}{\partial x^\nu} \tag{2.3}$$

The calculation of the exterior derivative $dx'^\mu$ gave a different transformation law:

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu \tag{2.4}$$

The quantities that transform according to the transformation law of equation of Equation 2.3 are called **covariant** vectors while the quantities that transform under the law of Equation 2.4 are called **contravariant** vectors. In order to distinguish the two different types of vectors,

covariant vectors are often denoted with a lower index, while for contravariant vectors upper indices are used.

$$A'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} A^{\nu} \tag{2.5}$$

$$B'_{\mu} = \frac{\partial x^{\nu}}{\partial x'^{\mu}} B_{\nu} \tag{2.6}$$

It immediately follows that quantities of the form $A^{\mu}B_{\mu} = A_{\mu}B^{\mu}$ transform like the differential $d\phi$, which means that they remain invariant in any coordinate frame of choice and are known as **scalar or inner products**. Of course, the fact that they remain invariant does not mean that they do not appear changed as functions. They just produce the same values for a specific point on the manifold, independently of the coordinates used to describe this point.

A product of the form $A^{\mu}B^{\nu}C^{\rho}\dots W_{\rho}V_{\kappa}\dots$ that contains covariant and contravariant components is called a tensor and transforms kind of independently for each covariant and contravariant component. For example, assume a tensor $T^{\mu\nu}_{\rho}$. According to the diffeomorphism of Equation 2.1 the components of this tensor transform as:

$$T'^{\mu\nu}_{\rho} = \frac{\partial x'^{\mu}}{\partial x^{\kappa}} \frac{\partial x'^{\nu}}{\partial x^{\lambda}} \frac{\partial x^{\sigma}}{\partial x'^{\rho}} T^{\kappa\lambda}_{\sigma}$$

A tensor is a generalized version of vectors. If a tensor contains only components that transforms covariantly (lower indices) then it is called a **covariant tensor** and respectively the term contravariant is used when it contains only contravariant components (upper indices). The total number of indices (upper and lower) appearing in a tensor characterizes its **rank**. A common convention is to call a scalar quantity a tensor of zero rank. Vectors are just tensors of rank 1.

Having introduced tensors in such a way[3], it immediately follows that the product $T_{\mu\nu\kappa}T^{\mu\nu\kappa}$ constitutes an invariant. Of course, it is generalized to tensors of any rank. Lastly, there exist some specific tensors with the property:

$$T_{\mu\nu} = \pm T_{\nu\mu} \tag{2.7}$$

The meaning of such equation is that some tensors remain unchanged (or appear multiplied by a minus sign) after interchanging two of their indices. If Equation 2.7 holds with the plus sign then the tensor is called symmetric. The tensor, which satisfies the previous equation with a minus sign is called antisymmetric or skew-symmetric. Of course, these property holds for tensors of higher rank, but for them the symmetry might not hold for all the components-indices. Then the symmetric or antisymmetric indices must be pointed out explicitly. For example one can say that a tensor $T_{\mu\nu\rho\sigma}$ is antisymmetric in its first two covariant indices. Then $T_{\mu\nu\rho\sigma} = -T_{\nu\mu\rho\sigma}$.

[3]which is not the most mathematically rigour way to do so.

### 2.1.3 Integration and Densities

By definition, invariant quantities produce the same value in every possible coordinate frame that can be described by the diffeomorphism of Equation 2.1. As a result, they can be integrated in any frame. Trying to do such an integration for an invariant quantity $A$ in a spacetime region one observes the following [4]:

$$\int A' dx'^4 = \int A \left| \frac{\partial x'^\mu}{\partial x^\nu} \right| dx^4 \neq \int A dx^4$$

So, one needs to define a new object $\mathscr{A}$, which transform as:

$$\mathscr{A}' = \left| \frac{\partial x^\nu}{\partial x'^\mu} \right| \mathscr{A} \tag{2.8}$$

such that:

$$\begin{aligned}
\int \mathscr{A}' dx'^4 &= \int \mathscr{A} \left| \frac{\partial x^\nu}{\partial x'^\mu} \right| \left| \frac{\partial x'^\mu}{\partial x^\nu} \right| d^4x \\
\int \mathscr{A}' dx'^4 &= \int \mathscr{A} \left| \frac{\partial x^\nu}{\partial x'^\mu} \frac{\partial x'^\mu}{\partial x^\nu} \right| dx^4 \\
\int \mathscr{A}' dx'^4 &= \int \mathscr{A} dx^4
\end{aligned}$$

The quantities that preserve integration and by definition transform with the law of Equation 2.8 are called **scalar densities**. But as mentioned in the previous section, scalar fields are just special kind of tensors. So, one can define in general tensor densities. Inspired by Equation 2.8, tensor densities can be defined as tensors whose transformation law has an extra term, which is the Jacobian of the diffeomorphism 2.1. For example, a third rank tensor density transforms as:

$$\mathscr{T}'^{\mu\nu}_{\rho} = \left| \frac{\partial x^\epsilon}{\partial x'^\delta} \right| \frac{\partial x'^\mu}{\partial x^\kappa} \frac{\partial x'^\nu}{\partial x^\lambda} \frac{\partial x^\sigma}{\partial x'^\rho} \mathscr{T}^{\kappa\lambda}_{\sigma}$$

Densities do not appear only in integration. Consider an antisymmetric fourth rank tensor $T_{\mu\nu\rho\sigma}$. In this context antisymmetry means that a mutual exchange of any two of its indices yields a minus sign upfront. Consequently, the only value of this tensor that is not zero is $T_{1234}$, because if a same index appears twice the value is zero by antisymmetry. For example, $T_{1123} = -T_{1123} \Rightarrow T_{1123} = 0$. Moreover, again by antisymmetry, all the values of $T_{\mu\nu\rho\sigma}$ can be expressed in terms of $T_{1234}$. For example, $T_{2314} = -T_{2134} = T_{1234}$. The transformation rule for the element $T_{1234}$ is:

[4] The notation $\left| \frac{\partial x'^\mu}{\partial x^\nu} \right|$ is used for the Jacobian Determinant

$$T'_{1234} = \frac{\partial x^\mu}{\partial x'_1}\frac{\partial x^\nu}{\partial x'_2}\frac{\partial x^\rho}{\partial x'_3}\frac{\partial x^\sigma}{\partial x'_4}T_{\mu\nu\rho\sigma}$$

Because of antisymmetry this calculation can be carried out, but it is quite long and cumbersome that is why it is omitted. The result of this calculation is:

$$T'_{1234} = \left|\frac{\partial x^\mu}{\partial x'^\nu}\right| T_{1234} \tag{2.9}$$

The essence of this result is that antisymmetric 4-rank tensors can be seen as one component $T_{1234} = \mathscr{T}$ behaving like a scalar density. The inverse of this result is also true. Assume a quantity[5] $\mathscr{G}^{\mu\nu\rho\sigma}$, such that:

$$\mathscr{G}^{\mu\nu\rho\sigma} = \begin{cases} A, & \text{even permutation of } \{1,2,3,4\} \\ -A, & \text{odd permutation of } \{1,2,3,4\} \\ 0, & \text{otherwise} \end{cases} \tag{2.10}$$

where $A$ is an invariant quantity. A correct way to express the invariant property of $A$ is to consider the following transformation rule for $\mathscr{G}_{\mu\nu\rho\sigma}$.

$$\mathscr{G}^{\mu\nu\rho\sigma} = \left|\frac{\partial x^\kappa}{\partial x^\lambda}\right| \frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial x'^\nu}{\partial x^\beta}\frac{\partial x'^\rho}{\partial x^\gamma}\frac{\partial x'^\sigma}{\partial x^\delta}\mathscr{G}'^{\alpha\beta\gamma\delta} \tag{2.11}$$

In the above equation the summation implied by repeated indices give the inverse determinant, which cancels out with the other determinant upfront. So, Equation 2.11 is another way of expressing that $A$ is a diffeomorphism invariant, but also, according to what described above, says that $\mathscr{G}^{\mu\nu\rho\sigma}$ is a contravariant completely antisymmetric tensor density of the fourth rank. A special case, which is also very useful, is the case of $\epsilon^{\mu\nu\rho\sigma}$, where $A = 1$. This is usually referred to as the Levi-Civita tensor density.

$$\epsilon^{\mu\nu\rho\sigma} = \begin{cases} 1, & \text{even permutation of } \{1,2,3,4\} \\ -1, & \text{odd permutation of } \{1,2,3,4\} \\ 0, & \text{otherwise} \end{cases} \tag{2.12}$$

The Levi-Civita tensor is used to transform antisymmetric covariant tensors to antisymmetric contravariant tensor densities. For example consider an antisymmetric tensor $T_{\mu\nu}$. Then, using this tensor and the Levi-Civita tensor, two densities can be constructed.

[5]The calligraphic notation is used, because this quantity will later be identified as a density. For now it is assumed as a random quite, meaning a quaintity without rejudice for its transformation rule.

$$\mathscr{A} = \frac{1}{8}\epsilon^{\mu\nu\rho\sigma}T_{\mu\nu}T_{\rho\sigma} = T_{12}T_{34} + T_{23}T_{14} + T_{31}T_{24} \tag{2.13}$$

$$\mathscr{U}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}T_{\rho\sigma} = \epsilon^{\mu\nu12}T_{12} + \epsilon^{\mu\nu13}T_{13} + \epsilon^{\mu\nu12}T_{14} + \epsilon^{\mu\nu12}T_{23} + \epsilon^{\mu\nu24}T_{24} + \epsilon^{\mu\nu34}T_{34} \tag{2.14}$$

Equation 2.13 transforms as a scalar density, while Equation 2.14 transforms as a tensor density of the second rank. The tensor density $\mathscr{U}^{\mu\nu}$ is also antisymmetrical, as one can see by exchanging indices $\mu$ and $\nu$. The following equations show that $\mathscr{A}$ and $\mathscr{U}^{\mu\nu}$ do indeed transform as scalar and tensor densities accordingly.

$$\begin{aligned}
\mathscr{A}' &= \frac{1}{8}\epsilon'^{\mu\nu\rho\sigma}T'_{\mu\nu}T'_{\rho\sigma} \\
\mathscr{A}' &= \frac{1}{8}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\frac{\partial x'^{\mu}}{\partial x^{\alpha}}\frac{\partial x'^{\nu}}{\partial x^{\beta}}\frac{\partial x'^{\rho}}{\partial x^{\gamma}}\frac{\partial x'^{\sigma}}{\partial x^{\delta}}\epsilon^{\alpha\beta\gamma\delta}\frac{\partial x^{\alpha}}{\partial x'^{\mu}}\frac{\partial x^{\beta}}{\partial x'^{\nu}}T_{\alpha\beta}\frac{\partial x^{\gamma}}{\partial x'^{\rho}}\frac{\partial x^{\delta}}{\partial x'^{\sigma}}T_{\gamma\delta} \\
\mathscr{A}' &= \frac{1}{8}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\left(\frac{\partial x'^{\mu}}{\partial x^{\alpha}}\frac{\partial x^{\alpha}}{\partial x'^{\mu}}\right)\left(\frac{\partial x'^{\nu}}{\partial x^{\beta}}\frac{\partial x^{\beta}}{\partial x'^{\nu}}\right)\left(\frac{\partial x'^{\rho}}{\partial x^{\gamma}}\frac{\partial x^{\gamma}}{\partial x'^{\rho}}\right)\left(\frac{\partial x'^{\sigma}}{\partial x^{\delta}}\frac{\partial x^{\delta}}{\partial x'^{\sigma}}\right)\epsilon^{\alpha\beta\gamma\delta}T_{\alpha\beta}T_{\gamma\delta} \\
\mathscr{A}' &= \frac{1}{8}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\epsilon^{\alpha\beta\gamma\delta}T_{\alpha\beta}T_{\gamma\delta} \\
\mathscr{A}' &= \frac{1}{8}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\mathscr{A}
\end{aligned}$$

$$\begin{aligned}
\mathscr{U}'^{\mu\nu} &= \frac{1}{2}\epsilon'^{\mu\nu\rho\sigma}T'_{\rho\sigma} \\
\mathscr{U}'^{\mu\nu} &= \frac{1}{2}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\frac{\partial x'^{\mu}}{\partial x^{\alpha}}\frac{\partial x'^{\nu}}{\partial x^{\beta}}\frac{\partial x'^{\rho}}{\partial x^{\gamma}}\frac{\partial x'^{\sigma}}{\partial x^{\delta}}\epsilon^{\alpha\beta\gamma\delta}\frac{\partial x^{\gamma}}{\partial x'^{\rho}}\frac{\partial x^{\delta}}{\partial x'^{\sigma}}T_{\gamma\delta} \\
\mathscr{U}'^{\mu\nu} &= \frac{1}{2}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\frac{\partial x'^{\mu}}{\partial x^{\alpha}}\frac{\partial x'^{\nu}}{\partial x^{\beta}}\left(\frac{\partial x'^{\rho}}{\partial x^{\gamma}}\frac{\partial x^{\gamma}}{\partial x'^{\rho}}\right)\left(\frac{\partial x'^{\sigma}}{\partial x^{\delta}}\frac{\partial x^{\delta}}{\partial x'^{\sigma}}\right)\epsilon^{\alpha\beta\gamma\delta}T_{\gamma\delta} \\
\mathscr{U}'^{\mu\nu} &= \frac{1}{2}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\frac{\partial x'^{\mu}}{\partial x^{\alpha}}\frac{\partial x'^{\nu}}{\partial x^{\beta}}\epsilon^{\alpha\beta\gamma\delta}T_{\gamma\delta} \\
\mathscr{U}'^{\mu\nu} &= \frac{1}{2}\left|\frac{\partial x^{\kappa}}{\partial x'^{\lambda}}\right|\mathscr{U}^{\mu\nu}
\end{aligned}$$

As illustrated with the above examples the Levi-Civita tensor density can be used to connect antisymmetric contravariant densities with antisymmetric covariant tensors. This is a general result. By multiplying a covariant antisymmetric tensor with the Levi-Civita tensor density and doing the appropriate summations between the indices, the result is a contravariant antisymmetric tensor density. Intuitively, we can think that by multiplying with $\epsilon^{\mu\nu\rho\sigma}$ we can raise the index of an antisymmetric tensor. By multiplying with appropriate factors[6], the densities will have the same components as the tensors, but they will appear in different indices. Table 2.1 summarizes the connection described above.

[6]Like the factors $\frac{1}{2}$ and $\frac{1}{8}$ in Equation 2.13 and 2.14.

| Tensor | Density | Relation |
|---|---|---|
| $T$ | $\mathscr{U}^{\mu\nu\rho\sigma}$ | $\epsilon^{\mu\nu\rho\sigma}A = \mathscr{U}^{\mu\nu\rho\sigma}$ |
| $T_{\mu}$ | $\mathscr{U}^{\mu\nu\rho}$ | $\epsilon^{\mu\nu\rho\sigma}T_{\sigma} = \mathscr{U}^{\mu\nu\rho}$ |
| $T_{\mu\nu}$ | $\mathscr{U}^{\mu\nu}$ | $\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}T_{\rho\sigma} = \mathscr{U}^{\mu\nu}$ |
| $T_{\mu\nu\rho}$ | $\mathscr{U}^{\mu}$ | $\frac{1}{6}\epsilon^{\mu\nu\rho\sigma}T_{\nu\rho\sigma} = \mathscr{U}^{\mu}$ |
| $T_{\mu\nu\rho\sigma}$ | $\mathscr{U}$ | $\frac{1}{24}\epsilon^{\mu\nu\rho\sigma}T_{\mu\nu\rho\sigma} = \mathscr{U}$ |

**Table 2.1:** Relating antisymmetrical covariant tensors with antisymmetrical contravariant tensor densities.

### 2.1.4 Differentiation

Integration of scalar quantities gave rise to the new definition of densities. It is rational, then, to see what is happening when one tries to differentiate a scalar field $\phi$ over the transformed frame $x'$ up to the second order.

$$\begin{aligned}
\frac{\partial\phi}{\partial x'^{\mu}} &= \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial\phi}{\partial x^{\kappa}} \\
\frac{\partial^2\phi}{\partial x'^{\nu}\partial x'^{\mu}} &= \frac{\partial}{\partial x'^{\nu}}\left\{\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial\phi}{\partial x^{\kappa}}\right\} \\
\frac{\partial^2\phi}{\partial x'^{\nu}\partial x'^{\mu}} &= \frac{\partial}{\partial x'^{\nu}}\left\{\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\right\}\frac{\partial\phi}{\partial x^{\kappa}} + \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial}{\partial x'^{\nu}}\left\{\frac{\partial\phi}{\partial x^{\kappa}}\right\} \\
\frac{\partial^2\phi}{\partial x'^{\nu}\partial x'^{\mu}} &= \frac{\partial^2 x^{\kappa}}{\partial x'^{\mu}\partial x'^{\nu}}\frac{\partial\phi}{\partial x^{\kappa}} + \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial^2\phi}{\partial x^{\kappa}\partial x^{\nu}} \\
\frac{\partial^2\phi}{\partial x'^{\nu}\partial x'^{\mu}} &= \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial^2\phi}{\partial x^{\kappa}\partial x^{\rho}} + \frac{\partial^2 x^{\kappa}}{\partial x'^{\mu}\partial x'^{\nu}}\frac{\partial\phi}{\partial x^{\kappa}}
\end{aligned} \tag{2.15}$$

It is obvious from the last equation that the second order derivative ruins the linear and homogeneous characteristics endowed by the transformations 2.4 and 2.3. A similar result can be obtained by differentiating Equation 2.3.

$$\begin{aligned}
A'_{\mu} &= \frac{\partial x^{\nu}}{\partial x'^{\mu}}A_{\nu} \\
\frac{\partial A'_{\mu}}{\partial x'^{\kappa}} &= \frac{\partial x^{\nu}}{\partial x'^{\mu}}\frac{\partial A_{\nu}}{\partial x'^{\kappa}} + \frac{\partial^2 x^{\nu}}{\partial x'^{\mu}\partial x'^{\kappa}}A_{\nu} \\
\frac{\partial A'_{\mu}}{\partial x'^{\kappa}} &= \frac{\partial x^{\nu}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\kappa}}\frac{\partial A_{\nu}}{\partial x^{\rho}} + \frac{\partial^2 x^{\nu}}{\partial x'^{\mu}\partial x'^{\kappa}}A_{\nu}
\end{aligned} \tag{2.16}$$

From this equation it becomes clearer that the components of the derivative of a covariant vector $A_{\mu}$ do not behave like a tensor because of the second non-homogeneous term appearing on the right-hand side of the last equation. One can now think of a correspondence of our new

derivatives on the manifold with the classic definitions of derivatives of vector calculus. Be careful, though, because when differentiating on a manifold with respect to some coordinate system, we are actually differentiating on a small region of the manifold, where the manifold can be thought as locally Euclidean and taking differences (even infinitesimal) of fields has a definite meaning. That being said, let' s see this correspondence with classical vector calculus.

The gradient[7] of a scalar field $\phi$ is:

$$\nabla\phi = \frac{\partial\phi}{\partial x^{\mu}} \tag{2.17}$$

This is a covariant vector that is why it is common to write it using index notation as:

$$\nabla\phi = \partial_{\mu}\phi \tag{2.18}$$

In Equation 2.15 we have calculated the derivative of this gradient, where one additional non-homogeneous term has appeared. Because, the partial derivatives can commute the following equation holds:

$$\partial_{\mu}\partial_{\nu}\phi - \partial_{\nu}\partial_{\mu}\phi = 0 \tag{2.19}$$

Inspired by the above equation one can try to evaluate a similar equation for a covariant vector $A_{\mu}$, in order to see whether the non-homogeneous terms of Equation 2.16 cancel out.

$$\begin{aligned}
&\frac{\partial A'_{\nu}}{\partial x'^{\mu}} - \frac{\partial A'_{\mu}}{\partial x'^{\mu}} = \\
&= \left(\frac{\partial^2 x^k}{\partial x'^{\mu}\partial x'^{\nu}}A_{\kappa} + \frac{\partial x^k}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial A_{\kappa}}{\partial x^{\rho}}\right) - \left(\frac{\partial^2 x^k}{\partial x'^{\nu}\partial x'^{\mu}}A_{\kappa} + \frac{\partial x^k}{\partial x'^{\nu}}\frac{\partial x^{\rho}}{\partial x'^{\mu}}\frac{\partial A_{\kappa}}{\partial x^{\rho}}\right) = \\
&= \frac{\partial x^k}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial A_k}{\partial x^{\rho}} - \frac{\partial x^k}{\partial x'^{\nu}}\frac{\partial x^{\rho}}{\partial x'^{\mu}}\frac{\partial A_{\kappa}}{\partial x^{\rho}} = \\
&= \frac{\partial x^k}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial A_k}{\partial x^{\rho}} - \frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial A_{\rho}}{\partial x^{\kappa}}
\end{aligned}$$

In the last equality the dummy indices $\kappa, \rho$ of the second term can be mutually exchanged as $\kappa \leftrightarrow \rho$. The final result is:

$$\frac{\partial A'_{\nu}}{\partial x'^{\mu}} - \frac{\partial A'_{\mu}}{\partial x'^{\nu}} = \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\left(\frac{\partial A_{\kappa}}{\partial x^{\rho}} - \frac{\partial A_{\rho}}{\partial x^{\kappa}}\right) \tag{2.20}$$

This resulting equation shows that the quantity $\partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ transforms as a second rank tensor. This is the equivalent of the curl.

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \tag{2.21}$$

If the curl of a vector happens to be zero then $\partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} = 0$, which is true if, according to Equation 2.19, the vector field shall be replaced by the gradient of a scalar field. This is a

[7] with 4 components, because we are working on a 4-dimensional manifold

well known result from vector calculus and its retrieval with our new way of expressing curl increases its validity.

Last but not least, by exchanging the indices $\mu$ and $\nu$ the tensor becomes:

$$\begin{aligned} F_{\nu\mu} &= \partial_\nu A_\mu - \partial_\mu A_\nu \\ F_{\nu\mu} &= -\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) \\ F_{\nu\mu} &= -F_{\mu\nu} \end{aligned}$$

which shows that the tensor $F_{\mu\nu}$ is antisymmetrical. Now, having defined the tensorial character of $F_{\mu\nu}$, one could try to differentiate this new tensor.

$$F'_{\mu\nu} = \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}F_{\kappa\rho} \tag{2.22}$$

$$\frac{\partial F'_{\mu\nu}}{\partial x'^\sigma} = \frac{\partial^2 x^\kappa}{\partial x'^\sigma \partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial^2 x^\rho}{\partial x'^\sigma \partial x'^\nu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \tag{2.23}$$

As in the case of vectors, just by differentiating the antisymmetric tensor, the result is not a tensor, which means that it does not transform homogeneously. More specifically, the homogeneous part is the last term of Equation 2.23. So, the goal is to form a proper differentiation procedure that will make the other two terms cancel, as it happened in the case of vectors with Equation 2.20. As in the case of vectors, a good idea is to permute the indices in Equation 2.23.

$$\begin{aligned} \frac{\partial F'_{\mu\nu}}{\partial x'^\sigma} &= \frac{\partial^2 x^\kappa}{\partial x'^\sigma \partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial^2 x^\rho}{\partial x'^\sigma \partial x'^\nu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\ \frac{\partial F'_{\sigma\mu}}{\partial x'^\nu} &= \frac{\partial^2 x^\kappa}{\partial x'^\nu \partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\mu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial^2 x^\rho}{\partial x'^\nu \partial x'^\mu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\ \frac{\partial F'_{\nu\sigma}}{\partial x'^\mu} &= \frac{\partial^2 x^\kappa}{\partial x'^\mu \partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\sigma}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\sigma}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \end{aligned}$$

The terms appearing in the above equations are similar, except for the indices $\kappa, \rho$. But, taking advantage of the antisymmetric character of $F_{\mu\nu}$, this can be fixed by doing the change $\kappa \leftrightarrow \rho$ in the second equation. So, the three of them now read:

$$\begin{aligned} \frac{\partial F'_{\mu\nu}}{\partial x'^\sigma} &= \frac{\partial^2 x^\kappa}{\partial x'^\sigma \partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial^2 x^\rho}{\partial x'^\sigma \partial x'^\nu}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\ \frac{\partial F'_{\sigma\mu}}{\partial x'^\nu} &= -\frac{\partial^2 x^\rho}{\partial x'^\nu \partial x'^\sigma}\frac{\partial x^\kappa}{\partial x'^\mu}F_{\kappa\rho} - \frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial^2 x^\kappa}{\partial x'^\nu \partial x'^\mu}F_{\kappa\rho} - \frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\ \frac{\partial F'_{\nu\sigma}}{\partial x'^\mu} &= \frac{\partial^2 x^\kappa}{\partial x'^\mu \partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\sigma}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\sigma}F_{\kappa\rho} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \end{aligned}$$

Adding them altogether, the first term of the second equation cancels with the second term of the first, while its second term cancels with the first term of the third equation. At the moment the terms containing the differentiated tensor all stay.

$$\frac{\partial F'_{\mu\nu}}{\partial x'^{\sigma}} + \frac{\partial F'_{\sigma\mu}}{\partial x'^{\nu}} + \frac{\partial F'_{\nu\sigma}}{\partial x'^{\mu}} = \frac{\partial^2 x^{\kappa}}{\partial x'^{\sigma}\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}F_{\kappa\rho} + \frac{\partial x^{\kappa}}{\partial x'^{\nu}}\frac{\partial^2 x^{\rho}}{\partial x'^{\mu}\partial x'^{\sigma}}F_{\kappa\rho} + \\ + \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} - \frac{\partial x^{\rho}}{\partial x'^{\sigma}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\lambda}}{\partial x'^{\nu}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} + \frac{\partial x^{\kappa}}{\partial x'^{\nu}}\frac{\partial x^{\rho}}{\partial x'^{\sigma}}\frac{\partial x^{\lambda}}{\partial x'^{\mu}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}}$$

The first two terms can easily mutually cancel by doing again the change $\kappa \leftrightarrow \rho$ and using the antisymmetry $F_{\kappa\rho} = -F_{\rho\kappa}$, meaning that the only surviving terms are the ones containing the derivative of $F_{\mu\nu}$. The main goal of this calculation is to show that the the sum of these derivatives transforms tensorially. So, in order to see whether the remaining three terms reduce to one it is wise to rename the sum of partial derivatives to $T_{\mu\nu\sigma}$, without prejudice to its tensorial nature.

$$T_{\mu\nu\sigma} = \frac{\partial F_{\mu\nu}}{\partial x^{\sigma}} + \frac{\partial F_{\sigma\mu}}{\partial x^{\nu}} + \frac{\partial F_{\nu\sigma}}{\partial x^{\mu}} \tag{2.24}$$

$$T'_{\mu\nu\sigma} = \frac{\partial F'_{\mu\nu}}{\partial x'^{\sigma}} + \frac{\partial F'_{\sigma\mu}}{\partial x'^{\nu}} + \frac{\partial F'_{\nu\sigma}}{\partial x'^{\mu}} \tag{2.25}$$

$$T'_{\mu\nu\sigma} = \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} + \frac{\partial x^{\kappa}}{\partial x'^{\sigma}}\frac{\partial x^{\rho}}{\partial x'^{\mu}}\frac{\partial x^{\lambda}}{\partial x'^{\nu}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} + \frac{\partial x^{\kappa}}{\partial x'^{\nu}}\frac{\partial x^{\rho}}{\partial x'^{\sigma}}\frac{\partial x^{\lambda}}{\partial x'^{\mu}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} \tag{2.26}$$

The goal is to show that $T_{\mu\nu\sigma}$, defined in Equation 2.24, transforms as a third rank covariant tensor. That would be the transformation described by Equations 2.27 and 2.28 bellow.

$$T'_{\mu\nu\sigma} = \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}T_{\kappa\rho\lambda} \tag{2.27}$$

$$T'_{\mu\nu\sigma} = \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\left(\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} + \frac{\partial F_{\lambda\kappa}}{\partial x^{\rho}} + \frac{\partial F_{\rho\lambda}}{\partial x^{\kappa}}\right) \tag{2.28}$$

To show that $T_{\mu\nu\sigma}$ does indeed transforms as such, Equation 2.26 can be rewritten by renaming the dummy indices leading, at fist, to Equation 2.29 and lastly to Equation 2.30, which, in a closer look, is the same with Equation 2.28 and that proves that the quantity of Equation 2.24, where $F_{\mu\nu}$ is any covariant antisymmetric tensor[8] of rank 2, transforms as a third rank covariant tensor.

$$T'_{\mu\nu\sigma} = \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} + \frac{\partial x^{\rho}}{\partial x'^{\sigma}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\lambda}}{\partial x'^{\nu}}\frac{\partial F_{\rho\kappa}}{\partial x^{\lambda}} + \frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\kappa}}{\partial x'^{\sigma}}\frac{\partial x^{\lambda}}{\partial x'^{\mu}}\frac{\partial F_{\rho\kappa}}{\partial x^{\lambda}} \tag{2.29}$$

$$T'_{\mu\nu\sigma} = \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\frac{\partial F_{\kappa\rho}}{\partial x^{\lambda}} + \frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial F_{\lambda\kappa}}{\partial x^{\rho}} + \frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x^{\lambda}}{\partial x'^{\sigma}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial F_{\rho\lambda}}{\partial x^{\kappa}} \tag{2.30}$$

[8]Note that Equation 2.21 was not used in this proof.

The object $T_{\mu\nu\sigma}$ is also fully antisymmetrical in its indices as it can be seen by exchanging the indices $\mu \leftrightarrow \nu$ and $\nu \leftrightarrow \sigma$.

$$
\begin{aligned}
T'_{\nu\mu\sigma} &= \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} + \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\
T'_{\nu\mu\sigma} &= -\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} - \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} - \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\
T'_{\nu\mu\sigma} &= -T'_{\mu\nu\sigma}
\end{aligned}
$$

$$
\begin{aligned}
T'_{\mu\sigma\nu} &= \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} + \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\
T'_{\mu\sigma\nu} &= -\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} - \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} - \frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial F_{\kappa\rho}}{\partial x^\lambda} \\
T'_{\mu\sigma\nu} &= -T'_{\mu\nu\sigma}
\end{aligned}
$$

Now is the turn for tensors of the fourth rank $A_{\mu\nu\sigma\tau}$, formed as derivatives of third rank antisymmetric covariant tensors. The derivative of the last transforms as follows:

$$
\begin{aligned}
\frac{\partial T'_{\mu\nu\sigma}}{\partial x'^\tau} = &\frac{\partial^2 x^\kappa}{\partial x'^\sigma \partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\sigma}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial^2 x^\rho}{\partial x'^\nu \partial x'^\tau}\frac{\partial x^\lambda}{\partial x'^\sigma}T_{\kappa\rho\lambda} \\
&+ \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial^2 x^\lambda}{\partial x'^\sigma \partial x'^\tau}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\frac{\partial x^\lambda}{\partial x'^\sigma}\frac{\partial x^\alpha}{\partial x'^\tau}\frac{\partial T_{\kappa\rho\lambda}}{\partial x^\alpha}
\end{aligned} \tag{2.31}
$$

Following the same recipe as in tensors of the third rank, one can form the following derivatives by permuting the indices of the previous equation.

$$
\begin{aligned}
\frac{\partial T'_{\tau\mu\nu}}{\partial x'^\sigma} = &\frac{\partial^2 x^\kappa}{\partial x'^\sigma \partial x'^\tau}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\tau}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\nu}T_{\kappa\rho\lambda} \\
&+ \frac{\partial x^\kappa}{\partial x'^\tau}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial^2 x^\lambda}{\partial x'^\sigma \partial x'^\nu}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\tau}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial x^\alpha}{\partial x'^\sigma}\frac{\partial T_{\kappa\rho\lambda}}{\partial x^\alpha}
\end{aligned} \tag{2.32}
$$

$$
\begin{aligned}
\frac{\partial T'_{\sigma\tau\mu}}{\partial x'^\nu} = &\frac{\partial^2 x^\kappa}{\partial x'^\sigma \partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\tau}\frac{\partial x^\lambda}{\partial x'^\mu}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial^2 x^\rho}{\partial x'^\nu \partial x'^\tau}\frac{\partial x^\lambda}{\partial x'^\mu}T_{\kappa\rho\lambda} \\
&+ \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\tau}\frac{\partial^2 x^\lambda}{\partial x'^\mu \partial x'^\nu}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\sigma}\frac{\partial x^\rho}{\partial x'^\tau}\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\alpha}{\partial x'^\nu}\frac{\partial T_{\kappa\rho\lambda}}{\partial x^\alpha}
\end{aligned} \tag{2.33}
$$

$$
\begin{aligned}
\frac{\partial T'_{\nu\sigma\tau}}{\partial x'^\mu} = &\frac{\partial^2 x^\kappa}{\partial x'^\nu \partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\tau}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\tau}T_{\kappa\rho\lambda} \\
&+ \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial^2 x^\lambda}{\partial x'^\mu \partial x'^\tau}T_{\kappa\rho\lambda} + \frac{\partial x^\kappa}{\partial x'^\nu}\frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial x^\lambda}{\partial x'^\tau}\frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial T_{\kappa\rho\lambda}}{\partial x^\alpha}
\end{aligned} \tag{2.34}
$$

But, because there are three derivatives before each tensor term in each equation, adding the four equations altogether will not lead to cancelling all the non-homogeneous terms. Instead, one can observe that, in order to cancel the first term of Equation 2.31 with the third term of the fourth equation, the last must appear with a minus sign. The same is true for the third term of Equation 2.31, which will cancel with the first term of the second equation, if the last is multiplied with a minus sign. As for the second term of the first equation, this will cancel by addition with the second term of the third equation. These observations do give the correct formula for forming a derivative that will actually transform as a covariant four rank tensor. The correct way to form this derivative, is to add all the possible permutations, but multiply with a minus sign when the permutation is odd. So, the following derivative transforms as a 4-rank covariant tensor, where the symbol $(-1)^{\dagger}$ is used to encode the previous remark about the sign.

$$A_{\mu\nu\sigma\tau} = \sum (-1)^{\dagger} \frac{\partial T_{\mu\nu\sigma}}{\partial x^{\tau}} = \frac{\partial T_{\mu\nu\sigma}}{\partial x^{\tau}} - \frac{\partial T_{\tau\mu\nu}}{\partial x^{\sigma}} + \frac{\partial T_{\sigma\tau\mu}}{\partial x^{\nu}} - \frac{\partial T_{\nu\sigma\tau}}{\partial x^{\mu}} \tag{2.35}$$

### 2.1.5 Connection and Covariant Derivatives

There are two famous ways to introduce the notion of covariant derivatives. The first one is more suitable for algebraic thinkers and has its roots on Equation 2.16, where it is shown that the derivative of a vector does not transform as a vector itself. In order to achieve such a homogeneous transformation, the derivative[9] should absorb in some way the non-homogeneous term.

At first, a very common convention in the literature is the following, concerning the derivative of a covariant vector.

$$A_{\mu,\nu} = \frac{\partial A_{\mu}}{\partial x^{\nu}} \tag{2.36}$$

Of course, this convention also works for covariant tensors, for example:

$$T_{\mu\nu,\sigma} = \frac{\partial T_{\mu\nu}}{\partial x^{\sigma}} \tag{2.37}$$

Following the same notation, the new derivative that absorbs the non-homogeneous term would be:

$$A_{\mu;\nu} = A_{\mu,\nu} - A_{\kappa}\Gamma^{\kappa}_{\mu\nu} \tag{2.38}$$

The term denoted by $\Gamma$ above is responsible for absorbing the non-homogeneous term that appears in the derivative of a covariant vector. The demand for this derivative is to transform homogeneously, meaning as a second rank tensor. Demanding this rule of transformation, the transformation formula for the $\Gamma$-term will arise naturally.

[9] We are actually looking for such a derivative.

$$A'_{\mu;\nu} = \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}A_{\kappa;\lambda} \tag{2.39}$$

$$A'_{\mu,\nu} - A'_\rho \Gamma'^\rho_{\mu\nu} = \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}A_{\kappa,\lambda} - \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}A_\rho \Gamma^\rho_{\kappa\lambda} \tag{2.40}$$

Substituting the transformation of simple derivatives ($A'_{\mu,\nu}$) from Equation 2.16 and the transformation of the covariant vectors ($A'_\rho$) from Equation 2.6, Equation 2.40 becomes:

$$\frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}A_{\kappa,\lambda} + \frac{\partial^2 x^\kappa}{\partial x'^\mu \partial x'^\nu}A_\kappa - \frac{\partial x^\kappa}{\partial x'^\rho}A_\kappa \Gamma'^\rho_{\mu\nu} = \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}A_{\kappa,\lambda} - \frac{\partial x^\kappa}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}A_\rho \Gamma^\rho_{\kappa\lambda} \tag{2.41}$$

$$\left(\frac{\partial x^\kappa}{\partial x'^\rho}\Gamma'^\rho_{\mu\nu} - \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}\Gamma^\kappa_{\rho\lambda} - \frac{\partial^2 x^\kappa}{\partial x'^\mu \partial x'^\nu}\right)A_\kappa = 0 \tag{2.42}$$

For the last equation to be true for every covariant vector $A_\kappa$, the first factor must be equal to zero. After some displacements and a multiplication by the inverse Jacobian matrix, the transformation law of the $\Gamma$-term arises.

$$\Gamma'^\sigma_{\mu\nu} = \frac{\partial x'^\sigma}{\partial x^\kappa}\frac{\partial^2 x^\kappa}{\partial x'^\mu \partial x'^\nu} + \frac{\partial x'^\sigma}{\partial x^\kappa}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\lambda}{\partial x'^\nu}\Gamma^\kappa_{\rho\lambda} \tag{2.43}$$

Equation 2.43 is really important because it shows how the $\Gamma$- term should transform, in order for the new derivative of Equation 2.38 to transform as a second rank tensor.

Equation 2.38 is known as the **Covariant Derivative** of a covariant vector. The second way of introducing the covariant derivative might be more appropriate for a physical thinker, because it requires to visualize how would a vector move if it were restricted to a curved space.

The easiest way to visualize this is, as always, to consider the simplest case. In this situation, the simplest curved background is a curved line. First, consider two points, $A, B$, linked with a straight line. Then, consider a vector anchored at the point $A$. If one wants to transfer this vector to the point $B$, but simultaneously the vector has to move through the straight line, then it means that the angle of the vector with the line must remain the same, through all the positions the vector shall take during this transfer.

On the other hand consider that the points $A, B$ are lying on a (non-straight) curve. Then, to transfer a vector from $A$ to $B$ the angle that needs to stay intact is that between the vector at each point of the transfer and the tangent line of the curve at this point. Figure 2.1 might help in visualization.

Therefore, in the case of the straight line, the vector remains unchanged when comparing the original vector, attached to the point $A$, with the final vector in the point $B$. In contrast, in the case of the curved line, the vector changes just by "following" the line. There is no external force that changes the vector; only the geometry of its underline space is responsible for its change.

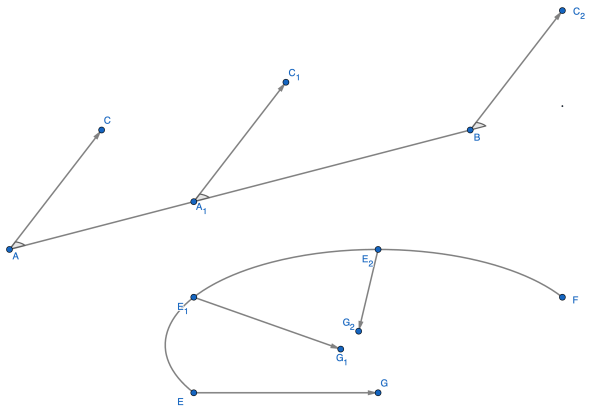



**Figure 2.1:** Visualization of Parallel Transport for a vector. The first case considers transporting through a straight line, which corresponds to the Euclidean space, while in the second case the vector is transfer through a curved line. In the second case, the vector points in a different direction after being transfered.

The job of the derivative is to quantify the change of functions between two points. But, in the case of curved backgrounds there will be two different kinds of changes. One coming from external forces/effects and another one only because the function is defined on a curved manifold.

It is a good idea, to define a new type of derivative that will absorb the change of the underlying geometry of the manifold and consider only the changes accounted to external parameters. This is important for physics where the physicist wants to distinguish the changes causes by "forces" to describe physical laws. That is exactly the job of the covariant derivative defined previously.

The difference between two vectors $A^\mu$, $A^\mu + dA^\mu$ defined in two infinitesimally close points $x^\rho$, $x^\rho + dx^\rho$ is:

$$dA^\mu = \frac{\partial A^\mu}{\partial x^\nu} dx^\nu \tag{2.44}$$

This difference contains the total change of the vector, in the sense that was discussed above. In that change there exist the change due to geometry, for which the symbol $\delta A^\mu$ is used. In other words, $\delta A^\mu$ is the change that is considered as no-change when transferring the vector from one point to another. This new quantity can only depend on the original vector $\delta A^\mu$ and the coordinates' one-form $dx^\nu$. The simplest and most compact way to write this dependence is:

$$\delta A^\mu = -\Gamma^\mu_{\kappa\nu} A^\kappa dx^\nu \tag{2.45}$$

To prove that this $\Gamma$-term used here is the same as the one introduced for the covariant derivative, Equation 2.43 must be reproduced using the notion of parallel transport. As mentioned above, the change $\delta A^\mu$ measures only the original change of the vector, without

taking into account the properties of the underlying manifold. As a result, this change must be the same if the local coordinate system changes, by the diffeomorphism 2.1.

$$\begin{aligned}
\delta A'^{\mu} &= (\delta A^{\mu}) \\
\delta\left(\frac{\partial x'^{\mu}}{\partial x^{\nu}} A^{\nu}\right) &= (-\Gamma^{\mu}_{\kappa\nu} A^{\kappa} dx^{\nu})' \\
\delta\left(\frac{\partial x'^{\mu}}{\partial x^{\nu}}\right) A^{\nu} + \frac{\partial x'^{\mu}}{\partial x^{\nu}} \delta A^{\nu} &= -\Gamma'^{\mu}_{\kappa\nu} A'^{\kappa} dx'^{\nu} \\
\frac{\partial^2 x'^{\mu}}{\partial x^{\nu} \partial x^{\kappa}} A^{\nu} dx^{\kappa} - \frac{\partial x'^{\mu}}{\partial x^{\nu}} \Gamma^{\nu}_{\kappa\sigma} A^{\kappa} dx^{\sigma} &= -\Gamma'^{\mu}_{\kappa\nu} \frac{\partial x'^{\kappa}}{\partial x^{\rho}} A^{\rho} \frac{\partial x'^{\nu}}{\partial x^{\sigma}} dx^{\sigma} \\
\left(\Gamma'^{\mu}_{\kappa\nu} \frac{\partial x'^{\kappa}}{\partial x^{\rho}} \frac{\partial x'^{\nu}}{\partial x^{\sigma}} - \frac{\partial x'^{\mu}}{\partial x^{\nu}} \Gamma^{\nu}_{\rho\sigma} - \frac{\partial^2 x'^{\mu}}{\partial x^{\rho} \partial x^{\sigma}}\right) A^{\rho} dx^{\sigma} &= 0
\end{aligned}$$

The vector $A^{\rho}$ was chosen arbitrarily, meaning that the last equation must hold for every possible vector $A^{\rho}$. After changing some dummy indices and doing a little algebraic manipulation the resulting equation is the following.

$$\Gamma'^{\sigma}_{\mu\nu} = \frac{\partial x'^{\sigma}}{\partial x^{\kappa}} \frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial x^{\lambda}}{\partial x'^{\nu}} \Gamma^{\kappa}_{\rho\lambda} - \frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial x^{\kappa}}{\partial x'^{\nu}} \frac{\partial^2 x'^{\sigma}}{\partial x^{\rho} \partial x^{\kappa}} \tag{2.46}$$

Equation 2.46 only differs with Equation 2.43 in the non-homogeneous term, but as one can see from the calculations following this paragraph there is no actual difference. Equation 2.46 is just another way of expressing Equation 2.43. It is obvious from this rule of transformation, that the $\Gamma$-term is not a tensor.

$$\begin{aligned}
\frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial x^{\kappa}}{\partial x'^{\nu}} \frac{\partial^2 x'^{\sigma}}{\partial x^{\rho} \partial x^{\kappa}} &= \frac{\partial x^{\kappa}}{\partial x'^{\nu}} \frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial}{\partial x^{\rho}} \left\{\frac{\partial x'^{\sigma}}{\partial x^{\kappa}}\right\} \\
&= \frac{\partial x^{\kappa}}{\partial x'^{\nu}} \frac{\partial}{\partial x'^{\mu}} \left\{\frac{\partial x'^{\sigma}}{\partial x^{\kappa}}\right\} \\
&= \frac{\partial}{\partial x'^{\mu}} \left\{\frac{\partial x^{\kappa}}{\partial x'^{\nu}} \frac{\partial x'^{\sigma}}{\partial x^{\kappa}}\right\} - \frac{\partial x'^{\sigma}}{\partial x^{\kappa}} \frac{\partial}{\partial x'^{\mu}} \left\{\frac{\partial x^{\kappa}}{\partial x'^{\nu}}\right\} \\
&= \frac{\partial}{\partial x'^{\mu}} \{\delta^{\sigma}_{\nu}\} - \frac{\partial x'^{\sigma}}{\partial x^{\kappa}} \frac{\partial}{\partial x'^{\mu}} \left\{\frac{\partial x^{\kappa}}{\partial x'^{\nu}}\right\} \\
\frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial x^{\kappa}}{\partial x'^{\nu}} \frac{\partial^2 x'^{\sigma}}{\partial x^{\rho} \partial x^{\kappa}} &= -\frac{\partial x'^{\sigma}}{\partial x^{\kappa}} \frac{\partial x^{\kappa}}{\partial x'^{\mu} x'^{\nu}}
\end{aligned}$$

In conclusion, in order to compare a vector in two different points on a manifold, one needs to calculate its covariant derivative, which only takes into consideration the "true" change of the vector. The covariant derivative is given by Equation 2.38. In this relation, the term labeled with the letter $\Gamma$ is responsible for compensating the inhomogeneous transformation of

a simple derivative [10] This term is often referred to as **affine connection**. The connection is also used to allow the parallel transport of a vector from one point of the manifold to another. This is done by using Equation 2.45, which is the original change of the vector.

$$A^{\mu} + \delta A^{\mu} = A^{\mu} - \Gamma^{\mu}_{\kappa\nu} A^{\kappa} dx^{\nu} \tag{2.47}$$

Equation 2.47 shows the parallel transported vector $A^{\mu}$, can be used to define the covariant derivative as the following limit.

$$\begin{aligned} A^{\mu}_{;\nu} &= \frac{(A^{\mu} + dA^{\mu}) - (A^{\mu} + \delta A^{\mu})}{dx^{\nu}} \\ &= \frac{dA^{\mu} - \delta A^{\mu}}{dx^{\nu}} \\ &= \frac{dA^{\mu} + \Gamma^{\mu}_{\kappa\rho} A^{\kappa} dx^{\rho}}{dx^{\nu}} \\ &= \frac{dA^{\mu}}{dx^{\nu}} + \Gamma^{\mu}_{\kappa\rho} A^{\kappa} \delta^{\rho}_{\nu} \\ A^{\mu}_{;\nu} &= A^{\mu}{}_{,\nu} + \Gamma^{\mu}_{\kappa\nu} A^{\kappa} \end{aligned} \tag{2.48}$$

Equation 2.48 is the covariant derivative of a contravariant vector, which differs from the covariant derivative of a covariant vector only by the plus sign appearing in front of the connection. This relation will be validated in the next section, where it will be derived only by the covariant derivative of the covariant vector and some of its consequences.

Of course, the combinations defined as frame invariant in Section 2.1.4 are the covariant derivatives of themselves. This is easily shown by just replacing the ordinary derivative with a covariant one. As a matter of fact, one shall not forget that they were initially defined to absorb the non-homogeneous term of the derivative.

### 2.1.6 Covariant Derivative of Tensors

First of all, Equation 2.38 was given the name derivative in a naive way. More specifically, to characterize an operation with the name derivative, this operation must be linear in terms of itself and of its input. It should also have a Leibniz rule for an input product. Therefore, the appropriate first step, in order to generalize this new derivative, is to show that it is actually a derivative. A derivative is linear with respect to the field of real numbers and also possess a Leibniz rule for the product of two functions.

[10] see the second term of Equation 2.16.

Consider two covariant vectors $A_\mu$, $B_\mu$ and two real numbers $a, b$.

$$\begin{aligned}
(aA_\mu + bB_\mu)_{;\nu} &= (aA_\mu + bB_\mu)_{,\nu} - \Gamma^\kappa_{\mu\nu}(aA_\kappa + bB_\kappa) \\
&= aA_{\mu,\nu} + bB_{\mu,\nu} - a\Gamma^\kappa_{\mu\nu}A_\kappa - b\Gamma^\kappa_{\mu\nu}B_\kappa \\
&= \left(aA_{\mu,\nu} - \Gamma^\kappa_{\mu\nu}aA_\kappa\right) + \left(bB_{\mu,\nu} - \Gamma^\kappa_{\mu\nu}bB_\kappa\right) \\
(aA_\mu + bB_\mu)_{;\nu} &= aA_{\mu;\nu} + bB_{\mu;\nu}
\end{aligned}$$

This proves the linear property of the derivative. As for the Leibniz rule, it will be imposed as a first step and then it will be proved acceptable by its results. More specifically, it will be used to calculate the covariant derivative of a contravariant vector. If this calculation matches Equation 2.48 then the Leibniz rule is accepted as correct.

The proof starts from the fact that $A_\mu B^\mu$ is a scalar, meaning that its covariant derivative must be the same to its ordinary derivative.

$$\begin{aligned}
(A_\mu B^\mu)_{;\nu} &= (A_\mu B^\mu)_{,\nu} \\
A_{\mu;\nu}B^\mu + A_\mu B^\mu_{;\nu} &= A_{\mu,\nu}B^\mu + A_\mu B^\mu_{,\nu} \\
A_{\mu,\nu}B^\mu - A_\kappa\Gamma^\kappa_{\mu\nu}B^\mu + A_\mu B^\mu_{;\nu} &= A_{\mu,\nu}B^\mu + A_\mu B^\mu_{,\nu} \\
A_\kappa\left(B^\kappa_{;\nu} - \Gamma^\kappa_{\mu\nu}B^\mu - B^\kappa_{,\nu}\right) &= 0 \\
B^\kappa_{;\nu} &= B^\kappa_{,\nu} + \Gamma^\kappa_{\mu\nu}B^\mu
\end{aligned}$$

This last equation is exactly the same with Equation 2.48. This concludes the proof that the covariant derivative does indeed posses a Leibniz rule.

A very important aspect of this new derivative that allows it to be defined for all the objects introduced in section 2.1.2 is that if a covariant derivative is equal to zero in a certain frame then it would remain zero for every possible choice of frame. Of course, the tensorial/homogeneous kind of transformation appearing in Equation 2.39 is responsible for this fact. So, to show that a quantity is invariant over a manifold, one shall just demand that its covariant derivative is zero.

The previous procedure used to validate Leibniz rule can be generalized to define the covariant derivative of every tensor. More specifically, an invariant product is constructed by multiplying the tensor with the appropriate number of covariant and contravariant vectors and then the derivative of this invariant must be equal to its covariant derivative. As an example the derivative of a mixed tensor will be calculated.

$$\begin{aligned}
(T^\mu_\nu A_\mu B^\nu)_{;\rho} &= (T^\mu_\nu A_\mu B^\nu)_{,\rho} \\
T^\mu_{\nu;\rho}A_\mu B^\nu + T^\mu_\nu A_{;\mu}B^\nu + T^\mu_\nu A_\mu B^\nu_{;\rho} &= T^\mu_{\nu,\rho}A_\mu B^\nu + T^\mu_\nu A_{,\mu}B^\nu + T^\mu_\nu A_\mu B^\nu_{,\rho} \\
T^\mu_{\nu;\rho}A_\mu B^\nu + T^\mu_\nu A_{,\mu}B^\nu - T^\mu_\nu\Gamma^\kappa_{\mu\rho}A_\kappa B^\nu + T^\mu_\nu A_\mu B^\nu_{,\rho} + T^\mu_\nu A_\mu\Gamma^\nu_{\kappa\rho}B^\kappa &= T^\mu_{\nu,\rho}A_\mu B^\nu + T^\mu_\nu A_{,\mu}B^\nu + T^\mu_\nu A_\mu B^\nu_{,\rho} \\
A_\kappa B^\rho\left(T^\kappa_{\nu;\rho} - T^\mu_\nu\Gamma^\kappa_{\mu\rho} - T^\kappa_\sigma\Gamma^\sigma_{\nu\rho} - T^\kappa_{\nu,\rho}\right) &= 0
\end{aligned}$$

The last equation gives the covariant derivative of the mixed tensor $T^{\mu}_{\nu}$.

$$T^{\mu}_{\nu;\rho} = T^{\mu}_{\nu,\rho} + T^{\kappa}_{\nu}\Gamma^{\mu}_{\kappa\rho} - T^{\mu}_{\kappa}\Gamma^{\kappa}_{\nu\rho} \tag{2.49}$$

### 2.1.7 Geodesics and the Metric Tensor

The purpose of this subsection is to investigate if there is a natural way to compare lengths on a manifold, having only introduced the connection $\Gamma$. That is without any reference to a metric tensor.

To investigate this, assume a curve on the spacetime manifold $\gamma = \gamma^{\mu} \in \mathcal{M}$[11], which is described by a real parameter $\lambda \in I \subset \mathbb{R}$. The derivative of this curve is the velocity at every point on the curve, or more rigorously a vector that lies on the tangent plane at every point on the curve. Of course, the curve lies entirely on the manifold, therefore to compare the velocity vector the notion of parallel transport of section 2.1.5 is needed.

The idea is to assume that by transferring the velocity vector between two neighboring points on the curve the vector will remain on the tangent plane. The same will happen if the vector is parallel transported to the neighboring point. So, the parallel transported vector would be parallel to the regular transported one. This is desrcibed by the following equation.

$$\begin{aligned} K\left[\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda} + d\left(\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda}\right)\right] &= \frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda} + \delta\left(\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda}\right) \\ K\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda} + K\frac{\mathrm{d}^2\gamma^{\mu}}{\mathrm{d}\lambda^2}d\lambda &= \frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda} - \Gamma^{\mu}_{\kappa\nu}\frac{\mathrm{d}\gamma^{\kappa}}{\mathrm{d}\lambda}\frac{\mathrm{d}\gamma^{\nu}}{\mathrm{d}\lambda}d\lambda \\ (1-K)\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda} &= K\frac{\mathrm{d}^2\gamma^{\mu}}{\mathrm{d}\lambda^2}d\lambda + \Gamma^{\mu}_{\kappa\nu}\frac{\mathrm{d}\gamma^{\kappa}}{\mathrm{d}\lambda}\frac{\mathrm{d}\gamma^{\nu}}{\mathrm{d}\lambda}d\lambda \end{aligned}$$

Taking a look at this equation, the right hand side is an one-form with respect to the differential $d\lambda$, while the left hand side is just a derivative. To fix this inconsistency the term $1-K$ must be also an one-form, which generally can be written as $1-K = \phi(\lambda)d\lambda$. Substituting in the last equation, the result is:

$$\phi(\lambda)\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda}d\lambda = \frac{\mathrm{d}^2\gamma^{\mu}}{\mathrm{d}\lambda^2}d\lambda - \phi(\lambda)\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda}d\lambda^2 + \Gamma^{\mu}_{\kappa\nu}\frac{\mathrm{d}\gamma^{\kappa}}{\mathrm{d}\lambda}\frac{\mathrm{d}\gamma^{\nu}}{\mathrm{d}\lambda}d\lambda$$

First, by doing a first order approximation the middle term of the right hand side vanishes. After that, the equation is just an equality between one-forms. The condition that must hold is the following:

$$\phi(\lambda)\frac{\mathrm{d}\gamma^{\mu}}{\mathrm{d}\lambda} = \frac{\mathrm{d}^2\gamma^{\mu}}{\mathrm{d}\lambda^2} + \Gamma^{\mu}_{\kappa\nu}\frac{\mathrm{d}\gamma^{\kappa}}{\mathrm{d}\lambda}\frac{\mathrm{d}\gamma^{\nu}}{\mathrm{d}\lambda} \tag{2.50}$$

[11] Mathematically the correct expression is $\gamma^{\mu} = x^{\mu} \circ \gamma$, where $x^{\mu}$ is the local coordinate system, near each point of the curve.

Equation 2.50 is the general form of the geodesics equation. A simpler and more intuitive way of expressing the above equation is by changing the parametrization of the curve $\gamma^\mu$ for $\lambda$ to $s(\lambda)$. Using the chain rule the geodescis equation can be written as:

$$\left(s'(\lambda)\right)^2 \frac{\mathrm{d}^2\gamma^\mu}{\mathrm{d}s(\lambda)^2} + \left(s'(\lambda)\right)^2 \Gamma^\mu_{\kappa\nu} \frac{\mathrm{d}\gamma^\kappa}{\mathrm{d}s(\lambda)} \frac{\mathrm{d}\gamma^\nu}{\mathrm{d}s(\lambda)} = \left(\phi(\lambda)s'(\lambda) - s''(\lambda)\right) \frac{\mathrm{d}\gamma^\mu}{\mathrm{d}s(\lambda)}$$

By demanding for the new parametrization to fulfill the equation $\phi s' = s''$ the geodesics equation takes the form:

$$\frac{\mathrm{d}^2\gamma^\mu}{\mathrm{d}s^2} + \Gamma^\mu_{\kappa\nu} \frac{\mathrm{d}\gamma^\kappa}{\mathrm{d}s} \frac{\mathrm{d}\gamma^\nu}{\mathrm{d}s} = 0 \tag{2.51}$$

The deeper meaning of this equation is that it is derived only by using the connection $\Gamma$ and, as any curve, can be used to measure distances by integrating the velocity. This remark gives rise to the idea that the connection must have a relation with the metric tensor of the manifold. Of course, this last sentence does not imply that every connection can be written in terms of a metric tensor.

A very basic introduction of the metric tensor is given bellow. The reason that there are no details on this subject, is that the purposes of this chapter is to give a basic introduction of differential geometry, in order to relate its ideas to the Yang-Mills theory over the Minkowski spacetime, where the metric tensor is only used in basic calculations.

As stated above, the idea arises from the need to measure the distance of two arbitrary points on a manifold. The metric is introduced as:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{2.52}$$

This equation an be used to prove that the metric transforms as a second rank covariant tensor. The distance of two arbitrary points on the manifold must remain the same under the diffeomorphic transformation of Equation 2.1.

$$\begin{aligned}
(ds')^2 &= (ds)^2 \\
g'_{\kappa\lambda} dx'^\kappa dx'^\lambda &= g_{\mu\nu} dx^\mu dx^\nu \\
g'_{\kappa\lambda} \frac{\partial x'^\kappa}{\partial x^\mu} dx^\mu \frac{\partial x'^\lambda}{\partial x^\nu} dx^\nu &= g_{\mu\nu} dx^\mu dx^\nu \\
g'_{\kappa\lambda} \frac{\partial x'^\kappa}{\partial x^\mu} \frac{\partial x'^\lambda}{\partial x^\nu} &= g_{\mu\nu} \\
\frac{\partial x^\mu}{\partial x'^\kappa} \frac{\partial x^\nu}{\partial x'^\lambda} g_{\mu\nu} &= g'_{\kappa\lambda}
\end{aligned} \tag{2.53}$$

The last equation is exactly the rule of transformation a covariant second rank tensor should follow, so it concludes the proof. Now that the tensorial character of the metric is established,

an obvious result is that the product $g_{\mu\nu}A^\mu B^\nu$ is an invariant quantity. This result is often used to convert a contravariant vector to a corresponding covariant one.

$$g_{\mu\nu}A^\mu B^\nu = A_\nu B^\nu$$

The meaning of the above equation is that the contravariant vector $A^\mu$ can be replace by the covariant vector $g_{\mu\nu}A^\mu$.

$$A_\nu = g_{\mu\nu}A^\mu \tag{2.54}$$

It can be proved that the inverse of the metric tensor exists and in any frame they are related as:

$$g^{\mu\kappa}g_{\mu\lambda} = \delta^\kappa_\lambda \tag{2.55}$$

The inverse metric $g^{\mu\kappa}$ transforms as a contravariant second rank rank and using the same arguments as before, it can be used to transform a covariant vector into a contravariant one.

$$A^\nu = g^{\mu\nu}A_\mu \tag{2.56}$$

Another important property of the metric tensor is its relation with the Jacobian defined by the diffeomorphism 2.1. Starting from the transformation of the metric, implied by Equation 2.53 and the determinants $g$ and $g'$ of the tensors $g_{\mu\nu}$ and $g'_{\kappa\lambda}$ respectively one gets the following formula.

$$g' = \left|\frac{\partial x^\mu}{\partial x'^\kappa}\right|^2 g$$

The middle term of this equation is the Jacobian determinant of the frame transformation, which gives the final result, that relates the metric tensor between two frames with the Jacobian.

$$J = \sqrt{\frac{g'}{g}} \tag{2.57}$$

This result is really important for the subject of this thesis, because the Yang-Mills equations are studied in a Minkowski spacetime background, where the determinant of the metric is $g = 1$ and as a result the Jacobian is also $J = 1$. Consequently, in the background of special relativity, the relation between tensors and tensor densities described in Section 2.1.3 is trivial, meaning there is no distinction between tensor and densities.

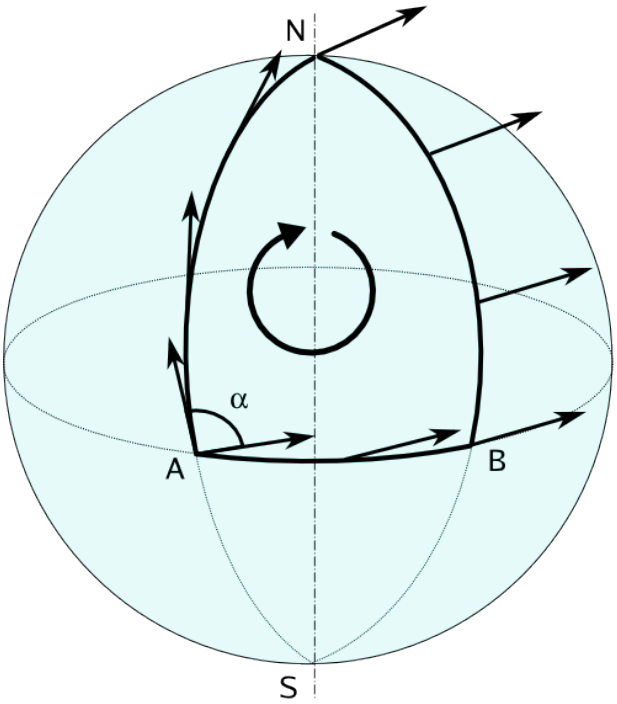



**Figure 2.2:** Enter Caption

### 2.1.8 Curvature and Torsion Tensors

In section 2.1.5 the notion of parallel transfer was introduced. An interesting conclusion is that when parallel transporting a vector over a closed path the final vector does not coincides with the initial vector, because of the underlying structure of the manifold.

The appropriate measure of this result is, of course, the covariant derivative. For this section, a different and more famous notation for the covariant is used. Specifically, the covariant derivative of a contravariant vector is denoted as[12]:

$$\nabla_\mu A^\nu = \partial_\mu A^\nu + \Gamma^\nu_{\kappa\mu} A^\kappa$$

Going back to measuring the effect of the underlying geometry of the spacetime manifold $\mathcal{M}$, when parallel transporting a vector $A^\mu$ from a point $A \in \mathcal{M}$ the same point using two different closed paths, as shown in Figure 2.2. The difference between the initial vector at $A$ and the parallel transferred vector at $A$ is measured by the commutator of the covariant derivatives in the $\mu$ and $\nu$ directions.

[12]The symbol $\partial_\mu$ is just shorthand notation for the usual derivative operator $\frac{\partial}{\partial x^\mu}$

$$
\begin{aligned}
[\nabla_\mu, \nabla_\nu]\, A^\rho &= \nabla_\mu(\nabla_\nu A^\rho) - \nabla_\nu(\nabla_\mu A^\rho) \\
&= \partial_\mu(\nabla_\nu A^\rho) - \Gamma^\sigma_{\nu\mu}(\nabla_\sigma A^\rho) + \Gamma^\rho_{\kappa\mu}(\nabla_\nu A^\kappa) - \\
&- \partial_\nu(\nabla_\mu A^\rho) - \Gamma^\sigma_{\mu\nu}(\nabla_\sigma A^\rho) + \Gamma^\rho_{\kappa\nu}(\nabla_\mu A^\kappa) \\
&= (\partial_\mu \Gamma^\rho_{\kappa\nu} - \partial_\nu \Gamma^\rho_{\kappa\mu}) A^\kappa + (\Gamma^\sigma_{\mu\nu} - \Gamma^\sigma_{\nu\mu})\nabla_\sigma A^\rho + \\
&+ \Gamma^\rho_{\kappa\nu}(\partial_\mu A^\kappa - \nabla_\mu A^\kappa) + \Gamma^\rho_{\kappa\mu}(\partial_\nu A^\kappa - \nabla_\nu A^\kappa) \\
[\nabla_\mu, \nabla_\nu]\, A^\rho &= (\partial_\mu \Gamma^\rho_{\kappa\nu} - \partial_\nu + \Gamma^\rho_{\kappa\mu}\Gamma^\rho_{\sigma\mu}\Gamma^\sigma_{\kappa\nu} - \Gamma^\rho_{\sigma\nu}\Gamma^\sigma_{\kappa\mu}) A^\kappa + (\Gamma^\sigma_{\mu\nu} - \Gamma^\sigma_{\nu\mu})\nabla_\sigma A^\rho
\end{aligned}
$$

From the last equation the two following two tensors can be defined:

$$R^\rho_{\kappa\mu\nu} = \partial_\mu \Gamma^\rho_{\kappa\nu} - \partial_\nu + \Gamma^\rho_{\kappa\mu}\Gamma^\rho_{\sigma\mu}\Gamma^\sigma_{\kappa\nu} - \Gamma^\rho_{\sigma\nu}\Gamma^\sigma_{\kappa\mu} \tag{2.58}$$

$$T^\sigma_{\mu\nu} = \Gamma^\sigma_{\mu\nu} - \Gamma^\sigma_{\nu\mu} \tag{2.59}$$

The tensor $R^\rho_{\kappa\mu\nu}$ is called the Riemann curvature tensor or just curvature of the manifold. The tensor $T^\sigma_{\mu\nu}$ is the torsion tensor. The important fact is that even in flat manifolds where all the components of the curvature are equal to zero, covariant derivatives may not commute, meaning that there is an urge for a connection. In classical general relativity the spacetime manifold is considered torsion-free, which is achieved by assuming symmetric affine connections. This kind of connections are often called Christoffel symbols.

## 2.2 Group Theory

### 2.2.1 General Definitions

Group theory is the mathematical theory that studies sets equipped with some special properties. The first important thing to do is to define what are these properties a set must have to be promoted into a group.

**Definition 2.2.1** (Group)**.** A group is a set $G = \{g\}$ equipped with a binary operation $\odot$, which is actually a map:

$$\odot : \; G \times G \longrightarrow G$$

$$(g_1, g_2) \longmapsto g_1 \odot g_2$$

The binary operation must fulfill the following properties:

(i) Closure: $\forall\, g_1, g_2 \in G,\; g_1 \odot g_2 \in G$

(ii) Associativity: $\forall g_1, g_2, g_3 \in G,\ g_1 \odot (g_2 \odot g_3) = (g_1 \odot g_2) \odot g_3$

(iii) Identity: $\exists e \in G$ such that $: \forall g \in G,\ e \odot g = g \odot e = g$

(iv) Inverse: $\exists g^{-1} \in G\ \forall g \in G$ such that $: g^{-1} \odot g = g \odot g^{-1} = e$

If the set $G$ equipped with the operation $\odot$ form a group the notation $(G, \odot)$ is widely used.

The term binary operation in the above Definition 2.2.1 is used as a fancy term to describe a map that takes two elements of a group $G$ and maps them to another element of the group. The exact form of the operation should be chosen according to the properties of Definition 2.2.1 and, of course, it depends on the elements of the group. For example, the set of integer numbers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ *equipped with addition* forms a group, because:

(i) The sum of two integers is an integer.

(ii) Addition is associative.

(iii) The number $0 \in \mathbb{Z}$ is the identity element of the addition $0 + a = a + 0 = a\ \forall a \in \mathbb{Z}$

(iv) The inverse element is just the opposite number, which belongs in the group: $\forall a \in \mathbb{Z}$ : $\exists\, (-a) \in \mathbb{Z}$ such that $a + (-a) = -a + a = 0$

The above example is very illustrative, because it practically emphasizes the fact that when defining a group two things must be defined: the set and the operation. If multiplication were chosen as the binary operation instead of the addition, then the combination $(\mathbb{Z}, \cdot)$ would not form a group, because the inverse of an integer number is not an integer itself. Moreover, the combination $(\mathbb{N}, +)$ do not form a group, because the opposite number of a natural number (that would be the inverse in terms of group theory) is not a natural number.

Two important properties of a group $(G, \odot)$ are the uniqueness of the identity element $e \in G$ and the uniqueness of the inverse $g^{-1} \in G$ of every element $g \in G$.

*Uniqueness of the identity.* Suppose that in a group $(G, \odot)$ exist two different identity elements $e_1, e_2 \in G,\ e_1 \neq e_2$.
From the identity property defined in Definition 2.2.1 $\forall g \in G$:

$$e_1 \odot g = g \odot e_1 = g$$

$$e_2 \odot g = g \odot e_2 = g$$

Because the above equations are true $\forall g \in G$, then for $g = e_2$ the first equation gives:

$$e_1 \odot e_2 = e_2$$

and the second equation for $g = e_1$ gives:

$$e_1 \odot e_2 = e_1$$

Combining the last two results:

$$e_1 = e_2$$

which is a contradiction according to the original hypothesis $e_1 \neq e_2$, which means that $e_1 = e_2$. This procedure can be followed in a similar manner to prove that $e_1 = e_2 = \ldots = e_n$, which means that the identity element of a group is unique.

□

*Uniqueness of the inverse.* Suppose that in a group $(G, \odot)$ exist two different inverse elements $a, b \in G,\ a \neq b$ for each element of the group $g \in G$.
From the inverse property defined by the Definition 2.2.1:

$$a \odot g = g \odot a = e$$

$$b \odot g = g \odot b = e$$

So, using the identity element $e \in G$ and the associativity of the group' s binary operation:

$$a = a \odot e = a \odot (b \odot a) = (a \odot b) \odot a = e \odot b = b$$

The last, statement is a contradiction, because in the original hypothesis it was $a \neq b$. This shall prove that the inverse $g^{-1} \in G$ is unique for each $g \in G$. □

**Definition 2.2.2** (Subgroup)**.** Let $(G, \odot)$ and $(H, \odot)$ be groups equipped with the same binary operation $\odot$ and underlying sets $G = \{g\}$, $H = \{h\}$. If $H \subseteq G$ then $(H, \odot)$ is called a subgroup of $(G, \odot)$.

Note that by the uniqueness of the identity proven above, the identity element $e \in G$ should also be an element of the set $H$, in order for the set $H$ to be considered a group. This means, that if $H$ is a subgroup of $G$ then, these two groups have the same identity element $e$.

**Definition 2.2.3** (Abelian Group)**.** Let $(G, \odot)$ be a group. The group $G$ is called Abelian if:

$$\forall g_1, g_2 \in G,\ g_1 \odot g_2 = g_2 \odot g_1$$

In other words, if a group' s binary operation is commutative then the group is called Abelian. For example, the set of integers equipped with addition form an Abelian group. Of course, even if $g_1 \odot g_2 \neq g_2 \odot g_1$ for two specific $g_1, g_2 \in G$ then the group is called non-Abelian.

At this point, it should not be forgotten that groups are sets with some special properties, which means that maps can be defined with groups playing the role of the domain and the target set. Some special cases of such maps are given by the definitions bellow.

**Definition 2.2.4** (Homomorphism)**.** Given two groups $(G, \odot)$ and $(H, \bullet)$, a group homomorphism is a map $\phi : G \longrightarrow H$ such that $\forall g_1, g_2 \in G$:

$$\phi(g_1 \odot g_2) = \phi(g_1) \bullet \phi(g_2)$$

**Definition 2.2.5** (Isomorphism)**.** Let $\phi : G \longrightarrow H$ be a group homomorphism. If the map $\phi$ is a bijective map then the map $\phi$ is called isomorphism and the groups $(G, \odot)$, $(H, \bullet)$ are called isomorphic groups if such map exists. The notations $(G, \odot) \cong (H, \bullet)$ and $G \cong H$ are equivalently used.

The isomorphism map is really important, because it is what a mathematician would call the "structure preserving map".

### 2.2.2 Matrix Groups

From the era of classical physics it has been known that matrices play a very important role in transformations and, as a result, in symmetries of a system. To be exact, every matrix express a linear transformation. [13] For this reason, the most important matrix sets are defined in this section. It has to be clear that the following definitions define the matrices as sets and not groups. Although, each definition is accompanied with a proof that shows it also forms a group with binary operation the matrix multiplication. So, the term matrix set or matrix group is used equivalently when referring to one of the following definitions.

**Definition 2.2.6** (General Linear Group)**.** The General Linear Group over $\mathbb{R}$ is the set of all invertible square $n \times n$ matrices of real values $(M^{\mathbb{R}}_{n\times n})$, with binary operation the matrix multiplication.

$$GL(n, \mathbb{R}) = \{A \in M^{\mathbb{R}}_{n\times n} \mid \det(A) \neq 0\}$$

*Proof.* (i) Closure: Assume two matrices $A_{n\times n}, B_{n\times n} \in GL(n, \mathbb{R})$. It is known from linear algebra that the result of the multiplication of two square $n \times n$ matrices is an $n \times n$ square matrix.[14]

$$A_{n\times n} \cdot B_{n\times n} = C_{n\times n} \tag{2.60}$$

So, for $C_{n\times n}$ to belong in the set $GL(n, \mathbb{R})$, it needs to have $\det(C) \neq 0$.

Again from linear algebra:[15]

[13] The inverse of this statement is not true.

[14] The symbol "$\cdot$" denotes ordinary matrix multiplication

[15] The subscript $n \times n$ denoting the dimensions of each matrix is omitted, because all matrices mentioned in the present section are $n \times n$ square matrices

$$det(A \cdot B) = \det(A)\det(B) \tag{2.61}$$

The result of these two equations is:

$$\det(C) = \det(A)\det(B)$$

Because, $A, B \in GL(n, \mathbb{R})$, $\det(A) \neq 0$ and $\det(B) \neq 0$. What follows from the last equations is that $\det(C) \neq 0$. As a consequence, $\forall A, B \in GL(n, \mathbb{R})$, $A \cdot B = C \in GL(n, \mathbb{R})$.

(ii) Associativity: From basic linear algebra it is known that matrix multiplication between square matrices of the same dimensionality is an associative operation. $\forall A, B, C \in GL(n, \mathbb{R})$, $(A \cdot B) \cdot C = A \cdot (B \cdot C)$.

(iii) Identity: The identity element is the identity $n \times n$ square matrix ($\mathbb{1}_{n\times n} \equiv \mathbb{1}$), because $\forall A \in GL(n, \mathbb{R})$, $A \cdot \mathbb{1} = \mathbb{1} \cdot A = A$ . Of course, $\det(\mathbb{1}) \neq 0$, so $\mathbb{1} \in Gl(n, \mathbb{R})$.

(iv) Inverse: Every square matrix $A \in GL(n, \mathbb{R})$ is an invertible matrix, because $\det(A) \neq 0$, which means that $\exists A^{-1}$, $A^{-1} \cdot A = A \cdot A^{-1} = \mathbb{1}$. The inverse matrix $A^{-1}$ must be an element of $GL(n, \mathbb{R})$. Again, from basic linear algebra:

$$\det\Big(A \cdot A^{-1}\Big) = \det(A)\det\Big(A^{-1}\Big)$$

but $A \cdot A^{-1} = \mathbb{1}$, $\det(\mathbb{1}) = 1$ and $\det(A) \neq 0$, so:

$$\det\Big(A^{-1}\Big) = \frac{1}{\det(A)} \neq 0 \tag{2.62}$$

The last equation proves that $A^{-1} \in GL(n, \mathbb{R})$

□

**Definition 2.2.7** (Special Linear Group)**.** The Special Linear Group over $\mathbb{R}$ is the set of all square $n \times n$ matrices with real values and determinant equal to one, with binary operation the matrix multiplication.

$$SL(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) \mid \det(A) = 1\}$$

Consequently, according to the Definition 2.2.2, $SL(n, \mathbb{R}) \subseteq GL(n, \mathbb{R})$

*Proof.* One can prove that the Special Linear Group is indeed a group by following the same steps presented above, where it was shown that the General Linear Group is indeed a group.

(i) Closure: $\forall A, B \in SL(n, \mathbb{R})$ and $C$ defined as $C = A \cdot B$, one should prove that $C$ belongs in the set $SL(n, \mathbb{R})$. So, it should be proven that $C$ is a square matrix with $\det(C) = 1$. As implied in the previous proof (see Equation 2.60) $C$ is a square $n \times n$ matrix. As for

the determinant: $\det(A) = 1$, $\det(B) = 1$ and using Equation 2.61 $\det(C) = 1$, which means that $C \in SL(n, \mathbb{R})$.

(ii) Associativity: Multiplication of square matrices is an associative operation

(iii) Identity: It must be proven that the identity matrix $\mathbb{1}$ belongs in $SL(n, \mathbb{R})$, which is true because every identity matrix has $\det(\mathbb{1}) = 1$.

(iv) Inverse: It can be proven, following the same steps presented in the same section of the previous proof, that the inverse matrix $A^{-1}$ exists $\forall A \in SL(n, \mathbb{R})$. Last but not least, using Equation 2.62 and substituting the $\det(A) = 1$, the result is $\det(A^{-1}) = 1$, which proves that every inverse $A^{-1}$ of every element $A \in SL(n, \mathbb{R})$ belongs in the set $SL(n, \mathbb{R})$.

□

Sets of Definition 2.2.6 and Definition 2.2.7 are defined as sets of matrices, whose elements are real numbers. They can be extended to the groups $GL(n, \mathbb{C})$ and $SL(n, \mathbb{C})$ respectively, so that matrices can contain complex numbers as well. These expansions do not change anything in the proofs provided above.

**Definition 2.2.8** (Orthogonal Group)**.** The set of the Orthogonal Group is:

$$O(n) = \{A \in GL(n, \mathbb{R}) \,|\, A^T A = A A^T = \mathbb{1}_{n \times n}\}$$

where the condition:

$$A^T A = A A^T = \mathbb{1} \tag{2.63}$$

is known as the *orthogonality condition*.

Directly by the definition, it should be clear that $O(n) \subseteq GL(n, \mathbb{R})$.

*Proof.* (i) Closure: According to the Equation 2.60, if a matrix $C$ is defined as $A \cdot B = C$, with $A, B \in O(n)$, then $C$ is a square matrix. Moreover, it must be proven that $C \cdot C^T = C^T \cdot C = \mathbb{1}$. Using the property:

$$(M \cdot N)^T = N^T \cdot M^T \tag{2.64}$$

and the associativity of matrix multiplication between square matrices of the same dimension:

$$C \cdot C^T = (A \cdot B) \cdot (A \cdot B)^T = (A \cdot B) \cdot (B^T \cdot A^T) = A \cdot (B \cdot B^T) \cdot A^T \tag{2.65}$$

Now, knowing that $A, B \in O(n)$ means that the orthogonality condition 2.63 holds for the matrices $A, B$ and Equation 2.65 gives:

$$C \cdot C^T = A \cdot \mathbb{1} \cdot A^T = A \cdot A^T = \mathbb{1}$$

Following exactly the same steps one can prove that $C^T \cdot C = \mathbb{1}$. This means that $C \in O(n)$.

(ii) Associativity: As mentioned before matrix multiplication between square matrices of the same dimensions is associative.

(iii) Identity: the identity element is the identity matrix $\mathbb{1}$. For this to be true, the identity matrix must be an element of the group $O(n)$. This is true, because $\mathbb{1}^T = \mathbb{1}$ and consequently $\mathbb{1} \cdot \mathbb{1}^T = \mathbb{1}$.

(iv) Inverse: Because $O(n) \subseteq GL(n, \mathbb{R})$, the inverse $A^{-1}$ of every element $A \in O(n)$ exists and is $A^{-1} \in GL(n, \mathbb{R})$, but this inverse element should also be an element of the group $O(n)$. So, the following Equation 2.66 must be proven.

$$A^{-1} \cdot \left(A^{-1}\right)^T = \left(A^{-1}\right)^T \cdot A^{-1} = \mathbb{1} \tag{2.66}$$

To prove this equation Property 2.64 combined with the following Property 2.67 will be useful.

$$(M \cdot N)^{-1} = N^{-1} \cdot M^{-1} \tag{2.67}$$

Starting from the equation $A \cdot A^T = \mathbb{1}$, which is true $\forall A \in O(n)$ and substituting $A = A^{-1}$, the equation $\left(A^{-1}\right)^T \cdot A^{-1} = \mathbb{1}^{-1} = \mathbb{1}$ is obtained. The last equation concludes the proof that $\forall A \in O(n) \exists A^{-1} \in O(n)$

□

The significance of orthogonal matrices is already known from linear algebra. Two matrices are called orthogonal if their rows and columns are orthonormal vectors.[16] Mathematically this is expressed with the orthogonality condition 2.63. Column matrices of $n$ values can represent a vector in an n-dimensional vector space. When an orthogonal matrix $O_{n\times n}$ acts on a column matrix $A_{n\times 1}$, which represents an n-component vector, a new n-component vector is produced which is rotated in the basis formed by the orthonormal vectors, appearing as the rows and columns of the orthogonal matrix $O$. As a general statement it can be said that rotations of real vectors are represented by orthogonal matrices.

Another important property of orthogonal matrices can be proved after having established that they form a group. The fact that the inverse element of a group is unique means that the inverse $O^{-1}$ of an orthogonal matrix $O$ is unique. In a more mathematical language this translates as: "There exists exactly one matrix $O^{-1} \in O(n)$, that satisfies the equations $O \cdot O^{-1} = O^{-1} \cdot = \mathbb{1}$." Combining the previous statement with the orthogonality condition of Equation 2.63, it can be concluded that **for every orthognal matrix**:

[16] That is why orthogonal matrices are also referred to as *orthonormal matrices*

$$O^{-1} = O^T \tag{2.68}$$

An important property that should be taken into consideration is that **for every matrix** $A$, the following equation holds.

$$\det\Big(A^T\Big) = \det(A) \tag{2.69}$$

So, for two orthogonal matrices, taking the determinant on both sides of the Equation 2.63 results in:

$$\det\Big(O \cdot O^T\Big) = \det(O)\det\Big(O^T\Big) = \det(O)^2 = \det(\mathbb{1}) = 1$$

which is:

$$\det(O)^2 = 1 \Rightarrow \det(O) = \pm 1 \tag{2.70}$$

This result is very important, because it restricts the determinant of an orthogonal matrix to the values $\pm 1$. The special case of $\det(O) = 1$ is the case of the Special Orthogonal Group described in Definition 2.2.9 right bellow.

**Definition 2.2.9** (Special Orthogonal Group)**.** Following the same pattern as in Definition 2.2.7, the set of the Special Orthogonal Group is defined as:

$$SO(n) = \{A \in O(n) \mid \det(A) = 1\}$$

Of course, $SO(n) \subseteq O(n)$

*Proof.* The proof that the Special Orthogonal Group does indeed form a group is trivial combining the proofs provided for the Orthogonal Group and the Special Linear Group. □

Notice that when defining the Orthogonal and the Special Orthogonal groups there is no parameter $\mathbb{R}$ inside $O(n)$ and $SO(n)$, as there was in $GL(n, \mathbb{R})$ and in $SL(n, \mathbb{R})$. This is happening, because we need orthogonal matrices to represent rotations, which are expressed by the condition 2.63, only in real vector spaces $\mathbb{R}^n$. The analogue in complex vector spaces is expressed by the unitary matrices defined bellow.

**Definition 2.2.10** (Unitary Group)**.** The Unitary Group is a matrix group with set:[17]

$$U(n) = \{A \in GL(n, \mathbb{C}) \mid A \cdot A^\dagger = A^\dagger \cdot A = \mathbb{1}_{n\times n}\}$$

The condition:

[17] $A^\dagger$ is used to denote the Hermitian conjugate of the matrix $A$, which is $A^\dagger \equiv \big(A^T\big)^\star = (A^\star)^T$

$$A \cdot A^{\dagger} = A^{\dagger} \cdot A = \mathbb{1} \tag{2.71}$$

is known as the *unitarity condition*

*Proof.* The proof is actually very similar to that provided for the group $O(n)$, but there are some differences arising from the use of the Hermitian conjugate.

(i) Closure: Let' s define again the matrix $C = A \cdot B$, with $A, B \in C$. As described many times above, $C$ is a square matrix with the same dimensions as the matrices $A, B$. As a second step to prove closure, the unitarity condition 2.71 must be proven. This goes as follows:

$$C \cdot C^{\dagger} = (A \cdot B) \cdot (A \cdot B)^{\dagger} \tag{2.72}$$

Now using the property:

$$(M \cdot N)^{\dagger} = N^{\dagger} M^{\dagger} \tag{2.73}$$

and the associativity of matrix multiplication the equation 2.72 is continued as:

$$C \cdot C^{\dagger} = A \cdot (B \cdot B^{\dagger}) \cdot A^{\dagger}$$

Now, remembering that $A, B \in U(n)$, $A, B$ fulfill the unitarity condition 2.71. Using this condition the last equation becomes:

$$C \cdot C^{\dagger} = A \cdot \mathbb{1} \cdot A^{\dagger} = A \cdot A^{\dagger} = \mathbb{1}$$

One can start from the side $C^{\dagger} \cdot C$, follow the exact same steps and finally prove the closure property of the group.

(ii) Associativity: Matrix multiplication is an associative operation

(iii) Identity: The Hermitian conjugate of the unitary matrix is the unitary matrix itself $\mathbb{1}^{\dagger} = \mathbb{1} \Rightarrow \mathbb{1} \in U(n)$.

(iv) Inverse: Again, $U(n) \subseteq GL(n, \mathbb{C})$, so $\forall A \in U(n)$, $\det(A) \neq 0 \Rightarrow \exists A^{-1} \,|\, A^{-1} \cdot A = A \cdot A^{-1} = \mathbb{1}$. This inverse matrix should be an element of the group $U(n)$, meaning that it must have the property $(A^{-1})^{\dagger} \cdot A^{-1} = A^{-1} \cdot (A^{-1})^{\dagger} = \mathbb{1}$. To prove this, linear algebra is needed again. Starting from the term $A \cdot A^{\dagger}$ and using the fact that if two invertible matrices are equal then their inverse matrices must be equal combined with the properties 2.73, 2.67 and $\mathbb{1}^{-1} = \mathbb{1}$, the equation takes the form:

$$\left(A \cdot A^{\dagger}\right) = \left(A^{\dagger}\right)^{-1} \cdot A^{-1} = \left(A^{-1}\right)^{\dagger} \cdot A^{-1} = \mathbb{1}$$

By this last equation the proof that the set $U(n)$ does indeed form a group is finished.

□

Unitary matrices have a similar property to that of orthogonal matrices concerning their determinant. In fact, taking into consideration the property:

$$\det\left(U^{\dagger}\right) = \det(U)^{\star} \tag{2.74}$$

and then taking the determinant of $U^{\dagger} \cdot U = 1$:

$$\det\left(U^{\dagger} \cdot U\right) = \det\left(U^{\dagger}\right)\det(U) = \det(U)^{\star}\det(U) = |\det(U)|^2 = 1$$

The last equation means that the determinant of a unitary matrix has the the form:

$$\det(U) = e^{ia},\ a \in [0, 2\pi) \tag{2.75}$$

**Definition 2.2.11** (Special Unitary Group)**.** The Special Unitary Group is a subgroup of the group $U(n)$ defined in Definition 2.2.10, with set:

$$SU(n) = \{A \in U(n) \mid \det(A) = 1\}$$

*Proof.* Proof is again trivial if one combines the techniques used for proving that $U(n)$ and $SL(n, \mathbb{R})$ form groups. □

### 2.2.3 Free Parameters of Matrix Groups

An important feature of the matrix groups defined in Section 2.2.2 is the *free parameters* of each group. This term refers to the number of matrix elements in each matrix of a matrix group, that are not fixed by any condition. First of all, it should be noted that $SL(n, \mathbb{R}), O(n), SO(n)$ are subgroups of the $GL(n, \mathbb{R})$ group, while $U(n), SU(n)$ are subgroups of $GL(n, \mathbb{C})$. As a result, each subgroup will have less free parameters than $GL(n, \mathbb{R})$ and $GL(n, \mathbb{C})$ respectively.

The group $GL(n, \mathbb{R})$ contains all invertible $n \times n$ square matrices, so its free parameters are the $n^2$ elements contained in each matrix. The condition $\det(A) \neq 0$ does not fix a parameter, it just rejects specific simultaneous values of the free parameters. In that sense, each matrix of $GL(n, \mathbb{R})$ can describe a point in a vector space $\mathbb{R}^{n^2}$, but every point of $\mathbb{R}^{n^2}$ could not be described as a matrix of $GL(n, \mathbb{R})$, because $\mathbb{R}^{n^2}$ also contains the points $(a, b, c, d)$ that lie on the surface $ac - bd = 1$. These points correspond to non-invertible $n \times n$ square matrices.[18]

To calculate the number of free parameters needed to fully define each matrix of the $SL(n, \mathbb{R})$ group, one should start from the $n^2$ free parameters of a matrix of the $GL(n, \mathbb{R})$ group. The

[18]This is a first try to connect the idea of matrix groups and the idea of surfaces of $\mathbb{R}^{n^2}$

condition for the determinant is one equation. As a result, it fixes only one parameter, so the number of free parameters of $SL(n, \mathbb{R})$ is $n^2 - 1$.

The same procedure can be followed to define the number of free parameters for the groups $O(n)$ and $SO(n)$. Although, the conditions of these groups are a little more complicated, it should be helpful to see explicitly what is happening for the matrices $O \in O(n)$ and $SO \in SO(n)$. One can start from $2 \times 2$ and $3 \times 3$ matrices and then generalize the result for arbitrary $n \times n$ square matrices.

$$O = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad O^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

The orthogonality condition 2.63 then gives:

$$O \cdot O^T = \begin{pmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Notice that the off-diagonal elements of this last equation give only one equation for the free parameters. In the case of $2 \times 2$ matrices the free parameters started from $2^2 = 4$ and reduced to 1, by the orthogonality condition. To generalize this result, let' s see what happens in $3 \times 3$ matrices.

$$O = \begin{pmatrix} a & b & c \\ e & f & g \\ h & i & j \end{pmatrix} \quad O^T = \begin{pmatrix} a & e & h \\ b & f & i \\ c & g & j \end{pmatrix}$$

Applying the orthogonality condition:

$$O \cdot O^T = \begin{pmatrix} a^2 + b^2 + c^2 & ae + bf + cg & ah + bi + cj \\ ae + bf + cg & e^2 + f^2 + g^2 & eh + fi + gj \\ ah + bi + cj & eh + fi + gj & c^2 + g^2 + j^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

By increasing the dimension of the matrix it becomes clear that the orthogonality condition gives one symmetric matrix, which must be equal to the identity matrix. For an arbitrary $n \times n$ matrix $n$ equations arise from the diagonal. The off-diagonal elements of an $n \times n$ square matrix count $n(n-1)$. But, because the matrix is symmetrical, only half of those elements give different equations. In total the different equations are:

$$n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$$

The free parameters were $n^2$ at first, so after fixing they are:

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

Now, according to the Equation 2.70 the determinant of an orthogonal matrix has two possible values: $\pm 1$. If an orthogonal matrix has its determinant equal to 1 then it is called special orthogonal matrix and it is an element of $SO(n)$. That being said, the condition $\det(A) = 1$ in the Definition 2.2.9 does not fix any free parameters, because there are infinite sets of free parameters that satisfy this condition. As a result, the number of free parameters of $SO(n)$ is also $\frac{n(n-1)}{2}$.

As mentioned above all the elements of a unitary matrix $U \in U(n)$ are complex numbers, because $U(n)$ is a subgroup of $GL(n, \mathbb{C})$. A matrix of $GL(n, \mathbb{C})$ has $n^2$ free complex parameters. Each complex number is composed by two real numbers, meaning that each matrix of $GL(n, \mathbb{C})$ consists of $2n^2$ free real parameters. Again, one can think of each matrix of $GL(n, \mathbb{C})$ as a point in $\mathbb{C}^{n^2}$. The vector space $\mathbb{C}^{n^2}$ without all the points $x = (a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, \ldots, a_{2n}, \ldots a_{n1}, a_{n2}, \ldots, a_{nn}) \in \mathbb{C}^{n^2}$, $a_{11}, a_{12}, \ldots, a_{nn} \in \mathbb{C}$, that satisfy the condition $det(A) \neq 1$, $A = [a_{ij}]$ , $i, j = 1, 2, \ldots, n$.[19]

At first, let' s calculate the free parameters of the unitary $2 \times 2$ square matrix. A general complex $2 \times 2$ square matrix and its hermitian conjugate can be seen bellow:

$$U = \begin{pmatrix} a_r + ia_i & b_r + ib_i \\ c_r + ic_i & d_r + id_i \end{pmatrix} \quad U^\dagger = \begin{pmatrix} a_r - ia_i & c_r - ic_i \\ b_r - ib_i & d_r - id_i \end{pmatrix}$$

Now, using the condition, which appeared in Definition 2.2.10:

$$U \cdot U^\dagger = \begin{pmatrix} a_r^2 + a_i^2 + b_r^2 + b_i^2 & (a_r c_r + a_i c_i + b_r d_r + b_i d_i) + i(a_i c_r - a_r c_i + b_i d_r - b_r d_i) \\ (a_r c_r + a_i c_i + b_r d_r + b_i d_i) - i(a_i c_r - a_r c_i + b_i d_r - b_r d_i) & c_r^2 + c_i^2 + d_r^2 + d_i^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

From the last equation it is quite obvious that the resulting matrix is not symmetrical. The off-diagonal elements are complex conjugate numbers. Because of this relation between the off-diagonal elements the last equation gives four separate equations between the free parameters of the unitary matrix. Two of them originate from the diagonal, while the other two come from the real and the imaginary part of the off-diagonal elements, that must be both equal to zero.

For a general $n \times n$ unitary matrix the equation $U \cdot U^\dagger = \mathbb{1}_{n \times n}$ gives $n$ equations from the diagonal elements and two equations from half of the off-diagonal elements. The equations that arise in total are:

$$n + 2\frac{n(n-1)}{2} = n^2$$

Initially, the real parameters were $2n^2$. Subtracting the $n^2$ fixed parameters, the result is $n^2$ and this is the final number for the free real parameters for a unitary matrix of the group $U(n)$.

[19] The notation $A = [a_{ij}]$ is used to denote the square matrix: $A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \\ a_{n1} & & a_{nn} \end{bmatrix}$

| Group | No of Free Parameters |
|---|---|
| $GL(n, \mathbb{R})$ | $n^2$ |
| $GL(n, \mathbb{C})$ | $2n^2$ |
| $SL(n, \mathbb{R})$ | $n^2 - 1$ |
| $SL(n, \mathbb{C})$ | $2n^2 - 1$ |
| $O(n)$ | $\frac{n(n-1)}{2}$ |
| $SO(n)$ | $\frac{n(n-1)}{2}$ |
| $U(n)$ | $n^2$ |
| $SU(n)$ | $n^2 - 1$ |

**Table 2.2:** The number of free **real** parameters for each matrix group, that was mentioned in Section 2.2.2.

Here, the special case of the group $SU(n)$ is different than the special case $SO(n)$. According to the Equation 2.75 the determinant of a unitary matrix can be any complex number of norm equal to 1, and that exactly is the difference with the orthogonal matrices, where the determinant can only take two values. That being said, the condition $\det(U) = 1$, $U \in SU(n)$, actually fixes a free parameter. As a result, the number of free parameters of the group $SU(n)$ is $n^2 - 1$, where $n^2$ are the free real parameters of $U(n)$ calculated above and 1 parameter is fixed from the determinant condition. The same is true for the group $SL(n, \mathbb{C})$, where the free parameters where $2n^2$ at first and after the determinant condition they are $2n^2 - 1$.

All the above calculations of the free parameters of each matrix group are summarized in Table 2.2.

### 2.2.4 Lie Groups, Lie Algebras and Representations

Lie groups are a special type of groups that connect the idea of a smooth manifold with group theory.

**Definition 2.2.12** (Lie Group)**.** A group $(G, \odot)$ is a Lie Group if the underlying set $G$ is a smooth manifold and the maps:

$$\begin{aligned} \odot : \ & G \times G \longrightarrow G \\ & (g_1, g_2) \longmapsto g_1 \odot g_2 \end{aligned}$$

$$\begin{aligned} i : & G \longrightarrow G \\ & g \longmapsto g^{-1} \end{aligned}$$

are smooth maps.

Of course, by the definition of manifolds the set must be continuous, meaning that a group has to be continuous in order to be a Lie Group. The term continuous when used for a set is a little controversial and is strictly defined in Topology. Intuitively a continuous set[20] can be thought as a set, where it is possible to "travel" from one element of the set to another without leaving the set.

Using more simple words, for a group $G$ to be characterized as a Lie Group, two conditions must be fulfilled. First, each one of its elements $g_i \in G$ must be described with some map $g: \mathbb{R}^n \longrightarrow G$, such that $(\theta_1, \theta_2, \ldots, \theta_n) \longmapsto g_i \in G$. The dimension $n$ is the number equal to the number of free real parameters of each group presented in Table 2.2.

All matrix groups studied in Section 2.2.2 are proven to be Lie Groups. As an example consider the group $SL(2,\mathbb{R})$. Every matrix $A \in SL(2,\mathbb{R})$ has 4 entries, say $(a,b,c,d)$. These entries can be thought as a point in $\mathbb{R}^4$. Now, according to Definition 2.2.7 the entries must satisfy the equation $ad - bc = 1$, which according to the implicit function theorem describe a surface of $\mathbb{R}^4$. So, a manifold is associated with the elements of the group $SL(2,\mathbb{R})$. The rest of the groups can be also associated with manifolds following the same thinking procedure. Of course, the groups of complex entries can be thought as complex manifolds as well.

**Definition 2.2.13** (Dimension of a Lie Group)**.** The dimension of a Lie Group is the dimension of the manifold.

It can be proven that the dimension of all matrix Lie Groups is the number of the free real parameters of each group given in Table 2.2. For example, $SU(n)$ is a $n^2 - 1$ Lie Group. A very useful property of Lie Groups is that they can be associated with a corresponding Lie Algebra, which is nothing but a vector space equipped with a special map defined bellow.

**Definition 2.2.14** (Lie Algebra)**.** A Lie Algebra $\mathfrak{g}$[21] is a vector space over some field $\mathbb{F}$[22] equipped with a map:

$$[,] : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}$$

which possesses the following properties:

(i) Bilinearity: $\forall \mathfrak{g}_1, \mathfrak{g}_2, \mathfrak{g}_3 \in \mathfrak{g}$ and $\forall a, b \in \mathbb{F}$:

$$[a\mathfrak{g}_1 + b\mathfrak{g}_2, \mathfrak{g}_3] = a[\mathfrak{g}_1, \mathfrak{g}_2] + b[\mathfrak{g}_2, \mathfrak{g}_3] \tag{2.76}$$

(ii) Antisymmetry: $\forall \mathfrak{g}_1, \mathfrak{g}_2 \in \mathfrak{g}$:

$$[\mathfrak{g}_1, \mathfrak{g}_2] = -[\mathfrak{g}_2, \mathfrak{g}_1] \tag{2.77}$$

[20] When referring to the continuity of a group we are actually referring to the continuity of its underlying set.

[21] When referring to a Lie Algebra the "mathfrak" notation will be used

[22] The term field here is used with its mathematical meaning. For our purposes of studying the matrix groups of Section 2.2 the field will be either $\mathbb{C}$ or $\mathbb{R}$

(iii) Jacobi Identity: $\forall \mathfrak{g}_1, \mathfrak{g}_2, \mathfrak{g}_3 \in \mathfrak{g}$:

$$[\mathfrak{g}_1, [\mathfrak{g}_2, \mathfrak{g}_3]] + [\mathfrak{g}_3, [\mathfrak{g}_1, \mathfrak{g}_2]] + [\mathfrak{g}_2, [\mathfrak{g}_3, \mathfrak{g}_1]] = 0 \tag{2.78}$$

The map of Definition 2.2.14 is called Lie Bracket and by definition it closes the algebra, meaning that every element of the target set of this map belongs to the Lie Algebra itself. The dimension of the Lie Algebra is equal to the dimension of the underlying vector space. Note that in Mathematics the term "algebra" is used to describe a **vector space** which is equipped with a map that closes the algebra. The difference between an algebra and a group is that groups do not have the structure of a vector space, they are just sets, that are equipped with the properties of Definition 2.2.1. Closure is included in those properties but it is not the only one. The importance and elegance of Lie Groups and Lie Algebras comes from the fact that they are connected in a way that is going to be described bellow, achieving in this way a connection between two different mathematical concepts.

As every vector space a Lie Algebra can be spanned by a basis. Also, the target set of the Lie Bracket is the Lie Algebra itself, meaning that the result of every Lie Bracket can be expanded in a basis of the Lie Algebra. Suppose a basis $T = \{T_a\} = \{T_1, T_2, \ldots, T_d\}, a = 1, 2, \ldots, d$ of a Lie Algebra $\mathfrak{g}$ of dimension $\dim(\mathfrak{g}) = d$. The Lie Bracket between two elements of the basis is $[T_a, T_b] = \mathfrak{g}_o$, where $\mathfrak{g}_o \in \mathfrak{g}$. Therefore, $\mathfrak{g}_o$ can be expanded as $\mathfrak{g}_o = f^c T_c$[23]. This remark encouraged the following definition for the structure constants of a Lie Algebra, which proved to be very useful for the connection between a Lie Algebra with a Lie Group and vice versa.

**Definition 2.2.15** (Structure Constants)**.** Assume a Lie Algebra $\mathfrak{g}$ of dimension $\dim(\mathfrak{g}) = d$ and a basis $T = \{T_a\}, a = 1, 2, \ldots, d$ of this algebra. The structure constants of the Lie Algebra are defined via the Lie Bracket as:

$$[T_a, T_b] = \sum_{c=1}^{d} f_{ab\,c} T_c = f_{ab}^c T_c \tag{2.79}$$

where $a, b, c \in \{1, 2, \ldots, d\}$

It should be noted that structure constants are the coefficients of an orthonormal expansion on the basis $T = \{T_a\}$, meaning that they are just complex numbers[24]. Applying the properties of the Lie Bracket from Definition 2.2.14 on the structure constants one can derive the properties that the structure constants must fulfill themselves. Firstly, from the antisymmetry property of the Lie Bracket follows:

[23] Einstein' s summation convention is implied

[24] Complex numbers are the most general case for the purposes of this thesis. The indices $a, b$ are there to remind that the structure constants do relate to the elements $T_a, T_b$ appearing in the Lie Brackets. Generally, the structure constants are elements of the field $\mathbb{F}$ appearing in Definition 2.2.14

$$\begin{aligned} [T_a, T_b] &= -[T_b, T_a] \\ f^c_{ab} T_c &= -f^c_{ba} T_c \\ f^c_{ab} &= -f^c_{ba} \end{aligned} \tag{2.80}$$

where the last line follows from the fact that $T = \{T_c\}$, so for two expansions on the same basis to equal, the coefficients of every respective elements $T_c$ must be equal. Equation 2.80 shows that the structure constants of every Lie Algebra are totally antisymmetric in their two bottom indices.

From the Jacobi identity of the Lie Bracket follows a kind of Jacobi identity for the structure constants. Applying the Jacobi identity on the basis $T = \{T_a\}$ and working term by term:

$$\begin{aligned} [T_a, [T_b, T_c]] &= \left[T_a, f^d_{bc} T_d\right] = f^d_{bc} \left[T_a, T_d\right] = f^d_{bc} f^e_{ad} T_e \\ [T_c, [T_a, T_b]] &= \left[T_c, f^d_{ab} T_d\right] = f^d_{ab} \left[T_c, T_d\right] = f^d_{ab} f^e_{cd} T_e \\ [T_b, [T_c, T_a]] &= \left[T_b, f^d_{ca} T_d\right] = f^d_{ca} \left[T_b, T_d\right] = f^d_{ca} f^e_{bd} T_e \end{aligned}$$

$$\begin{aligned} [T_a, [T_b, T_c]] + [T_c, [T_a, T_b]] + [T_b, [T_c, T_a]] &= 0 \\ f^d_{bc} f^e_{ad} T_e + f^d_{ab} f^e_{cd} T_e + f^d_{ca} f^e_{bd} T_e &= 0 \\ \left[f^d_{bc} f^e_{ad} + f^d_{ab} f^e_{cd} + f^d_{ca} f^e_{bd}\right] T_e &= 0 \end{aligned} \tag{2.81}$$

Because $T = \{T_c\}$ is a basis of the Lie Algebra, all elements of the basis are linearly independent:

$$\sum_{i=1}^{d} \lambda_i T_i = 0 \;\Rightarrow\; \lambda_i = 0 \;\forall\; i \in \{1, 2, \ldots, \dim(\mathfrak{g})\} \tag{2.82}$$

As a result, the only solution for Equation 2.79 is the following Equation 2.83, which is usually referring to as a Jacobi identity for the structure constants of a Lie Algebra.

$$f^d_{bc} f^e_{ad} + f^d_{ab} f^e_{cd} + f^d_{ca} f^e_{bd} = 0 \tag{2.83}$$

At this point, having defined Lie Groups and Lie Algebras it should be wise to understand the connection between them, implied by their common name. In terms of Differential Geometry a Lie Group $G$ has a corresponding Lie Algebra $\mathfrak{g}$, which is the tangent space $T_e(G)$ to the identity element of the group $e \in G$. It is a mathematical result that the exponential map is "responsible" for this connection. In fact, the exponential map is defined as $\exp : \mathfrak{g} \longrightarrow G$, via the Taylor expansion:

$$\exp\{iX\} = 1 + iX + i^2\frac{X^2}{2!} + i^3\frac{X^3}{3!} + \ldots,\ X \in \mathfrak{g} \tag{2.84}$$

The exponential map connects every $g_i = g_i(\theta_1, \theta_2, \ldots, \theta_n) \in G$ with an element $X \in \mathfrak{g}$ in the following way:

$$g_i = g_i(\theta_1, \theta_2, \ldots, \theta_n) = \exp\{i\theta^a T_a\} = 1 + i\theta^a T_a + i^2\frac{(\theta^a T_a)^2}{2!} + i^3\frac{(\theta^a T_a)^3}{3!} + \ldots \tag{2.85}$$

A very important remark on the above equation is that the real parameters $(\theta_1, \theta_2, \ldots, \theta_n)$ that describe (smoothly) a group element are the coefficients of the orthonormal expansion $X = \theta^a T_a$ describing the corresponding element $X \in \mathfrak{g}$ of the group element $g_i \in G$. Another important remark is that the identity element 1 appearing in the right hand side of Equation 2.85 could not in the general case be the identity element of the group, because in the general mathematical concept of an algebra there is no obligation for an identity to exist. Being more specific, if 1 is the identity element of the group how could it be possible to add this element with the polynomials that follow, which contain only elements of the Lie Algebra?

In fact, the Taylor expansion of Equation 2.85 only makes sense when one refers to a matrix representation of a Group. It can be mathematically proven that every compact group can be represented with elements in a matrix group. Every Lie Group is compact and the connection between a Lie Algebra and a Lie Group requires that every Lie Algebra has a representation on a matrix group.

Last but not least, according to Equation 2.85 if one knows a basis $T = \{T_a\}$ of the Lie Algebra (s)he can generate every element $g_i = g_i(\theta_1, \theta_2, \ldots, \theta_n) \in G$. That is why the elements $T_a$ are usually referred to as generators of the group.

A very important property of Lie Groups is that they can be represented as set of $n \times n$ square matrices. Knowing how to do this representation is really important for physics, because all Lie Groups that express symmetries of the physical systems can be thought as appropriate subsets square matrices. But first we need to define mathematically what a representation is.

**Definition 2.2.16** (Representation)**.** A representation $\mathcal{R}$ of a group $G$ is a map:

$$\mathcal{R}:\ G \longmapsto GL(n, \mathbb{C})$$

[25] For which group axioms require the following properties:

- if $g_i \in G \longmapsto M_j \in GL(n, \mathbb{C})$ then $g_i^{-1} \in G \longmapsto M_j^{-1} \in GL(n, \mathbb{C})$
- $e \in G \longmapsto \mathbb{1}_{n\times n} \in GL(n, \mathbb{C})$
- $g_1 \odot g_2 = g_3 \longmapsto M_1 M_2 = M_3$ where $g_1, g_2, g_3 \in G$ and $M_1, M_2, M_3 \in GL(n, \mathbb{C})$

The bullet points in the above definition are the conditions needed to show that a map $\mathcal{R}:\ G \longmapsto GL(n, \mathbb{C})$ is a representation of the group $G$. Note, also, that by definition the map $\mathcal{R}$ is not necessarily invertible. Two arbitrary group elements can be represented by the

[25] or $GL(n, \mathbb{R})$

same matrix element of $GL(n, \mathbb{C})$. If the map $\mathcal{R}$ is injective then the representation is called **faithful**.

Another important aspect of representations is that for a specific group $G$ there exist an infinite number of representations $\mathcal{R}$. More particularly, consider a representation $\mathcal{R}$ of a group $G$. We shall prove that the representation $\tilde{\mathcal{R}} = S\mathcal{R}S^{-1}$ where $S$ is a constant invertible matrix, is also a representation.The representations $\tilde{\mathcal{R}}$ and $\mathcal{R}$ are often called **equivalent representations**. Moreover, if $S$ happens to be a unitary matrix the term *unitarily equivalent* is used to describe the representations $\tilde{\mathcal{R}}$ and $\mathcal{R}$.

*Proof.* Because $\mathcal{R}$ is a representation of a group $G$, Definition 2.2.16 requires:

(i) $\mathcal{R}(e) = \mathbb{1}$

(ii) $\mathcal{R}(g_i^{-1}) = \mathcal{R}^{-1}(g_i),\ \forall g_i \in G$

(iii) if $g_3 = g_1 \odot g_2$, then $\mathcal{R}(g_3) = \mathcal{R}(g_1)\mathcal{R}(g_2)$

To prove that $\tilde{\mathcal{R}} = S\mathcal{R}S^{-1}$ is also a representation of $G$, one has to prove that it holds similar properties with the representation $\mathcal{R}$.

(i) $\tilde{\mathcal{R}}(e) = S\mathcal{R}(e)S^{-1} = S\mathbb{1}S^{-1} = \mathbb{1}$

(ii) $\tilde{\mathcal{R}}(g_i) = S\mathcal{R}^{-1}(g_i)S^{-1} = \tilde{\mathcal{R}}^{-1}(g_i)$

(iii) $\tilde{\mathcal{R}}(g_3) = S\mathcal{R}(g_3)S^{-1} = S\mathcal{R}(g_1)S^{-1}S\mathcal{R}(g_2)S^{-1} = \tilde{\mathcal{R}}(g_1)\tilde{\mathcal{R}}(g_2)$

□

The above proposition leads to the result that every group can have infinitely many representations, which brings on the obvious question regarding which representation contains all the information needed by a physicist to study a physical system.

**Definition 2.2.17** (Completely Reducible Representation)**.** A representation $\mathcal{R}$ of a group $G$ is called completely reducible if it is equivalent to the form:

$$S\mathcal{R}(g_i)S^{-1} = \begin{pmatrix} A(g_i) & \mathbb{0} \\ \mathbb{0} & B(g_i) \end{pmatrix},\ \forall g_i \in G \tag{2.86}$$

Opposing to the above definition a representation is called *irreducible (irrep)* if the matrices representing each $g_i \in G$ cannot be simultaneously written in the block-diagonal form of Equation 2.86.

All the above are basic definitions for representations of groups. To study a physical system, physicists are concerned with two main issues. Fist of all, they need to find which group expresses the symmetry of the system. After that, all the irreps of the group must be found. Thus, finding irreps is an aspect of great interest in physics. As a consequence, there are some representations that play a much greater role in representation theory than others.

**Definition 2.2.18** (Trivial Representation)**.** Let $G$ be a Lie Group. The trivial representation is just the mapping:

$$\mathcal{R}(g_i) = \mathbb{1},\ \ \forall g_i \in G \tag{2.87}$$

**Definition 2.2.19** (Fundamental Representation)**.** The fundamental representation of a matrix Lie Group is the map:

$$\mathcal{R}(g_i) = g_i, \ \ \forall g_i \in G \tag{2.88}$$

It is more practical, from a physical point of view, to define these representations via their act on vector spaces. For example, consider a scalar field $\phi = \phi(x_\mu) = (\phi_1(x_\mu), \dots \phi_N(x_\mu))$ living in an N-dimensional vector space. This field can be transformed only by $N \times N$ matrices, because that is what is allowed by matrix multiplication. So, if we need to transform it by a specific group we must choose a representation of dimension N. The fundamental representation of the group $SU(N)$ is an accepted choice, because in matrix notation an element $U \in SU(N)$ acts on the scalar vector field $\phi$ as:

$$\begin{bmatrix} \phi'_1 \\ \vdots \\ \phi'_N \end{bmatrix} = \begin{bmatrix} u_{11} & \dots & u_{1N} \\ \vdots & \ddots & \\ u_{n1} & & u_{NN} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_N \end{bmatrix}$$

This notation is time and space consuming, so the index notation $\phi'^i = U^i_j \phi^j$ is always preferred.

The idea of representations also applies to Lie Algebras, as expected from Equation 2.85. Because, a Lie Algebra is a vector space, it is sufficient to define a representation for its basis, the generators. The fundamental representation of a Lie Algebra is the set of $N \times N$ matrices that fulfill the conditions presented in Table 2.3 discussed in the next Section.

There is one more very important irreducible representation, which is really important for physics. It is easier to define this representation for a Lie Algebra and then translate the results for the corresponding Lie Group.

**Definition 2.2.20** (Adjoint Representation)**.** The adjoint representation of a Lie Algebra $\mathfrak{g}$ with generators $T = \{T^a\}, \ a = 1, \dots, \dim(\mathfrak{g})$ is the map:

$$\mathcal{R}(T^a) = \left(T^a_{\text{adj}}\right)^{bc} = -if^{abc}, \ \ \forall T^a \in T \tag{2.89}$$

where $f^{abc}$ are the structure constants of the Lie Algebra.

The indices $a, b, c$ in the above definition run in the set $\{1, 2, \dots, \dim(\mathfrak{g})\}$, where $\dim(\mathfrak{g})$ is the dimension of the group manifold (see Table 2.2), meaning that this representation makes it possible for the group and the algebra to act on fields "living" on vector spaces with dimension equal to $\dim(\mathfrak{g})$.

Last but not least, when transforming a vector $v$ under a Lie Group $G$ the following transformation law is implied:

$$v'^i = U^i_j v^j$$

where the element $U$ of the group $G$ is represented in the appropriate representation of the group that can act on $v$. The above equation can be also expressed in terms of Lie Algebra with the substitution $U = \exp\{i\theta^a T_a\}$.

$$v'^i = \left(e^{i\theta^a T_a}\right)^i_j v^j \tag{2.90}$$

Using a Taylor expansion $U = \mathbb{1} + i\theta^a T^a + \mathcal{O}(\theta^2)$, which in index notation is $U_{ij} = \delta_{ij} + i\theta^a (T^a)_{ij} + \mathcal{O}(\theta^2)$, the above expression can be written in an infinitesimal form:

$$v'^i = (\delta_{ij} + i\theta^a (T^a)_{ij})^i_j \, v^j, \quad |\theta_a| << 1 \tag{2.91}$$

where $i, j \in \{1, \ldots, \dim(\mathcal{R})\}$ [26] and $a \in \{1, \ldots, \dim(\mathfrak{g})\}$.

So, when physicists say that a field transform in the fundamental representation what they mean is that the generators $(T^a)^{ij}$ appearing in Equation 2.91 are in the fundamental representation. Of course, if the $(T^a)^{ij}$ are in the adjoint then the filed transforms in the adjoint.

### 2.2.5 Lie Algebras of Matrix Groups

In the previous Section, the main points of the connection between a Lie Algebra and a Lie Group were presented. The aim of this section is to collocate what is exactly the connection between the Matrix Lie Groups presented in Section 2.2.2 and their corresponding Lie Algebras. Actually, by knowing what the corresponding Lie Algebra of a Lie Group is, studying Lie Groups becomes a lot more easier because, all the elements of the group can be studied only by a finite set $T = \{T_a\}$ of generators.

First of all, in the previous section, two main properties for the structure constants of a Lie Algebra were proven only by using the definitive properties of the Lie Bracket included in Definition 2.2.14. More explicitly, Equation 2.80 shows that structure constants have to be antisymmetric and Equation 2.83 is a kind of Jacobi identity. The use of the word "antisymmetric" is kind of confusing, because structure constants $f^c_{ab}$ are just numbers, not matrices. The antisymmetry here refers to their to lower indices as Equation 2.80 implies. A spot that is left unanswered till this section is understanding and connecting to an actual operation between matrices the Lie Bracket.

So, to find what operation the Lie Bracket is in terms of matrices one should start by writing two different group elements $g_1, g_2 \in G$ as exponentials of elements of a Lie Algebra $g_1 = g_1(\theta_1, \ldots, \theta_n) = \exp\{i\theta^a T_a\}$ and $g_2 = g_1(\omega_1, \ldots, \omega_n) = \exp\Big\{i\omega^b T_b\Big\}$. Then, by demanding closure of the group, the element $g_3 = g_1 g_2 = \exp\{i\theta^a T_a\} \exp\Big\{i\omega^b T_b\Big\}$ should be in the group. From this last statement follows that the element $g_3 \in G$ should also be expressed as an exponential $g_3 = \exp\{i\lambda^c T_c\}$, where the parameters $\lambda^c$ will be expressed as functions of the parameters $\theta^a$ and $\omega^b$. The key to connect these three exponentials is the following proposition, broadly known as the *Baker-Campbell-Hausdorff (BCH)* formula.

**Proposition 2.2.1** (Baker-Campbell-Hausdorff (BCH) Formula)**.** *For any square $n \times n$ matrices $A, B$ the following equation holds:*

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\frac{1}{12}[A,[A,B]]-\frac{1}{12}[B,[A,B]]+\ldots} \tag{2.92}$$

[26] $\dim(\mathcal{R})$ is the dimension of the chosen representation for the Lie Group $G$ or the Lie Algebra $\mathfrak{g}$.

*where the notation* $[A, B]$ *is the commutator between the matrices* $A, B$ *defined as:*

$$[A, B] = AB - BA \tag{2.93}$$

*The omitted terms in Equation 2.92 are higher order commutators, that can always be expressed as function of the commutator* $[A, B]$

As mentioned several times above, the interest of this thesis restricts to matrix Lie Groups and their corresponding matrix Lie Algebras. Using the BCH formula for the group elements $g_1, g_2, g_3 \in G$ one gets:

$$e^{i\theta^a T_a} e^{i\omega^b T_b} = e^{i(\theta^a T_a + \omega^b T_b) + \frac{i^2}{2}\theta^a \omega^b [T_a, T_b]} \tag{2.94}$$

The right hand side of the above equation must be a group element because of closure, which leads to the conclusion that the commutator $[T_a, T_b] = T_a T_b - T_b T_a$ must be the Lie Bracket of the Lie Algebra.

In conclusion, every Lie Group has a corresponding Lie Algebra. The difference between the corresponding Lie Algebras of different Lie Groups, should originate from the definition of every Lie Group. The main purpose of this section is to analyze how the definitive conditions of every Matrix Lie Group presented in Section 2.2.2 are "translated" in conditions for the generators of the corresponding Lie Algebras.

The method for this "translation" begins by expressing arbitrary elements $g_i$ of every group in terms of the corresponding Lie Algebra via the exponential map of Equation 2.85 and then imposing the definitive condition of the Lie Group.

We will start from the $GL(n, \mathbb{R})$. Consider an element $A = A(\theta_1, \ldots, \theta_{n^2}) \in GL(n, \mathbb{R})$, where by Definition 2.2.6 $\det(A) \neq 0$. Now, according to Equation 2.94, the element $A$ can be written as $A = \exp\{i\theta^a T_a\}$, where $T_a$ are the generators of the corresponding Lie Algebra $\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$. The condition for the determinant is:

$$\det(A) = \det\left(e^{i\theta^a T_a}\right) \neq 0$$

A very useful property of linear algebra states that for every $n \times n$ (real or complex) square matrix the following equation holds:

$$\det\left(e^{A}\right) = e^{\mathrm{Tr}(A)}, \;\; \forall A \in M_{n\times n} \tag{2.95}$$

This property is the key identity to finish the calculation for $\det(A)$:

$$\det(A) = e^{\mathrm{Tr}(i\theta^a T_a)} = e^{i\theta^a \,\mathrm{Tr}(T_a)} \neq 0 \tag{2.96}$$

Of course, Equation 2.96 holds for every $T_a \in M_{n\times n}$, leading to the very interesting result that the corresponding Lie Algebra $\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$ of the Lie Group $GL(n, \mathbb{R})$ is the set of $n \times n$ square

matrices. This set also possesses a vector space structure as is required from the definition of an algebra.

In the case of an element $A = A(\theta_1, \ldots, \theta_{n^2-1}) \in SL(n, \mathbb{R})$ the condition of Definition 2.2.7 can be expressed as:

$$\det(A) = e^{\mathrm{Tr}(i\theta^a T_a)} = e^{i\theta^a \,\mathrm{Tr}(T_a)} = 1$$

The last equality is fulfilled only if:

$$i\theta^a \,\mathrm{Tr}(T_a) = i\left(\theta_1 \,\mathrm{Tr}\left(T^1\right) + \cdots + \theta_{n^2-1} \,\mathrm{Tr}\left(T^{n^2-1}\right)\right) = 0$$

The last equation must be true for every $(\theta_1, \ldots, \theta_{n^2-1}) \in \mathbb{R}^{n^2-1}$, because every element $A \in SL(n, \mathbb{R})$ must be described by the Lie Algebra. This means that the generators of the algebra $\mathfrak{sl}(n, \mathbb{R})$ are **traceless** matrices: $\mathrm{Tr}(T^i) = 0 \;\forall\; i \in 1, 2, \ldots, n^2 - 1$.

Considering, now, the orthogonal group $O(n)$. According to Definition 2.2.8 an element $A \in O(n)$ always fulfills the orthogonal condition $AA^T = A^\top A = \mathbb{1}$. Rewriting this property using the fact that $A^\top = A^{-1}$ and then applying the BCH formula (Equation 2.94) one gets:

$$\begin{aligned} AA^T &= \mathbb{1} \\ e^{i\theta^a T_a} e^{\left(i\theta^b T_b\right)^\top} &= \mathbb{1} \\ e^{i\theta^a T_a} e^{i\theta^b T_b^\top} &= \mathbb{1} \\ e^{i\left(\theta^a T_a + \theta^b T_b^\top\right) - \frac{1}{2}\left[\theta^a T_a, \theta^b T_b^\top\right]} &= \mathbb{1} \\ e^{i\theta^a\left(T_a + T_a^\top\right) - \frac{1}{2}\left[\theta^a T_a, \theta^b T_b^\top\right]} &= \mathbb{1} \end{aligned} \tag{2.97}$$

In order for Equation 2.97 to hold for every value the parameters $\theta = \{\theta_a\},\ a = 1, 2, \ldots, \frac{n(n-1)}{2}$ can take, the generators can be chosen to be be antisymmetric:

$$T_a = -T_a^\top,\ \forall a \in \{1, 2, \ldots, \frac{n(n-1)}{2}\} \tag{2.98}$$

With that choice Equation 2.97 is written:

$$e^{\frac{1}{2}\left[\theta^a T_a, \theta^b T_b\right]} = \mathbb{1}$$

By writing explicitly the sums of the above equation it can be easily seen that the equation holds for all antisymmetric matrices $T_a$.

$$\begin{aligned}
\Big[\theta^a T_a, \theta^b T_b\Big] &= \Big[\theta^1 T_1 + \theta^2 T_2 + \cdots + \theta^k T_k \,,\, \theta^1 T_1 + \theta^2 T_2 + \cdots + \theta^k T_k\Big] \\
\Big[\theta^a T_a, \theta^b T_b\Big] &= \Big[\theta^1 T_1, \theta^1 T_1\Big] + \Big[\theta^1 T_1, \theta^2 T_2\Big] + \Big[\theta^1 T_1, \theta^3 T_3\Big] + \cdots + \Big[\theta^1 T_1, \theta^k T_k\Big] + \\
&+ \Big[\theta^2 T_2, \theta^1 T_1\Big] + \Big[\theta^2 T_2, \theta^2 T_2\Big] + \Big[\theta^2 T_2, \theta^3 T_3\Big] + \cdots + \Big[\theta^2 T_2, \theta^k T_k\Big] + \cdots + \\
&+ \Big[\theta^k T_k, \theta^1 T_1\Big] + \Big[\theta^k T_k, \theta^2 T_2\Big] + \Big[\theta^k T_k, \theta^3 T_3\Big] + \cdots + \Big[\theta^k T_k, \theta^k T_k\Big]
\end{aligned}$$

The cross terms $[\theta^i T_i, \theta^j T_j]$ $i \neq j$ cancel out, because of the antisymmetry of the Lie Bracket: $[\theta^i T_i, \theta^j T_j] \;= -\,[\theta^j T_j, \theta^i T_i]$ $i \neq j$, meaning that $\Big[\theta^a T_a, \theta^b T_b\Big] = (\theta^a)^2\,[T_a, T_a] = \mathbb{0}$, which is true for every set of parameters $\theta = \{\theta_a\}$, $a = 1, 2, \ldots, \frac{n(n-1)}{2}$, because $[T_a, T_a] = \mathbb{0}$. This last equation validates that the generators of $\mathfrak{o}(\mathfrak{n})$ algebra must be antisymmetric $n \times n$ square matrices.

The Lie Algebra $\mathfrak{o}(\mathfrak{n})$ serves also as the corresponding Lie Algebra of the group $SO(n)$, because the condition $\det(A) = 1$, $\forall A \in SO(n)$ of Definition 2.2.9 translates to the condition $\mathrm{Tr}(T_i) = 0 \;\forall\; i \in 1, 2, \ldots, n^2 - 1$ for the generators, according to Equation 2.95. But, antisymmetric matrices are already traceless, meaning that no extra condition is introduced for the generators.

The condition referring to the generators of the algebras $\mathfrak{u}(\mathfrak{n})$ and $\mathfrak{su}(\mathfrak{n})$ are calculated following the same recipe. Using the equation of Definition 2.2.10 for an arbitrary element $U \in U(n)$:

$$\begin{aligned}
AA^\dagger &= \mathbb{1} \\
e^{i\theta^a T_a} e^{-i\theta^b T_b^\dagger} &= \mathbb{1} \\
e^{i\theta^a \left(T_a - T_a^\dagger\right) + \frac{1}{2}\left[\theta^a T_a, \theta^b T_b^\dagger\right]} &= \mathbb{1}
\end{aligned} \tag{2.99}$$

From Equation 2.99 follows that the generators of the Lie Algebra $\mathfrak{u}(\mathfrak{n})$ are Hermitian matrices $T_a = T_a^\dagger$. Same as before, $\mathfrak{u}(\mathfrak{n})$ coincides with the Lie Algebra $\mathfrak{su}(\mathfrak{n})$. The conditions for the elements of each Lie Algebra presented above is summarized in Table 2.3.

| Lie Group | Lie Algebra | Condition for $T_i$ |
|---|---|---|
| $GL(n,\mathbb{R})$ | $\mathfrak{gl}(\mathfrak{n},\mathbb{R})$ | $T_i \in M_{n\times n}$ |
| $SL(n,\mathbb{R})$ | $\mathfrak{sl}(\mathfrak{n},\mathbb{R})$ | $\mathrm{Tr}(T_i)=0$ |
| $O(n)$ | $\mathfrak{o}(\mathfrak{n})$ | $T_i=-T_i^\top$ |
| $SO(n)$ | $\mathfrak{so}(\mathfrak{n})$ | $T_i=-T_i^\top$ |
| $U(n)$ | $\mathfrak{u}(\mathfrak{n})$ | $T_i=T_i^\dagger$ |
| $SU(n)$ | $\mathfrak{su}(\mathfrak{n})$ | $T_i=T_i^\dagger$ |

**Table 2.3:** The corresponding Lie Algebras of the Lie Groups defined in Section 2.2.2. $T_i$ refers to the elements of the algebra.

# Chapter 3

# Gauge Theories

## 3.1 Abelian Gauge Theory

### 3.1.1 Gauge Invariance in Classical Electrodynamics

Gauge theories started from a simple property of Maxwell's equations for classical electrodynamics. In this section, we are going to illustrate this property and construct the first gauge theory, which possesses the famous $U(1)$ symmetry. The followung discussion is adapted from [13].

First of all, Maxwell's equations in natural units ($\hbar = c = 1$) are:

$$\nabla \cdot \vec{B} = 0 \tag{3.1}$$

$$\nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \tag{3.2}$$

$$\nabla \cdot \vec{E} = \varrho \tag{3.3}$$

$$\nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = \vec{j} \tag{3.4}$$

Equation 3.1 allows to write the magnetic field $\vec{B}$ as a function of a vector potential $\vec{A}$.

$$\vec{B} = \nabla \times \vec{A} \tag{3.5}$$

Substituting this new way of expressing the magnetic field in Equation 3.2 one gets the following result.

$$\nabla \times \left( \vec{E} + \frac{\partial \vec{A}}{\partial t} \right) = 0 \tag{3.6}$$

Now, we have found a quantity with zero curl. According to basic vector caclulus this quantity can be expressed as a function of a scalar potential.

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} = -\nabla V \tag{3.7}$$

Of course, the negative sign is just a common convention in classical electrodynamics. At this point the electric field $\vec{E}$ and the magnetic field $\vec{B}$ are both expressed as functions of the vector and scalar potential $\vec{A}$ and $V$. Substituting the electric field in Equation 3.3, we get an equation for the potentials $V, \vec{A}$.

$$\begin{aligned} \nabla \cdot \left( -\nabla V - \frac{\partial \vec{A}}{\partial t} \right) &= \varrho \\ \nabla^2 V + \frac{\partial}{\partial t}(\nabla \cdot \vec{A}) &= -\varrho \end{aligned} \tag{3.8}$$

Rewriting Equation 3.4 according to the potentials of Equations 3.5 and 3.7 we obtain another equation for the potentials:

$$\begin{aligned} \nabla \times (\nabla \times \vec{A}) - \frac{\partial}{\partial t} \left( -\nabla V - \frac{\partial \vec{A}}{\partial t} \right) &= \vec{j} \\ \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A} + \nabla \left( \frac{\partial V}{\partial t} \right) + \frac{\partial^2 \vec{A}}{\partial t^2} &= \vec{j} \\ \left( \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \vec{A} + \nabla \left( \nabla \cdot \vec{A} + \frac{\partial V}{\partial t} \right) &= \vec{j} \end{aligned} \tag{3.9}$$

The resulting Equations 3.8 and 3.9 form a system of partial differential equations with four degrees of freedom. The three coming from the components of the vector potential $\vec{A} = (A_1, A_2, A_3)$ and the fourth is just the scalar potential $V$. As a matter of fact, these two equations contain all the information of the full Maxwell' s equations, and simultaneously the number of degrees of freedom is reduced from six to four. So, a rational question should be "where did these extra two degrees of freedom go?". The answer to this question is the beginning of gauge theories.

The system of Equations 3.8 and 3.9 does not uniquely specify the potentials $\vec{A}$ and $V$ as they have four degrees of freedom specified only by two equations. So, let' s assume that we can substitute the potentials as:

$$\begin{aligned} \vec{A'} &= \vec{A} + \vec{c_A} \\ V' &= V + c_V \end{aligned}$$

where $\vec{c_A} = \vec{c_A}(t, \vec{x})$ and $c_V = c_V(t, \vec{x})$. In our assumption the new potentials should give the same electromagnetic field, meaning that:

$$\begin{aligned} \vec{B'} &= \vec{B} \\ \nabla \times \vec{A'} &= \nabla \times \vec{A} \\ \nabla \times \vec{A} + \nabla \times \vec{c_A} &= \nabla \times \vec{A} \\ \nabla \times \vec{c_A} &= 0 \end{aligned}$$

Again from basic vector calculus $c_A$ can be written as:

$$\vec{c_A} = \nabla \lambda$$

At this point the four degrees of freedom $(\vec{c_A}, c_V)$ in our initial assumption have been reduced to two $(\lambda, c_V)$. Demanding now, that the new potential must also give the same electric field and using Equation 3.7

$$\begin{aligned} \vec{E'} &= \vec{E} \\ \nabla V' + \frac{\partial \vec{A'}}{\partial t} &= \nabla V + \frac{\partial \vec{A}}{\partial t} \\ \nabla V + \nabla c_V + \frac{\partial \vec{A}}{\partial t} + \frac{\partial}{\partial t}(\nabla \lambda) &= \nabla V + \frac{\partial \vec{A}}{\partial t} \\ \nabla \left( c_V + \frac{\partial \lambda}{\partial t} \right) &= 0 \end{aligned}$$

The last equation means that the function $c_V + \frac{\partial \lambda}{\partial t}$ must be independent of the space coordinates:

$$\begin{aligned} c_V + \frac{\partial \lambda(t)}{\partial t} &= c(t) \\ c_V &= c(t) - \frac{\partial \lambda}{\partial t} \end{aligned}$$

It is more delicate to define a new function $\lambda'$, such that $\frac{\partial \lambda'}{\partial t} = c_V$, which according to the last equation is:

$$\begin{aligned} \frac{\partial \lambda'}{\partial t} &= c(t) - \frac{\partial \lambda}{\partial t} \\ \lambda' &= \int_0^t c(t) dt - \lambda \end{aligned}$$

Of course, we can at last just do a renaming $\lambda' \mapsto \lambda$. Finally, the new potentials $\vec{A'}, V$ can be written as:

$$\vec{A}' = \vec{A} + \nabla\lambda \tag{3.10}$$

$$V' = V + \frac{\partial\lambda}{\partial t} \tag{3.11}$$

This is exactly what is referred to as gauge symmetry. Maxwell' s equations remain invariant by the change instructed by Equations 3.10 and 3.11. So, the two degrees of freedom that disappeared are appearing again as the derivatives of the scalar function $\lambda$. By choosing appropriately this function $\lambda$ we can make Equations 3.8 and 3.9 easier to solve. The function $\lambda$ is called gauge and the process of choosing such a function is called gauge fixing. Of course, the solutions after gauge fixing are still valid in the general case because, as mentioned above, Maxwell' s equations are gauge invariant.

### 3.1.2 Lorentz Invariance of Classical Electrodynamics

Classical electrodynamics is the ancestor of modern gauge theories. As every physical theory it should be Lorentz invariant, or more generally frame invariant. So, it must be formulated using the notions of Section 2.1. Maxwell' s equations consist of two pairs of equations. Equations 3.3 and 3.4, which are inhomogeneous and the homogeneous pair 3.1, 3.2.

The purpose of this section is to rewrite Maxwell equations using a field that will transform homogeneously when passing from one frame to another. The first, critical observation is that the charge of a particle remains invariant when the frame of reference transforms or in plain words a moving particle will have the same charge as it would have whether considered stationary.

Secondly, Maxwell equations contain the continuity equation of charge. More specifically, the continuity equation appears by taking the divergence of Equation 3.4 and then substituting the charge density from Equation 3.3.

$$\nabla \cdot (\nabla \times \vec{B}) - \frac{\partial}{\partial t}\left\{\nabla \cdot \vec{E}\right\} = \nabla \cdot \vec{j}$$

$$\frac{\partial \varrho}{\partial t} + \nabla \cdot \vec{j} = 0 \tag{3.12}$$

The physical meaning of this equation is that the electric charge cannot be lost. As a result, any change of the electrical charge in a specific volume of space should be compensated by currents leaving or entering the volume. A more intuitive way to understand this physical interpretation is the integral form of Equation 3.12. The easiest way to obtain this form is by integrating the continuity equation over a 3-volume $V$ and then applying Gauss' theorem in the term containing the divergence of the current $\vec{j}$.

$$\int_V \frac{\partial \varrho}{\partial t} dx^3 + \int_V \nabla \cdot \vec{j} dx^3 = 0$$
$$\frac{\partial}{\partial t} \int_V \varrho dx^3 + \int_{\partial V} \vec{j} \cdot \hat{n}_S dx^2 = 0 \tag{3.13}$$

The physical interpretation of continuity equation, in combination with its integral form, which contains the integral of charge density, is a very strong argument to assume that the tuple $(\varrho, \vec{j})$ should be considered as a vector density in terms of tensor calculus of Section 2.1.3. If considered so, then the continuity equation can easily be derived by the divergence of this vector density, which, as proved in Section 2.1.4, is an invariant density. So, the first clue to rewrite Maxwell equations in terms of tensors is to consider the contrvariant vector density $\mathscr{J}^{\mu} = (\varrho, \vec{j})$. This fact oblige someone to look upon the inhomogeneous pair of Maxwell equations 3.4 and 3.3 as equations between vector densities.

The right hand side of Equations 3.4 and 3.3 contains the derivatives of the fields $\vec{E}$ and $\vec{J}$, but as mentioned, these derivatives should behave like a vector density as is $\mathscr{J}^{\mu}$. As proved in Section 2.1.4 the divergence of a contravariant scalar tensor density of rank-2, transforms as a scalar vector density. As a result, a logical assumption should be that the two inhomogeneous Maxwell equations can be written in the following form:

$$\frac{\partial \mathscr{F}^{\mu\nu}}{\partial x^{\nu}} = \mathscr{J}^{\mu} \tag{3.14}$$

Equation 3.14 contains one free index, which means that it is a system of 4 equations. By expanding this equation one gets the following 4 equations[1].

$$\frac{\partial \mathscr{F}^{11}}{\partial t} + \frac{\partial \mathscr{F}^{12}}{\partial x} + \frac{\partial \mathscr{F}^{13}}{\partial y} + \frac{\partial \mathscr{F}^{14}}{\partial z} = \varrho \tag{3.15}$$
$$\frac{\partial \mathscr{F}^{21}}{\partial t} + \frac{\partial \mathscr{F}^{22}}{\partial x} + \frac{\partial \mathscr{F}^{23}}{\partial y} + \frac{\partial \mathscr{F}^{24}}{\partial z} = j_x \tag{3.16}$$
$$\frac{\partial \mathscr{F}^{31}}{\partial t} + \frac{\partial \mathscr{F}^{32}}{\partial x} + \frac{\partial \mathscr{F}^{33}}{\partial y} + \frac{\partial \mathscr{F}^{34}}{\partial z} = j_y \tag{3.17}$$
$$\frac{\partial \mathscr{F}^{41}}{\partial t} + \frac{\partial \mathscr{F}^{42}}{\partial x} + \frac{\partial \mathscr{F}^{43}}{\partial y} + \frac{\partial \mathscr{F}^{44}}{\partial z} = j_z \tag{3.18}$$

Appropriate components of the tensor density $\mathscr{F}^{\mu\nu}$ must be chosen, in order for the above equations to produce the following 4 equations, which are the inhomogeneous Maxwell's equations expanded.

[1]Greek indices run in $\{1, 2, 3, 4\}$, where 1 corresponds to the times component $t$, and therefore $2, 3, 4$ correspond to the spatial components $x, y, z$.

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \varrho \tag{3.19}$$

$$-\frac{\partial E_x}{\partial t} + \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = j_x \tag{3.20}$$

$$-\frac{\partial E_y}{\partial t} + \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = j_y \tag{3.21}$$

$$-\frac{\partial E_z}{\partial t} + \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = j_z \tag{3.22}$$

By demanding, one to one correspondence between the previous equations, one can identify the components of $\mathscr{F}^{\mu\nu}$. A more convenient way to see these components is to represent this tensor on the group $GL(n, \mathbb{R})$. The matrix form also makes it easier to observe that $\mathscr{F}^{\mu\nu}$ is an antisymmetrical tensor density. Last but not least, as stated at the end of the previous chapter, in Minkowski spacetime there is no difference between tensor and tensor densities. So, from now on, there will be given no distinction between those two quantities and the symbol $F^{\mu\nu}$ replaces the old $\mathscr{F}^{\mu\nu}$.

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix} \tag{3.23}$$

Of course, the job is not finished, because Equation 3.14 combined with 3.23 only produces one pair of Maxwell equations. The homogeneous pair must be derived too and, of course, it cannot be derived by $F^{\mu\nu}$. So, we need a new tensorial quantity to derive this other pair. An easy guess is to use Equation 2.14 introduced in Section 2.1.3. According to this equation, a covariant tensor can be constructed from a tensor density, which in the present case is the same as the tensor $F^{\mu\nu}$. The newly constructed tensor is $\tilde{F}_{\mu\nu}$, which in matrix representation is the following.

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma} = \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z & E_y \\ -B_y & E_z & 0 & -E_x \\ -B_z & -E_y & E_x & 0 \end{pmatrix} \tag{3.24}$$

The new tensor can be extracted by the old one by replacing $\vec{B} \to -\vec{E}$ and $\vec{E} \to \vec{B}$. Mathematically speaking there is a great amount of details left. The new tensor is actually the hodge star dual of the first one. In terms of differential geometry the electromagnetic field tensor $F^{\mu\nu}$ is a 2-form, and so is the new dual tensor $\tilde{F}_{\mu\nu}$, but these details far outrun the purposes of this thesis. There are some nice references for them in the Bibliography section ([19], [1]).

Differentiating the dual tensor $\tilde{F}_{\mu\nu}$ the homogeneous pair of Maxwell' s equations appears. Moreover, the construction of these tensors, serves another really important cause. According

to Section 2.1.2 the products $F_{\mu\nu}F^{\mu\nu}$, $\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}$ are invariant quantities under frame transformations of the underlying spacetime manifold. As a result, the previous products are Lorentz invariant products, a remark that will be quite useful for the next section of this chapter.

According to all the above, the full set of Maxwell' s equations are equivalent to Equations 3.25 and 3.26.

$$\frac{\partial F^{\mu\nu}}{\partial x^{\mu}} = J^{\mu} \tag{3.25}$$

$$\frac{\partial \tilde{F}^{\mu\nu}}{\partial x^{\mu}} = 0 \tag{3.26}$$

Last but not least, a very important comment about this section concerns why the mathematical object of the tensor was chosen to describe the electromagnetic field. The electromagnetic tensor $F^{\mu\nu}$ contains all the information about the electromagnetic field, which is a measurable quantity. According to the principle of covariance every measurable quantity in physics should transform covariantly under the frame transformation, in order to avoid any ambiguity when measuring its numerical values.

Finally, closing this section, it should be mentioned that in Section 2.1.4 a way to construct a second rank tensor $F_{\mu\nu}$ was presented. This can be achieved by taking a certain combination of derivatives of a covariant vector $A_{\mu}$, as shown in the following equation.

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \tag{3.27}$$

### 3.1.3 The Lagrangian of Classical Electrodynamics

According to the Lagrangian Formalism of Classical Mechanics, the equations of motion for a field theory should arise by minimizing the action $S$ of the theory.

$$S = \int d^4x\mathcal{L} \tag{3.28}$$

where $\mathcal{L} = \mathcal{L}(q_i, \dot{q}_i, \nabla q_i)$ is a function depending only on the fields and their (time and spatial) derivatives, called the Lagrangian Density. The notation $q_i$ is used for the physical degrees of freedom. For a field theory $q_i$s are fields, which depend only on spacetime $q_i = q_i(x_{\mu})$. That is why the integration of the Lagrangian density is performed in spacetime ($d^4x$) and not only in the time component $dt$ as in classical mechanics.

According to subsection 3.1.1, the physical fields of electrodynamics are the potentials $V, \vec{A}$. It is convenient to define a 4-potential:

$$A^{\mu} = (A^1, A^2, A^3, A^4) = (V, \vec{A}) \tag{3.29}$$

or converting this contravariant vector to its dual covariant, using the Minkowski metric:

$$A_\mu = \eta_{\mu\nu} A^\nu = (A_1, A_2, A_3, A_4) = (-A^1, A^2, A^3, A^4) \tag{3.30}$$

Therefore the Lagrangian density $\mathcal{L}$, must be a function $\mathcal{L} = \mathcal{L}(A^\mu, \dot{A}^\mu, \nabla A^\mu)$ or in short-hand notation $\mathcal{L} = \mathcal{L}(A^\mu, \dot{A}^\mu, \partial_\nu A^\mu)$. As mentioned, the purpose of this subsection is to define the proper action $S$ that will produce the equations of motion for classical electrodynamics. Taking the variation of the action and demanding it to go to zero one gets[2]:

$$\begin{aligned}
\delta S &= 0 \\
\delta \int_\Omega d^4x \mathcal{L}(A^\mu, \partial_\nu A^\mu) &= 0 \\
\int_\Omega d^4x \delta\mathcal{L}(A^\mu, \partial_\nu A^\mu) &= 0 \\
\int_\Omega d^4x \left[ \frac{\partial \mathcal{L}}{\partial A^\mu} \delta A^\mu + \frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)} \delta(\partial_\nu A^\mu) \right] &= 0 \\
\int_\Omega d^4x \frac{\partial \mathcal{L}}{\partial A^\mu} \delta A^\mu + \int_\Omega d^4x \frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)} \partial_\nu(\delta A^\mu) &= 0
\end{aligned}$$

Integrating by parts the second integral of the last equation:

$$\int_\Omega d^4x \frac{\partial \mathcal{L}}{\partial A^\mu} \delta A^\mu + \int_\Omega d^4x \partial_\nu \left\{ \frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)} \delta A^\mu \right\} - \int_\Omega d^4x \partial_\nu \left\{ \frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)} \right\} \delta A^\mu = 0$$

The middle term according to the generalized Stokes' Theorem is the integral over the boundary $\partial\Omega$, where one can assume that this boundary is the time and spatial infinity where the fields go to zero.

$$\int_\Omega d^4x \left[ \frac{\partial \mathcal{L}}{\partial A^\mu} - \partial_\nu \left\{ \frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)} \right\} \right] \delta A^\mu = 0 \tag{3.31}$$

The last term must be independent of the way the fields $A_\mu$ are varied, which can be achieved if the following equation holds.

$$\frac{\partial \mathcal{L}}{\partial A^\mu} - \partial_\nu \left\{ \frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)} \right\} = 0 \tag{3.32}$$

Equation 3.32 is actually a set of equations (one for each value of $\mu = 0, 1, 2, 3$) known as Euler-Lagrange equations for classical electrodynamics. Note that the derivatives in these

[2]Einstein's summation convention is applied

equations are functional derivatives, where the functions $A^\mu$ and $\partial_\nu A^\mu$ are considered the independent variables of the theory.

As stated many times above Euler-Lagrange should produce the Equations of Motion of the classical theory of electromagnetism for every spacetime manifold, or limiting to the purposes of this thesis, the Lagrangian must be Lorentz invariant. The easiest way to achieve this is by using the invariant products $F_{\mu\nu}F^{\mu\nu}$, $\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}$ presented in Section 3.1.2.

In Minkowski spacetime structure, described by the metric $\eta_{\mu\nu}$, the tensors $F^\mu\nu$ and $\tilde{F}^{\mu\nu}$ can be transformed to covariant tensors, using the procedure bellow, where one can think of as each index transforms separately.

$$F_{\mu\nu} = \eta_{\mu\sigma}\eta_{\nu\kappa}F^{\sigma\kappa}$$

The previous conversion only changes the following components of the electromagnetic tensor.

$$\begin{array}{lll} F_{12} = -F^{12} & F_{13} = -F^{13} & F_{14} = -F^{14} \\ F_{21} = -F^{21} & F_{31} = -F^{31} & F_{41} = -F^{41} \end{array}$$

Because this transformation does not make any difference whether the tensors are used with lower or upper indices, it is usual to give no interest in the distinction between covariant and contravariant quantities when referring to Minksowski spacetime. This remark makes the calculations a lot simpler, and of course much more complicated if one wants to formulate the theory over curved spacetime manifolds.

At first let' s make the ansatz that the Lagrangian is:

$$\mathcal{L} = F_{\mu\nu}F^{\mu\nu} = (\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) = 2(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu)$$

or making no distinctions between covariant and contravariant vectors according to the above remark the Lagrangian density takes the following simpler form:

$$\mathcal{L} = 2[(\partial_\mu A_\nu)^2 - \partial_\mu A_\nu \partial_\nu A_\mu] \tag{3.33}$$

then the terms appearing in the Euler-Lagrange equations 3.32 give:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial A_\mu} &= 0 \\
\frac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)} &= 2\left[2(\partial_\kappa A_\lambda)\frac{\partial(\partial_\kappa A_\lambda)}{\partial(\partial_\nu A_\mu)} - \frac{\partial(\partial_\kappa A_\lambda)}{\partial(\partial_\nu A_\mu)} - (\partial_\kappa A_\lambda)\frac{\partial(\partial_\lambda A_\kappa)}{\partial(\partial_\nu A_\mu)}\right] = \\
&= 2\left[2(\partial_\kappa A_\lambda)\delta^\nu_\kappa\delta^\mu_\lambda - (\partial_\lambda A_\kappa)\delta^\nu_\kappa\delta^\mu_\lambda - (\partial_\kappa A_\lambda)\delta^\nu_\lambda\delta^\mu_\kappa\right] = \\
&= 2[2\partial_\nu A_\mu - \partial_\mu A_\nu - \partial_\mu A_\nu] = \\
&= 4F_{\nu\mu} \\
\frac{\partial \mathcal{L}}{\partial(\partial_\kappa A_\lambda)} &= -4F_{\mu\nu}
\end{aligned}$$

where the antisymmetric property of the electromagnetic tensor was used in the last line.

The resulting Equations of Motion are:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial A^\mu} - \partial_\nu\left\{\frac{\partial \mathcal{L}}{\partial(\partial_\nu A^\mu)}\right\} &= 0 \\
\partial_\nu\{-4F_{\mu\nu}\} &= 0
\end{aligned}$$

This is exactly the derivative introduced in the right hand side of Equation 3.25. So, the invariant product $F_{\mu\nu}F^{\mu\nu}$ is correct but except for a factor $-4$ appearing upfront, which can easily be eliminated by reforming the Lagrangian as $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$. This term is known as the kinetic term of Maxwell's Lagrangian. Now, in order to produce the full Equation 3.25, an additional term must added to the Lagrangian, that will derive the term $J^\mu$ of the right hand side. The easiest way of adding this term is with the invariant product $A_\mu J^\mu$

$$\begin{aligned}
\frac{\partial}{\partial A_\mu}\{A_\mu J^\mu\} &= J^\mu \\
\frac{\partial}{\partial(\partial_\nu A_\mu)}\{A_\mu J^\mu\} &= 0
\end{aligned}$$

Combining these two terms to correctly produce the inhomogeneous pair of Maxwell' s equations, represented by Equation 3.25, one gets the Lagrangian:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - A_\mu J^\mu \tag{3.34}$$

Last but not least, the homogeneous pair of Maxwell' s equation should be derived from the Lagrangian, but a really extraordinary result is that the electromagnetic tensor already contains the homogeneous equations, hidden in its antisymmetric properties. One can from Equation 3.26 substitute the definition of the dual tensor $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. This results in a nice Bianchi identity.

$$\partial_\mu F_{\nu\rho} + \partial_\rho F_{\mu\nu} + \partial_\nu F_{\rho\mu} = 0 \tag{3.35}$$

As a result the Lagrangian 3.34 is the full Lagrangian needed to derive the Equations of Motion for the classical theory of Electromagnetism, which, as it should, is a Lorentz invariant quantity, because of the invariant products.

The gauge invariance of this theory demands that the electromagnetic potential 4-vector $A_\mu$ transforms in shorthand covariant notation as:

$$A_\mu \to A_\mu + \partial_\mu \lambda \tag{3.36}$$

where the function $\lambda$ is a smooth function of the spacetime coordinates.

### 3.1.4 Gauge Theories

Gauge theory originate from the classical theory of electrodynamics as shown in Section 3.1.1, but its great significance is observed when the electromagnetic theory is coupled with a spinor field. The physical meaning of this coupling is the description of the motion of a charged fermion (as is the electron) in an electromagnetic field. This is the Lagrangian of Quantum Electrodynamics constructed using the kinetic term of Maxwell's theory with a Dirac Lagrangian for a spin-1/2 particle, which takes the place of the electromagnetic current.

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \tag{3.37}$$

The terms $\psi, \bar{\psi}$ are Dirac' s spinor fields, of in more mathematical terminology they are Grassman fields, describing the relativistic behaviour of the fermion, through spacetime. The term $\gamma^\mu$ represent the Dirac' s matrices, which obey the Clifford' s algebra and transform as a 4-vector.

The previous equation, does not contain any interaction between the electromagnetic field and the fermionic component, meaning there is no term containing simultaneously $\psi$ and $A_\mu$. This interaction, which is a physical demand for the system, will arise naturally from the gauge invariance of the system.

To introduce this gauge invariance a general result from Quantum Mechanics can be used. If one wants to use this Lagrangian to describe quantum mechanical effects of a fermion inside an electromagnetic field, the fermion should obey the quantum mechanical laws. This is implied by saying that it is represented by a Dirac' s spinor, because Dirac' s equation is the relativistic analogue of Schrodinger' s equation.

That being said, in quantum mechanics the complex phase appearing in front of a wave function of particle is insignificant. Meaning, that for an arbitrary real parameter $\theta \in \mathbb{R}$, the spinor field $\psi$ transforms as:

$$\psi \to e^{iq\theta}\psi \tag{3.38}$$

and the dual field $\bar{\psi} = \psi^\dagger\gamma^0$, where $\psi^\dagger$ is the Hermitian conjugate of $\psi$, must transform as:

$$\bar{\psi} \to \bar{\psi}e^{-iq\theta} \tag{3.39}$$

These transformations leave the Lagrangian 3.37 untouched, which is a nice property but it does not serve the initial physical demand to provide the coupling between electromagnetism and fermions. The parameter $q$ is just a real number. Its physical significance will become clear at the end of this Section.

It is another quantum mechanical result, that the parameter $\theta$ shall not be restricted to be a global parameter, where the term global parameter is being used to describe a real number independent of the point of spacetime the fermion field is evaluated. The above transformation rules must hold and keep the Lagrangian invariant even if the parameter theta is a smooth real function $\theta = \theta(x) = \theta(x^\mu), \forall x^\mu \in \mathbb{R}^{1,3}$[3]. In bibliography, the use of the function $g(x)$ is often called a local transformation.

What is implied in the above paragraph is that the transformation of the field $\psi$, that should leave the Lagrangian 3.37 is given by a function $g(x) = \exp\{iq\theta(x)\}$. Observing this function, is can be seen that is a special case of the exponential map, described in Section 2.2.4, that connects each element of a Lie Group with an element of a corresponding Lie Algebra. In this specific case the function $g$ maps an arbitrary point of spacetime with an element of the group $U(1)$. Therefore it is:

$$g : \mathbb{R}^{1,3} \longrightarrow U(1) \tag{3.40}$$
$$g(x) = e^{iq\theta(x)} \tag{3.41}$$

It is obvious that the mass term of the Lagrangian 3.37 remains the same after the local transformation. The kinetic term, though, transforms explicitly as:

$$\begin{aligned} i\bar{\psi}'\gamma^\mu\partial_\mu\psi' &= i\bar{\psi}e^{-iq\theta(x)}\gamma^\mu\partial_\mu(e^{iq\theta(x)}\psi) \\ &= i\bar{\psi}e^{-iq\theta(x)}\gamma^\mu\partial_\mu(e^{iq\theta(x)})\psi + i\bar{\psi}e^{-iq\theta(x)}\gamma^\mu e^{iq\theta(x)}(\partial_\mu\psi) \\ &= -q\bar{\psi}e^{-iq\theta(x)}e^{iq\theta(x)}\gamma^\mu\partial_\mu(\theta(x))\psi + i\bar{\psi}\gamma^\mu\partial_\mu\psi \\ i\bar{\psi}'\gamma^\mu\partial_\mu\psi' &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - q\bar{\psi}\gamma^\mu\partial_\mu(\theta(x))\psi \end{aligned}$$

The transformation implied gave an extra term in the Lagrangian that has to be eliminated. The easiest way to eliminate this term is by adding the extra term $q\bar{\psi}\gamma^\mu\partial_\mu(\theta(x))\psi$ to the

[3]The notation $\mathbb{R}^{1,3}$ is used for a Minkowksi spacetime manifold with metric $(-,+,+,+)$.

Lagrangian. But, of course, this term is not physical. The big idea is tha this term can be introduced naturally by the transformation implied in equations 3.40, 3.41.

The idea lies on the gauge invariance of classical electrodynamics. Specifically, the kinetic term of Maxwell' s Lagrangian remains intact if the derivative of a real valued function $\theta(x)$, is added to the potential $A_\mu$. This is exactly what is written in Equation 3.36 of the previous section.

Thus, the Lagrangian need to be modified, in order to contain a term that will be linear in the potential $A_\mu$. Of course, the extra term must remain Lorentz invariant, which implies the obligatory use of the invariant product $\gamma^\mu A_\mu$.

The necessary term is:

$$V_{int} = q\bar{\psi}\gamma^\mu A_\mu \psi \tag{3.42}$$

This term, has the physical meaning of the interaction[4] between the fermion and the electromagnetic field, which was one of the basic requirements of the theory.

If, under the transformation $g$ the potential $A_\mu$ transforms as:

$$A_\mu \to A_\mu + \partial_\mu \theta(x) \tag{3.43}$$

then the transformation of the extra term $V_{int}$ reads:

$$\begin{aligned} V'_{int} &= q\bar{\psi}e^{-iq\theta(x)}\gamma^\mu[A_\mu + \partial_\mu\theta(x)]e^{iq\theta(x)}\psi \\ &= q\bar{\psi}\gamma^\mu A_\mu\psi + q\bar{\psi}\gamma^\mu\partial_\mu\theta(x)\psi \\ V'_{int} &= V_{int} + q\bar{\psi}\gamma^\mu(\partial_\mu\theta(x))\psi \end{aligned}$$

This concludes that the following Lagrangian, which is the Lagrangian of Quantum Electrodynamics (QED) is invariant under the transformation implied by the map $g$ and also Lorentz invariant. The extra term not only serves mathematical purposes, but also physical, because it models the interaction between the two fields, implying that physical fields are the gauge invariant ones.

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma^\mu(\partial_\mu - iqA_\mu)\psi - m\bar{\psi}\psi \tag{3.44}$$

In the above Lagrangian the use of the parentheses to group the kinetic term for the fermion with $V_{int}$ is not by chance. There is a very deep geometrical meaning here. The field $\psi = \psi(x)$ is a function defined on Minkowski spacetime $\mathbb{R}^{1,3}$, which takes values on $\mathbb{C}^4$. After the transformation dictated by the function $g$ the field $\psi$ takes the new value $g\psi \in \mathbb{C}^4$.

[4]That explains the index "int" in Equation 3.42.

This reminds the concepts of differential geometry introduced in Chapter 2, with the use of diffeomorphisms $x' : \mathbb{R}^{1,3} \to \mathbb{R}^{1,3}$. In the present case the transformation is still a diffeormorphism but between different manifolds as one can see in Equation 3.40. So, if one considers the field $\psi$ as a vector then the derivative of this field $\partial_\mu \psi$ does not transform the same way as the vector. This was the reason the initial Lagrangian 3.37, did not remain invariant after the transformation $g$. This is a very rational result, the map $g$ takes a different value for every point in spacetime. As a result the transformed field $g\psi$ changes by a different factor. So, the ordinary derivative is not enough to compare the real change of the field $\psi$.

In more mathematical thinking the ordinary derivative of the field transforms inhomogeneously and this, as explained in Chapter 2 is not a correct behaviour for a derivative. The essence of the previous sentence is that when a derivative of a field is zero, the field is constant. But with the presence of the inhomogeneous term the derivative can go to zero with a non-constant field. With the addition of the new term $V_{int}$ in the Lagrangian, the functional remains invariant after the transformation, because the term appearing in the parentheses of 3.44 transforms homogeneously with the transformation $g$, hence the term in the parentheses is the definition for the covariant derivative. In more formal notation the covariant derivative is defined as:

$$\nabla_\mu = \partial_\mu - iqA_\mu \tag{3.45}$$

and when acting on a fermionic field $\psi$, it transforms homogeneously under the gauge transformation $g$ as shown bellow. According to Chapter 2.1.4 and especially Equation 2.6 the electromagnetic potential $A_\mu$ plays the role of connection and because it is the connection of gauge transformation it is commonly referred to as gauge field. The physical meaning of seeing the 4-vector potential as a connection is that when a fermion moves through spacetime with the presence of electromagnetic field, it interacts differently in every point. Of course, the parameter $q$ that appears in front of the connection in Equation 3.45 and on the exponential of the gauge transformation dictates how strongly the fermion interacts with the electromagnetic field. Therefore, $q$ is the electric charge.

$$\begin{aligned}
\nabla'_\mu \psi' &= \nabla'_\mu (g\psi) \\
&= (\partial_\mu - iqA'_\mu)(e^{iq\theta(x)}\psi) \\
&= \partial_\mu (e^{iq\theta(x)}\psi) - iq(A_\mu + \partial_\mu \theta(x))(e^{iq\theta(x)}\psi) \\
&= e^{iq\theta(x)}(\partial_\mu \psi) + iqe^{iq\theta(x)}(\partial_\mu \theta(x))\psi - iqA_\mu e^{iq\theta(x)}\psi - iqe^{iq\theta(x)}(\partial_\mu \theta(x))\psi \\
&= e^{iq\theta(x)}(\partial_\mu \psi - iqA_\mu \psi) \\
\nabla'_\mu \psi' &= g\nabla_\mu \psi
\end{aligned}$$

Following the geometrical point of view and having defined the covariant derivative, it is straightforward to define the curvature as the commutator of the covariant derivatives. As shown bellow, the curvature is proportionate to the electromagnetic field tensor $F_{\mu\nu}$.

$$
\begin{aligned}
[\nabla_\mu, \nabla_\nu]\psi &= \nabla_\mu(\nabla_\nu\psi) - \nabla_\nu(\nabla_\mu\psi) \\
&= \partial_\mu(\nabla_\nu\psi) - iqA_\mu\nabla_\nu\psi - \partial_\nu(\nabla_\mu\psi) + iqA_\nu\nabla_\mu\psi \\
&= \partial_\mu(\partial_\nu\psi - iqA_\nu\psi) - iqA_\mu(\partial_\nu\psi - iqA_\nu\psi) - \\
&- \partial_\nu(\partial_\mu\psi - iqA_\mu\psi) + iqA_\nu(\partial_\mu\psi - iqA_\mu\psi) \\
&= -iq(\partial_\mu A_\nu\psi) - iqA_\mu\partial_\nu\psi + iq(\partial_\nu A_\mu\psi) + iqA_\nu\partial_\mu\psi \\
&= -iq(\partial_\mu A_\nu)\psi + iq(\partial_\nu A_\mu)\psi \\
[\nabla_\mu, \nabla_\nu]\psi &= -iqF_{\mu\nu}\psi
\end{aligned} \tag{3.46}
$$

The previous discussion refers to the U(1) gauge theory, which is the symmetry behind QED. QED is a really important, elegant and successful theory of modern physics. Its success made it a model theory, that inspired a lot of new modifications. A very simple one is to change the target of the transformation $g$, to be an arbitrary Lie Group and not specifically U(1). This change is the beginning of Yang-Mills theory and it truly contains a lot of deep physical and mathematical results [31].

## 3.2 Classical Yang-Mills Theory

### 3.2.1 The Covariant Derivative and the Connection of Yang-Mills Theory

Yang-Mills theory stands on the modification of the $U(1)$ theory. The change Yang and Mills investigated is altering the gauge transformation $g$, by changing the target set to be an arbitrary Lie Group $G$, with Lie Algebra $\mathfrak{g}$. The map $U(x)$ takes each point $x \in \mathbb{R}^{1,3}$ and match it with an element $g \in G$, which as explained in Chapter 2.2.4, can be written as an exponential map of its Lie Algebra with generators $T_a$. The coefficients $\theta_a(x)$ are responsible for mapping the different points $x \in \mathbb{R}^{1,3}$ to different elements of the group $U \in G$.

$$U : \mathbb{R}^{1,3} \longrightarrow G \tag{3.47}$$

$$U(x) = e^{ig\theta_\alpha(x)T^\alpha} \tag{3.48}$$

More concretely, in Yang-Mills theory the group $G$ can be an arbitrary Lie Group, which in the general case can be a non-Abelian group. The group $G$ is often called the structure or the gauge group of the theory. The appropriate mathematical theory that studies Yang-Mills equations is the theory of principal bundles. The following results are not extremely mathematically rigour, but they do remain correct.

Following the same reasoning as in Chapter 2.1.4, the covariant derivative of Yang-Mills theory can be established in two ways. Firstly, the effect of the local gauge transformation $U$ on a field $\psi$ is:

$$\psi \to U\psi = e^{ig\theta_\alpha(x)T^\alpha}\psi \tag{3.49}$$

This transformation is not as simple as it was for the Abelian U(1) case. As established, the exponential is equal to a different group element in every spacetime point and of course this cannot in general "multiply" the map $\psi$. For example, if the group $G$ is chosen to be the group SU(3), how these 3x3 would act on the map $\psi$ that creates values that live on $\mathbb{C}^4$. The answer is that the group and the map $\psi$ are represented on a representation of the same dimension. The representation is realised from the generators $T_a$ of the Lie Algebra, so what should actually be written Equation 3.49 is $\mathcal{R}(T^a)$, but this is always omitted in the bibliography.

The transformation 3.49 can also be written in its infinitesimal version, using the Taylor expansion around infinitesimal functions $\theta^a(x)$ with $\|\theta\| \to 0^+$.

$$U = \exp\{ig\theta^a(x)T_a\} = \mathbb{1} + ig\theta^a T_a + \mathcal{O}(g^2, \theta^2) \tag{3.50}$$
$$\psi \to U\psi = \psi + ig\theta^a T_a\psi \tag{3.51}$$

The covariant derivative $\nabla_\mu$ must transform in the same way when it acts on a field $\psi$, meaning that it has to absorb the extra terms arising from ordinary differentiation. As described in Chapter 2.1.4 the covariant derivative has to contain an extra field that does not transform homogeneously[5], and its role is to absorb the non-homogeneous terms appearing in the ordinary derivative.

Let $A_\mu$ be the field that will compensate the non-homogeneous terms. Then the covariant derivative can be defined analogously to the Abelian gauge theory as:

$$\nabla_\mu = \partial_\mu - igA_\mu \tag{3.52}$$

The covariant derivative of Equation 3.52 will act on the field $\psi$, which as explained above is represented in an appropriate representation, such that each group element $U$ can act on. As a result, the field $A_\mu$ must be represented in this representation, and because the generators $T_a$ are a basis of this representation, the compensating field $A_\mu$ can be expanded on that basis as $A_\mu = A_\mu \cdot T = A_\mu^\alpha T^\alpha$.[6] Rigorously speaking, the field $A_\mu$ is a map from Minkowski space $\mathbb{R}^{1,3}$ to the Lie Algebra of the underlying group. The covariant derivative can be written as:

$$\nabla_\mu = \partial_\mu \mathbb{1} - igA_\mu^a T_a \tag{3.53}$$

Again, the correct notation is $\mathcal{R}(A_\mu)$ instead of just $A_\mu$, where the representation depends on the object that the derivative acts on. The coefficient $g$, which has taken the place of the electric charge $q$, is often referred to as the coupling constant. It is just a real number

[5]This result will be reproduced for the covariant derivative of Yang-Mills theory.

[6]The repeated index $a$ implies summation.

that describes how strong is the interaction of the field $\psi$ with the compensating field $A_\mu$. This last remark will be clear by the end of the present section. Using this definition the transformed covariant derivative acting on the transformed field is:

$$\begin{aligned}
\nabla'_\mu \psi' &= U \nabla_\mu \psi \\
(\partial_\mu - igA'_\mu)\psi' &= U \nabla_\mu \psi \\
(\partial_\mu U\psi) - igA'_\mu U\psi &= U \nabla_\mu \psi \\
(\partial_\mu U)\psi + U\partial_\mu \psi - igA'_\mu U\psi &= U\partial_\mu \psi - igUA_\mu \psi
\end{aligned} \tag{3.54}$$

After some simple algebra and demanding that the last equation should hold independently of the chosen field $\psi$ the result is Equation 3.55.

$$\begin{aligned}
A'_\mu U &= UA_\mu - \frac{i}{g}\partial_\mu U \\
A'_\mu U U^{-1} &= UA_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1} \\
A'_\mu &= UA_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1}
\end{aligned} \tag{3.55}$$

The last equation, shows how the field $A_\mu$ should transform in order for the covariant derivative, acting on a field $\psi$, to transform as the field $\psi$ itself. Using the Taylor expansion around the identity element $\mathbb{1} \in G$, one can find how the components $A^a_\mu$ transform.

Taking the first order approximation $\exp\{ig\theta^a(x)T_a\} = \mathbb{1} + ig\theta^a T_a + \mathcal{O}(g^2, \theta^2)$ with $\|\theta\| \to 0^+$ Equation 3.55 becomes:

$$\begin{aligned}
A^{a'}_\mu T_a &= (\mathbb{1} + ig\theta^a T_a)A^a_\mu T_a(\mathbb{1} - ig\theta^a T_a) - \frac{i}{g}\partial_\mu(\mathbb{1} + ig\theta^a T_a)(\mathbb{1} - ig\theta^a T_a) \\
&= (A^a_\mu T_a + ig\theta^a T_a A^b_\mu T_b)(\mathbb{1} - ig\theta^a T_a) + (\partial_\mu \theta^a T_a)(\mathbb{1} - ig\theta^a T_a) \\
&= A^a_\mu T_a - igA^a_\mu T_a \theta^b T_b + ig\theta^a T_a A^b_\mu T_b + \underbrace{g^2 \theta^a T_a A^b_\mu T_b \theta^c T_c}_{\mathcal{O}(\theta^2)} + \\
&\quad + \partial_\mu \theta^a T_a - \underbrace{ig(\partial_\mu \theta^a T_a)\theta^b T_b}_{\mathcal{O}(\theta^2)} \\
&= \partial_\mu \theta^a T_a + ig[\theta^a T_a, A^b_\mu T_b] + A^a_\mu T_a \\
&= \partial_\mu \theta^a T_a + ig\theta^a [T_a, T_b]A^b_\mu + A^a_\mu T_a \\
A^{a'}_\mu T_a &= \partial_\mu \theta^a T_a - g\theta^a A^b_\mu f^{abc} T_c + A^a_\mu T_a
\end{aligned} \tag{3.56}$$

In the last equation the formula $[T_a, T_b] = if^{abc}T_c$ was used, which differs from Equation 2.79 presented in Chapter 2.2.5 only by the $i$ factor extracted upfront, which is common in theoretical physics.

Another important form of the infinitesimal transformation of the connection can be formed from Equation 3.56. The advantage of the following equation is its independence on the generators $T_a$.

$$\delta A^a_\mu = A^{a'}_\mu - A^a_\mu = \partial_\mu \theta^a - g\theta^c A^b_\mu f^{cba} \tag{3.57}$$

### 3.2.2 The Curvature of Yang-Mills Theory

The next step to formulate properly the non-Abelian theory, is to compute the commutator of the covariant derivatives, which according to Equation 3.46 is proportional to the field strength tensor of the theory.

$$\begin{aligned}
[\nabla_\mu, \nabla_\nu]\psi &= (\partial_\mu - igA_\mu)(\partial_\nu\psi - igA_\nu\psi) - (\partial_\nu - igA_\nu)(\partial_\mu\psi - igA_\mu\psi) \\
&= \partial_\mu\partial_\nu\psi - ig(\partial_\mu A_\nu\psi) - igA_\mu\partial_\nu\psi - g^2A_\mu A_\nu\psi - \\
&- \partial_\nu\partial_\mu\psi + ig(\partial_\nu A_\mu\psi) + igA_\nu\partial_\mu\psi + g^2A_\nu A_\mu\psi \\
&= -ig(\partial_\mu A_\nu)\psi - igA_\nu\partial_\mu\psi - igA_\mu\partial_\nu\psi - g^2A_\mu A_\nu\psi + \\
&+ ig(\partial_\nu A_\mu)\psi + igA_\mu\partial_\nu\psi + igA_\nu\partial_\mu\psi + g^2A_\nu A_\mu\psi \\
&= \left\{-ig(\partial_\mu A_\nu - \partial_\nu A_\mu) - g^2[A_\mu, A_\nu]\right\}\psi \\
&= -ig\left\{(\partial_\mu A_\nu - \partial_\nu A_\mu) - ig[A_\mu, A_\nu]\right\}\psi \\
[\nabla_\mu, \nabla_\nu]\psi &= -igF_{\mu\nu}\psi
\end{aligned} \tag{3.58}$$

As expected from the comparison with the Abelian gauge theory the term inside the curly brackets, named $F_{\mu\nu}$, is just a function of the connection $A_\mu$ and not an operator as was the covariant derivative at first. This is the curvature or the field strength tensor of the non-Abelian gauge theory. Again the coupling constant $g$ appears in front of $F_{\mu\nu}$, which shows that the curvature of the field is dependent on this constant. If this constant takes bigger values then the covariant derivatives fail even more to commute, which results in bigger curvature.

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] \tag{3.59}$$

An easily seen fact for the curvature in the view of relativity is that the first two terms constitute the covariant curl of a four vector, which, as stated many times, is Lorentz covariant. The commutator of the connection with itself is just a product of two covariant 4-vectors, meaning that transforms as a covariant tensor of the second rank. As a wholesome the tensor $F_{\mu\nu}$ is indeed a covariant 2-tensor.

Yet, in the point of view of gauge theories the transformation law of the curvature must be investigated under the gauge transformation. For this investigation the rule of transformation for the covariant derivatie will be used as $\nabla'_\mu = U\nabla_\mu U^{-1}$

$$
\begin{aligned}
F_{\mu\nu} &= \frac{i}{g}[\nabla_\mu, \nabla_\nu] && (3.60)\\
F'_{\mu\nu} &= \frac{i}{g}(\nabla'_\mu \nabla'_\nu - \nabla'_\nu \nabla'_\mu) \\
&= \frac{i}{g}(U\nabla_\mu \underbrace{U^{-1}U}_{\mathbb{1}} \nabla_\nu U^{-1} - U\nabla_\nu \underbrace{U^{-1}U}_{\mathbb{1}} \nabla_\mu U^{-1}) \\
&= \frac{i}{g}U(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)U^{-1} \\
F'_{\mu\nu} &= \frac{i}{g}U[\nabla_\mu, \nabla_\nu]U^{-1} = UF_{\mu\nu}U^{-1} && (3.61)
\end{aligned}
$$

The last equation shows that the curvature transforms homogeneously under the gauge transformation. In fact, the transformation dictated by Equation 3.61 is the analogue of the homogeneous transformation for the second rank mixed tensor of general relativity.

The tensor $F_{\mu\nu}$ depends on the connection $A_\mu$, which is Lie Algebra valued, therefore the tensor takes values on the Lie Algebra as well and has its own components $F^a_{\mu\nu}$ for expansion on the basis $T^a$. The transformation formula for these components can be obtained by Equation 3.61. First the components can be calculated as follows.

$$
\begin{aligned}
F^a_{\mu\nu}T_a &= \partial_\mu A^a_\nu T_a - \partial_\nu A^a_\mu T_a - ig[A^a_\mu T_a, A^b_\nu T_b] \\
&= \partial_\mu A^a_\nu T_a - \partial_\nu A^a_\mu T_a - igA^a_\mu A^b_\nu [T_a, T_b] \\
&= \partial_\mu A^a_\nu T_a - \partial_\nu A^a_\mu T_a + gA^c_\mu A^b_\nu f^{cba} T_a \\
F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu && (3.62)
\end{aligned}
$$

The infinitesimal rule of transformation can be obtained as before using a Taylor' s expansion around some small parameters $\theta^a(x)$. Using Equation 3.61:

$$
\begin{aligned}
F'_{\mu\nu} &= (\mathbb{1} + ig\theta^a T_a)F_{\mu\nu}(\mathbb{1} - ig\theta^b T_b) \\
&= (F_{\mu\nu} + ig\theta^a T_a F_{\mu\nu})(\mathbb{1} - ig\theta^b T_b) \\
&= F_{\mu\nu} - igF_{\mu\nu}\theta^b T_b + ig\theta^a T_a F_{\mu\nu} + \underbrace{g^2\theta^a T_a F_{\mu\nu}\theta^b T_b}_{\mathcal{O}(\theta^2)} \\
&= F_{\mu\nu} + ig(\theta^a T_a F^b_{\mu\nu} T_b - F^a_{\mu\nu} T_a \theta^b T_b) \\
&= F_{\mu\nu} + ig(F^a_{\mu\nu}\theta^b T_b T_a - F^a_{\mu\nu}\theta^b T_a T_b) \\
F'^a_{\mu\nu}T_a &= F^a_{\mu\nu}T_a - igF^a_{\mu\nu}\theta^b[T_a, T_b] \\
F'^a_{\mu\nu}T_a &= F^a_{\mu\nu}T_a + gF^a_{\mu\nu}\theta^b f^{abc}T_c \\
F'^a_{\mu\nu} &= F^a_{\mu\nu} - g\theta^b f^{abc}F^c_{\mu\nu} && (3.63)
\end{aligned}
$$

Equation 3.63 shows a really important result for the curvature tensor $F_{\mu\nu}$. To make this result visible one should rewrite Equation 3.51 for a general field $\psi$ in an arbitrary representation using all the indices.

$$(\psi')_i = (\psi)_i + ig\theta^a (T_a)_{ij} (\psi)_j \tag{3.64}$$

This equation can coincide with the transformation rule for the curvature appearing in Equation 3.63 if the generators are given by $(T_a)_{ij} = -i(f^a)_{bc}$ or in the usual notation $T_a = -if^{abc}$, which is exactly the definition of the adjoint representation presented in Chapter 2.2.4.

In conclusion, the really important result from the transformation rule of the curvature is that its rule of transformation was instructed by the transformation of the connection $A_\mu$ and it naturally became clear that the curvature behaves like a field that transforms (and therefore represented) in the adjoint representation of the group. Therefore, whenever an operator (as the covariant derivative) has to act on the curvature, the adjoint representation in implied.

### 3.2.3 Lagrangian and Equations Of Motion

The next expected step to properly formalize Yang-Mills theory is to build the action functional, from which the equations of motion can be derived. The action is the spacetime integral of the Lagrangian density, as described in section 3.1.

The action should be Lorentz invariant as well as gauge invariant. The Lorentz invariant product is the analogously to the Abelian theory $F_{\mu\nu}F^{\mu\nu}$, where in the present case the field strength tensor takes values on the Lie Algebra of the group. Under gauge transformation the Lorentz invariant product transforms homogeneously, as one can see from the following equation.

$$F'_{\mu\nu}F'^{\mu\nu} = UF_{\mu\nu}U^{-1}UF^{\mu\nu}U^{-1} = UF_{\mu\nu}F^{\mu\nu}U^{-1}$$

Of course, homogeneous transformations are a good result, but as mentioned the action should be a gauge invariant scalar quantity. The only possible way to make this product invariant is to use its trace $\mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\}$. There is another more elegant form of this trace, in terms of the components of the tensor $F_{\mu\nu}$.

$$\mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\} = \mathrm{Tr}\Big\{F^a_{\mu\nu}T^a F^{b\,\mu\nu}T^b\Big\} = F^a_{\mu\nu}F^{b\,\mu\nu}\,\mathrm{Tr}\Big\{T^a T^b\Big\}$$

In theoretical physics the generators are normalized as:

$$\mathrm{Tr}\Big\{T^a T^b\Big\} = \frac{1}{2}\delta^{ab} \tag{3.65}$$

Meaning that the invariant trace can be used as the Lagrangian density of the theory in the following form:

$$\mathcal{L} = \mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\} = \frac{1}{2}F^a_{\mu\nu}F^{a\,\mu\nu} \tag{3.66}$$

The action is the spacetime integral of this product and taking the variation of this action equal to zero the Equations Of Motion should arise naturally.

$$\begin{aligned}
\delta\int d^4x\mathcal{L} &= \int d^4x\delta\mathcal{L} \\
&= \int d^4x\delta(\frac{1}{2}F^a_{\mu\nu}F^{a\,\mu\nu}) \\
&= \frac{1}{2}\int d^4x 2F^a_{\mu\nu}\delta F^{a\,\mu\nu} \\
&= \int d^4x F^a_{\mu\nu}\delta(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu) \\
&= \int d^4x F^a_{\mu\nu}[\partial_\mu\delta A^a_\nu - \partial_\nu\delta A^a_\mu + gf^{abc}(A^b_\mu(\delta A^c_\nu) + (\delta A^b_\mu)A^c_\nu)] \\
&= \int d^4x[F^a_{\mu\nu}\partial_\mu\delta A^a_\nu - F^a_{\mu\nu}\partial_\nu\delta A^a_\mu + g(F^a_{\mu\nu}f^{abc}A^b_\mu(\delta A^c_\nu) + F^a_{\mu\nu}f^{abc}(\delta A^b_\mu)A^c_\nu)]
\end{aligned}$$

The first two terms can be integrated my parts. The index b, appearing bellow, stands for boundary. The antisymmetric property $F^a_{\mu\nu} = -F^a_{\nu\mu}$ is also used.

$$\begin{aligned}
\int d^4x F^a_{\mu\nu}\partial_\mu\delta A^a_\nu &= F^a_{\mu\nu}\delta A^a_\nu\Big|_{\mathrm{b}} - \int d^4x(\partial_\mu F^a_{\mu\nu})\delta A^a_\nu \\
\int d^4x F^a_{\mu\nu}\partial_\nu\delta A^a_\mu &= F^a_{\mu\nu}\delta A^a_\mu\Big|_{\mathrm{b}} - \int d^4x(\partial_\nu F^a_{\mu\nu})\delta A^a_\mu \\
&= F^a_{\nu\mu}\delta A^a_\nu\Big|_{\mathrm{b}} - \int d^4x(\partial_\mu F^a_{\nu\mu})\delta A^a_\nu \\
&= -F^a_{\mu\nu}\delta A^a_\nu\Big|_{\mathrm{b}} + \int d^4x(\partial_\mu F^a_{\mu\nu})\delta A^a_\nu \\
\int d^4x F^a_{\mu\nu}\partial_\mu\delta A^a_\nu &= -\int d^4x F^a_{\mu\nu}\partial_\nu\delta A^a_\mu
\end{aligned}$$

Changing the indices in the second part of the remaining term one has:

$$\begin{aligned}
F^a_{\mu\nu}f^{abc}A^b_\mu(\delta A^c_\nu) + F^a_{\nu\mu}f^{abc}(\delta A^b_\nu)A^c_\mu &= F^a_{\mu\nu}f^{abc}A^b_\mu(\delta A^c_\nu) - F^a_{\mu\nu}f^{abc}(\delta A^b_\nu)A^c_\mu \\
&= F^a_{\mu\nu}f^{abc}A^b_\mu(\delta A^c_\nu) - F^a_{\mu\nu}f^{acb}(\delta A^c_\nu)A^b_\mu \\
&= F^a_{\mu\nu}f^{abc}A^b_\mu(\delta A^c_\nu) + F^a_{\mu\nu}f^{abc}(\delta A^c_\nu)A^b_\mu \\
&= 2F^a_{\mu\nu}f^{abc}A^b_\mu(\delta A^c_\nu) \\
&= 2F^c_{\mu\nu}f^{cba}A^b_\mu(\delta A^a_\nu) \\
&= -2F^c_{\mu\nu}f^{abc}A^b_\mu(\delta A^a_\nu)
\end{aligned}$$

These changes make it easier to calculate the variation of the action.

$$\delta S = 0$$
$$2\ F^a_{\mu\nu}\delta A^a_\nu\Big|_{\mathrm{b}} - 2\int d^4x(\partial_\mu F^a_{\mu\nu} + gF^c_{\mu\nu}f^{abc}A^b_\mu)\delta A^a_\nu = 0$$

The boundary terms cancel out, because a common convention is that fields tend to go to zero in spatial and time infinity. For the last equation to be true independently of the chosen connection, the equation of motion are:

$$\partial_\mu F^a_{\mu\nu} + gF^c_{\mu\nu}f^{abc}A^b_\mu = 0 \tag{3.67}$$

One can rewrite these equations without the components. The first step is to multiply by a generator of the Lie algebra $T^a$ from the left.

$$\partial_\mu F^a_{\mu\nu}T^a + gF^c_{\mu\nu}f^{abc}T^aA^b_\mu = 0$$

The term $f^{abc}T^a$ after some index manipulation and using the antisymmetric property of the structure constants $f^{abc}$ can be rewritten as:

$$f^{abc}T^a = f^{cba}T^c = f^{acb}T^c = -f^{abc}T^c = -i[T^a, T^b]$$

The equations of motion are:

$$\begin{aligned}
\partial_\mu F^a_{\mu\nu}T^a + gF^c_{\mu\nu}f^{abc}T^aA^b_\mu &= 0\\
\partial_\mu F^a_{\mu\nu}T^a + gF^a_{\mu\nu}f^{cba}T^cA^b_\mu &= 0\\
\partial_\mu F^a_{\mu\nu}T^a + igF^a_{\mu\nu}[T^a, T^b]A^b_\mu &= 0\\
\partial_\mu F^a_{\mu\nu}T^a + ig[F^a_{\mu\nu}T^a, A^b_\mu T^b] &= 0\\
\partial_\mu F_{\mu\nu} - ig[A_\mu, F_{\mu\nu}] &= 0 \qquad (3.68)\\
\nabla_\mu F_{\mu\nu} &= 0 \qquad (3.69)
\end{aligned}$$

Note that Equation 3.68 is exactly how the covariant should act on the curvature $F_{\mu\nu}$, because as explained in the previous section the curvature transforms in the adjoint representation, therefore the gauge field $A_\mu$ should be in the adjoint representation as well. This means that the term $A_\mu F^{\mu\nu}$ that appears in covariant differentiation of $F_{\mu\nu}$ implies the map $ad_A[F] = [A_\mu, F^{\mu\nu}]$.

The equations of motion for the Yang-Mills connection field $A_\mu$ are given compactly in Equation 3.69. As a result the Lagrangian density used to define these equations is correct and usually it is written with an extra factor of $-1/4$, in order to absorb the extra factors appearing when varying the action. The final lagrangian density is:

$$\mathcal{L}_{YM} = -\frac{1}{2}\,\mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\} = -\frac{1}{4}F^{a}_{\mu\nu}F^{a\,\mu\nu} \tag{3.70}$$

Finally, it should be mentioned that these equations are homogeneous, because in the Lagrangian density there is no matter fields (fermions). The purpose of this thesis is to study pure Yang-Mills action, the only remark stated here is that if one needs to add matter to this action the ordinary derivative has to be replaced with the gauge covariant derivative in the Dirac's Lagrangian density.

### 3.2.4 Wilson Lines

Another way of introducing the covariant derivative is through the notion of parallel transfer. The field $\psi$ acquires a different phase when moving through the spacetime manifold, making it impossible to compare concretely two different values of the field. For example, think of the difference $\psi(y) - \psi(x)$. After the gauge transformation this difference will be $U(y)\psi(y) - U(x)\psi(x)$, which has a dependence on both phases on points $x, y \in \mathbb{R}^{1,3}$.

What is needed is some kind of field that will make possible to transfer the field $\psi(x)$ to $\psi(y)$ and then the comparison will be valid. More precisely, the above difference should transform covariantly under the gauge transformation $U$. In other words the new field should make the difference transform as the field $\psi(x)$ This condition is:

$$\psi(y) - \psi(x) \to U(x)[W(x,y)\psi(y) - \psi(x)] \tag{3.71}$$

This can be achieved by demanding that the new field $W(x,y)$, follow the transformation rule:

$$W(x,y) \to U(x)W(x,y)U^{-1}(y) \tag{3.72}$$

The field $W(x,y)$ is actually the parallel transporter of the theory, and can be used to define the covariant derivative, because the derivative itself is a difference between two infinitesimally neighboring points.

$$\nabla_\mu\psi(x) = \lim_{dx^\mu \to 0} \frac{W(x, x+dx)\psi(x+dx) - \psi(x)}{dx^\mu} \tag{3.73}$$

Equation 3.72 meas that the covariant derivative transforms as $\nabla_\mu\psi(x) \to U(x)\nabla_\mu\psi(x)$, which was the basic requirement of the previous section. The field $W(x,y)$ depends on two points and it acts on the field $\psi$ and produces another new field of the same nature with $\psi$. This is true because this new field $W(x,y)\psi(y)$ is then subtracted from $\psi(x)$. Therefore, it acts on the field the same way as the group element $U(y)$, meaning that $W(x,y)$, takes values on the Lie Group.

A very important property is that the parallel transporter should not have any effect when there is no parallel transfer.

$$W(x,x) = \mathbb{1} \tag{3.74}$$

This equation may seem quite simple, but the true underlying meaning is that the neutral element of the group $\mathbb{1} \in G$ is a value of $W(x,y)$. As a result the values of the field $W(x,y)$ form a new group, which is a subgroup of $G$, according to Definition 2.2.2. An expansion around the identity element can be performed. In the following expansion the field $A_\mu$ is just an arbitrary field.

$$W(x, x+dx) = \mathbb{1} - igA_\mu dx^\mu + \mathcal{O}(dx^2) \tag{3.75}$$

As mentioned, the field $W(x,y)$ takes values on the Lie Group. In order for this to happen in Equation 3.75 the field $A_\mu$ must take values on the Lie Algebra, in order for this Taylor expansion around the identity element to map to the Lie Group. The real number $g$ has been extracted from the above product, in order to reproduce the same formulas for the covariant derivative as in Section 3.2.

Substituting the Taylor expansion of Equation 3.75 in the limit 3.73 and derive Equation 3.52 introduced in Section 3.2.

$$\begin{aligned}
\nabla_\mu \psi(x) &= \lim_{dx^\mu \to 0} \frac{(\mathbb{1} - igA_\mu dx^\mu)\psi(x+dx) - \psi(x)}{dx^\mu} \\
&= \lim_{dx^\mu \to 0} \frac{\psi(x+dx) - \psi(x)}{dx^\mu} - igA_\mu \lim_{dx^\mu \to 0}(\psi(x+dx)) \\
\nabla_\mu \psi(x) &= (\partial_\mu - igA_\mu)\psi(x)
\end{aligned}$$

The last equation is the same as that of the previous Section, concluding that the field $A_\mu$ is the gauge connection. Having defined the covariant derivative, a field $\psi$ can be parallel transferred from a point $x \in \mathbb{R}^{1,3}$ to a point $y \in \mathbb{R}^{1,3}$ along a curve $\gamma : \lambda \in \mathbb{R} \mapsto x^\mu(\lambda) \in \mathbb{R}^{1,3}$. The equation of this parallel transfer is analogous to the directional derivative of a scalar function.

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \nabla_\mu \psi(x) = 0 \tag{3.76}$$

The solution of the above equation can be decomposed on a product of the following form.

$$\psi(x) = W(x,y)\psi(y) \tag{3.77}$$

The meaning of this decomposition is that having defined the field at a initial point $y$, one is looking for a way to transfer this field along any point $x$ on the curve $\gamma$, such that Equation 3.76. The field $W(x,y)$ is the propagator and of course it depends on the form of the chosen curve $\gamma$.

Substituting Equation 3.77 into the differential equation 3.76 the result is:

$$\begin{aligned}
\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\psi(x) &= 0 \\
\frac{\mathrm{d}\psi(x)}{\mathrm{d}\lambda} - ig\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}A_\mu\psi(x) &= 0 \\
\frac{\mathrm{d}W(x,y)\psi(y)}{\mathrm{d}\lambda} - ig\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}A_\mu W(x,y)\psi(y) &= 0 \\
\frac{\mathrm{d}W(x,y)}{\mathrm{d}\lambda}\psi(y) - ig\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}A_\mu W(x,y)\psi(y) &= 0 \\
\frac{\mathrm{d}W(x,y)}{\mathrm{d}\lambda} - ig\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}A_\mu W(x,y) &= 0
\end{aligned} \tag{3.78}$$

The last equation is really important, because it shows that the $W(x,y)$, which in this context is named the parallel propagator satisfies the same differential equation with the field $\psi(x)$.

The solution of this equation is known and the program to solve it consists of two main steps. The first step to solve this differential equation is to integrate both sides in terms of the parameter $\lambda$. The integral of the right-hand side would still contain the unknown propagator $W(x,y)$. The next step, is to plug the resulting integral in $W(x,y)$ of the right-hand side. This will lead to an infinite series of integrals, which after manipulation they can be written in a closed form using the path-ordering operator $\mathcal{P}$. This procedure is actually the same with the one usually performed to obtain the Dyson' s formula in Quantum Field Theories.

$$W(x,y) = \mathcal{P}\left\{-\exp\left[ig\int_x^y A_\mu dz^\mu\right]\right\} \tag{3.79}$$

This field can be found in the bibliography with a lot of different names. The most famous ones are parallel transporting field and Wilson line. Its role is to parallel transfer any field $\psi(y)$ it acts from the point $y$ to the point $x$, making the difference $W(x,y)\psi(y) - \psi(x)$ to transform covariantly under the gauge transformation.

Because it acts on the fields in the same manner as the gauge transformation itself, the field must take group values, meaning that it can also be expanded on the Lie Algebra of the group, because as seen from Equation 3.74 the neutral element of the group is mapped to with the Wilson Line.

Last but not least, a really important property that is actually the beginning of Lattice Gauge theory is the parallel transport of a field to itself through a closed path. For such a loop, Equation 3.79 gives the following contour integral, which can be expressed in terms of the field strength tensor using the generalized Stokes' Theorem, which is no trivial result. .

$$W_{\text{loop}} = \mathcal{P}\left\{-\exp\left[ig\oint A_\mu dz^\mu\right]\right\} \tag{3.80}$$

In the next Chapter, the connection between the tensor $F_{\mu\nu}$ and the Wilson loop will be derived using lattice gauge theory. For now, the important remark is that the Wilson loop

transforms homogeneously, because the starting point of the loop has no relevant effect on its transformation.

$$W'_{\text{loop}} = UW_{\text{loop}}U^{-1} \tag{3.81}$$

Finally, taking the trace of the previous equation [7] and using its cyclic property, one can see that the trace of the Wilson loop is a gauge invariant quantity, therefore it can be used to construct the action of the theory.

$$\text{Tr}\Big\{W'_{\text{loop}}\Big\} = \text{Tr}\Big\{UW_{\text{loop}}U^{-1}\Big\} = \text{Tr}\Big\{U^{-1}UW_{\text{loop}}\Big\} = \text{Tr}\{W_{\text{loop}}\} \tag{3.82}$$

## 3.3 Quark Confinement

At the time being, non-Abelian gauge theories are believed to be behind the gauge interaction of the strong nuclear force between quarks and antiquarks. In the Standard Model the gauge group that describes these strong interactions is $SU(3)$. Quarks and antiquarks are realised as constituents of hadrons and mesons. They are massive electrically charged fermions that also carry the color charge. This new charge takes three different values, often referred to as red, green and blue, and, of course, for every value, the anticharge exists as well (antire, antigreen, antiblue). Every quark can be understood as the source of a chromomagnetic field, which, as the electromagnetic field, can be visualize as a set of field lines. Chromoelectric field lines emerge from quarks and end up in antiquarks.

There is commonly acceptable experimental evidence that quarks are the contituents of hadrons and mesons, with the first consisting of three quarks or antiquarks and the second consinsting of just a quark-antiquark pair. The experimental evidence does not consist of direct observations of free quarks in nature. The quarks are held together by strong nuclear forces inside the mesons and hadrons and are theoretically are understood only as bound states of such particles. The name of the phenomenon commonly accepted as the cause is quark confinement and it is believed that it is based on the non-Abelian nature of gauge theories. Although, still today there is no theoretical proof of confinement, making it an open problem for theoretical physics.

Moreover, a really important property of quarks in large energy scales, meaning in short distances is that they behave as free particles with no interaction in-between. Mathematically speaking the coupling constant, which encodes how strong is the interaction of quarks by the intermediate vector gluon field $A_\mu$ goes to zero in short distances, meaning when quarks are bound inside hadrons or mesons. This phenomenon is proved analytically and is known as asymptotical freedom [14, 21].

A consequence of confinement is that static quark-antiquark pairs are bound together by a string of constant chromoelectric flux, formed by their chromoelectric field lines, as visualized

[7] Remember that the Wilson line integrals take values on the Lie Group. As such, the trace value have a meaning on the chosen representation.

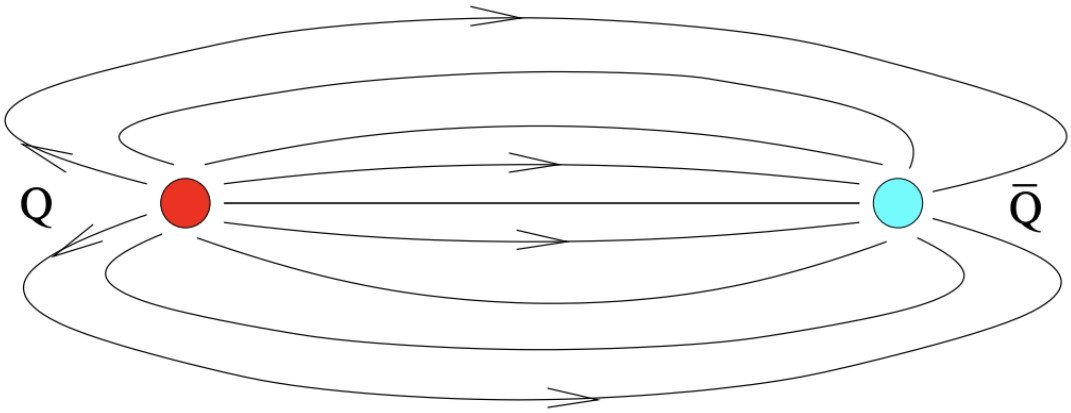



**Figure 3.1:** A quark-antiquark pair with the corresponding chromoelectric field lines, which concentrate in a string/tube of fixed chromoelectric flux [8].

in Figure 3.1. The energy of this string grows linearly with the separation of the pair. After a certain point it is cheaper in energy for a quark-antiquark pair to be created from the vacuum, than to extend more the string. This picture can be seen clearly by caclulating the quark-antiquark potential. This potential turns out to be linear in the separation $r$ of the quarks [27, 22, 7]:

$$V(r) = Kr \tag{3.83}$$

The coefficient $K$ is known as the string tension and in terms of dimensions it has to be $[K] = [\Lambda_{QCD}]^2$. This potential, as every potential in a quantum mechanical theory, can be interpreted as the ground state energy.

As mentioned this potential is expected only in non-Abelian theories, while in Abelian theories (as QED), the potential must be the Coulomb' s law. Therefore, a quantity that measures the existence of this linear potential is needed. Using path integral arguments for Abelian and non-Abelian theories it can be proven that an appropriate measure of the confinement effect is the expectation value of a Wilson loop. For non-confining theories this expectation value should obey a perimeter law, while for theories, where confinement exists, the expectation value should obey what is known as an area law.

More specifically in a theory, where charges confine, the Wilson loop will follow the area law presented bellow. $A$ is the area enclosed by the loop and $K$ is the string tension [27].

$$\langle 0|W_{\text{loop}}|0\rangle \sim e^{-KA} \tag{3.84}$$

For a theory containing unconfined charges the expectation value of the Wilson loop depends only on the perimeter of the loop. The constant $C$ is parameter depending on the loop shape, the details of regularization used to cutoff the divergences of the path integral and the coupling constant $g$ [27].

$$\langle 0|W_{\text{loop}}|0\rangle \sim e^{-CP} \tag{3.85}$$

The above results were calculated for pure Abelian and non-Abelian theories. In theories containing matter, where the effect of charge screening is there, the Wilson loop depends on the area as well as the perimeter with a weighting towards one, depending on the size of the loop.

# Chapter 4

# Pure Yang-Mills Theory on the Lattice

## 4.1 Lattice Formulation of Pure Yang-Mills Theory

The purpose of this section is to formulate gauge theories on a discrete spacetime lattice. The term discrete spacetime is used to describe the approximation of the Minkowski spacetime manifold $\mathbb{R}^{1,3}$ with discrete points. This can be achieved by defining $n$ sites in each one of the four dimensions, meaning $n^4$ sites in total, and demanding periodic boundary conditions on the last site in each dimension. A periodic 2-dimensional lattice can be seen in Figure 4.1.

The idea is that by assuming spacing $a$ between each site the continuum can be derived by taking the limit $a \to 0^+$. So, according to renormalization (see Section 4.6), every quantity $H$ defined on the lattice should have an appropriate dependency on the spacing $a$, such that when taking the limit $a \to 0^+$ $H$ has to approach its real physical value.

The basic quantity in theoretical physics and mathematics is the functional that gives the equations of motion, or in simpler terms the action $S$. For the lattice, the action was first pioneered by Wilson and the idea is that when defining a field on the $n^4$ sites the Wilson line

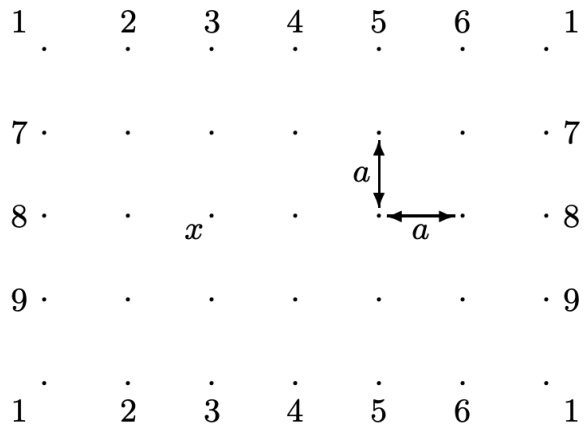



**Figure 4.1:** A $6 \times 3$ (2-dimensional) lattice with periodic boundary conditions.

is the appropriate field to use in order to compare these values gauge invariantly.

According to Equation 3.72 of the previous Chapter, the Wilson line must transform as $U(x)W(x,y)U^{-1}(y)$. With the discretization imposed the group elements $U$ can only be defined on the sites of the lattice and the most elementary Wilson line of this model can be thought as field that "lives" on the link between two neighboring sites. Therefore, on the lattice the field transforms as:

$$W_\mu(\nu) \to U(n)W_\mu(\nu)U^{-1}(n+\mu) \tag{4.1}$$

where $n$ denotes a lattice site and $n+\mu$ the next lattice site in the $\mu$ direction. A more formal notation could be $W(n, n+\mu)$ instead of $W_\mu(\nu)$, but it is not a common notation, so it will not be used.

The field $W_\mu(\nu)$ can be thought of as taking the value of a field at a lattice site $n$ and transferring this value on the site $n+\mu$. So, to compare the values of a field between two points on the lattice, the only requirement is a discrete path connecting neighboring sites and the field $W$ on this path that is actually a product of elementary fields $W_\mu(\nu)$ between all the sites of the lattice.

As stated in the previous chapter the Wilson line takes values on the group manifold. As a result the inverse field $W_\mu^{-1}(\nu)$ exists on every link and intuitively what it does on the lattice is to connect field values in the opposite way, meaning from a site $n$ to its neighbor $n-\mu$ in the $\mu$ direction or in the notation used so far $W_\mu^{-1}(n) = W_{-\mu}(n)$

To construct the action the trace of a Wilson loop must be used, because it is a gauge invariant quantity. The simpler loop on the lattice consists of two neighboring sites $n, n+\mu$. The loop on this link would be:

$$W_{\text{loop}} = W_\mu(\nu)W_\mu^{-1}(\nu) = \mathbb{1}$$

The simplest path gives the identity element of the group, so it there is no information that can be extracted from this path. The next simplest loop is a square lying on a two dimensional plane. This loop form an elementary square on the lattice and in Lattice QCD is referred to as "plaquette". The Wilson loop on an arbitrary plaquette laying on the $\mu\nu$ plane connects four lattice sites as shown in Figure 4.2:

$$W_\square = W_\mu(x)W_\nu(x+\mu)W_\mu^{-1}(x+\nu)W_\nu^{-1}(x)$$

According to Equation 3.79 of the previous Chapter the parallel transporter $W$ can be expressed in terms of the gauge field $A_\mu$ as path order exponential of a contour integral. Considering a link on the lattice the aforementioned exponential takes the following form:

$$W_\mu(x) = \mathcal{P}\exp\Big\{-ig\int_C A_\mu d^\mu z\Big\} = e^{-igaA_\mu(x)}$$

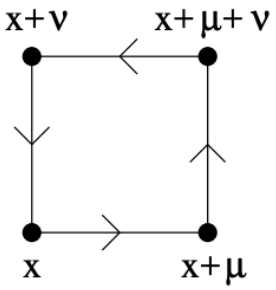



**Figure 4.2:** Elementary plaquette lying on the $\mu\nu$ plane on the lattice.

The essence of the lattice is that lattice spacing $a$ is a small parameter, meaning that the gauge field $A_\mu$ can be expanded around this parameter as:

$$A_\mu(x+\nu) = A_\mu + a\partial_\nu A_\mu + \mathcal{O}\Big(a^2\Big) \tag{4.2}$$

$$A_\nu(x+\mu) = A_\nu + a\partial_\mu A_\nu + \mathcal{O}\Big(a^2\Big) \tag{4.3}$$

The product presented above can be written as:

$$W_\square = e^{-igaA_\mu(x)} e^{-iga(A_\nu(x)+a\partial_\mu A_\nu(x))} e^{iga(A_\mu(x)+a\partial_\nu A_\mu(x))} e^{igaA_\nu(x)}$$

Using BCH formula the above equation takes the following form:

$$W_\square(x) = e^{-iga\left(A_\mu(x)+A_\nu(x)+a\partial_\mu A_\nu(x)-\frac{iga}{2}[A_\mu(x),A_\nu(x)]\right)+\mathcal{O}\left(a^3\right)} e^{iga\left(A_\mu(x)+A_\nu(x)+a\partial_\nu A_\mu(x)+\frac{iga}{2}[A_\mu(x),A_\nu(x)]\right)+\mathcal{O}\left(a^3\right)}$$

Finally, using one more time the BCH formula the product can be written as:

$$W_\square = e^{-iga^2 F_{\mu\nu}(x)+\mathcal{O}\left(a^4\right)} \tag{4.4}$$

Therefore, a Wilson loop around a plaquette is a sort of measure for the curvature of the theory, which is intuitively expected, because on the lattice the elementary squares are used to measure the curl and the curvature is defined as a generalised curl. Moreover, the curvature is Lie algebra valued, which makes possible for $W_\square$ to be expanded around the identity element.

$$W_\square(x) = \mathbb{1} - iga^2 F_{\mu\nu}(x) - \frac{g^2a^4}{2} F_{\mu\nu}F_{\mu\nu} + \mathcal{O}\Big(a^6\Big)$$

The gauge invariant quantity is the trace of the above expression.

$$\mathrm{Tr}\{W_{\square}(x)\} = \mathrm{Tr}\{\mathbb{1}\} - iga^2\,\mathrm{Tr}\{F_{\mu\nu}(x)\} - \frac{g^2a^4}{2}\,\mathrm{Tr}\{F_{\mu\nu}(x)F_{\mu\nu}(x)\} \tag{4.5}$$

The trace of the identity matrix depends, of course, on the representation chosen for the gauge group $G$. In the Standard Model, the symmetry used for QCD is $SU(3)$. In the present thesis, the general group $SU(N)$, will be used. For $SU(N)$ the trace of the identity matrix is the dimension $N$ of the group matrices. Moreover, the curvature $F_{\mu\nu}$ can be expanded as $F^a_{\mu\nu}T^a$, but the generators of $SU(N)$ are Hermitian, therefore traceless matrices.

Taking into consideration the above remarks the Wilson loop on an elementary plaquette [1] becomes proportional to the trace $\mathrm{Tr}\{F_{\mu\nu}F_{\mu\nu}\}$, which is exactly the term that appeared in the action of the Yang-Mills theory.

$$W_{\mathrm{loop}}(x) = N - \frac{g^2a^4}{2}\,\mathrm{Tr}\{F_{\mu\nu}(x)F_{\mu\nu}(x)\} \tag{4.6}$$

After some rescaling the action can be written as in Equation 4.7. The rescaling involves taking $a = 1$, which is equivalent to say, that on the lattice the dimenions are measured in lattice units, where the fundamental dimension is $[a] = [L]$.

$$S = \beta\sum_{\square}\left(1 - \frac{1}{N}\,\mathrm{Re}\{\mathrm{Tr}\{W_{\square}\}\}\right) = \beta\sum_{\square}\left(1 - \frac{1}{N}\,\mathrm{Tr}\left\{W_{\square} + W_{\square}^{-1}\right\}\right) \tag{4.7}$$

where the constant $\beta$ contains the number $N$, which is useful when calculating and comparing differences between different gauge groups of the Yang-Mills theory. Of course, the action can be formulated for any representation of the gauge group, but in theories without fermions as considered in the present thesis, the fundamental representation is an easy choice.

$$\beta = \frac{2N}{g^2} \tag{4.8}$$

Now, with the help of the path integral formulation the partition function can be calculated. The formulation used in the context of this thesis is different with the one introduced by Feynman [11], only by the use of the imaginary time $\tau = -it$. With this "rotation" from Minkowski to Euclidean space the action transforms as $S_E \to iS$, where the index $E$ stands for Euclidean, and for notation simplicity it will not be used from now on. One can question whether the use of Euclidean spacetime over Minkowski can cause a problem with the predictions of such a theory. The naive answer is that on the lattice the system can be studied in the confined region in low velocities, where Minkowski and Euclidean metric give the same results.

$$Z = \int \mathcal{D}W e^{-S[W]}$$

[1] That is what the value $x$ denotes in the above calculation according to Figure 4.2

The measure of the path integral is in terms of the parallel transporters $W$, between neighboring sites of the lattice, because that is the dependence of the action defined in Equation 4.7. Meaning that this integral is $d$-dimensional, where $d$ is the number of links. These parallel transporters are elements of the Lie Group $G$, which means that some extra attention must be given to the integration measure $\mathcal{D}W$, because the partition function must be gauge invariant, resulting in the obligatory use of an invariant measure as well. This will be discussed in the next section.

Last but not least, the expectation value of an arbitrary operator $H = H(W)$ can be calculated as usual from the path integral. The important fact here is that the operator has to be described as a gauge invariant function of $W$.

$$\langle H \rangle = \frac{\int \mathcal{D}W H(W) e^{-S[W]}}{\int \mathcal{D}W e^{-S[W]}} \tag{4.9}$$

## 4.2 Dimensional Analysis

Throughout the present thesis natural units ($\hbar = c = 1$) were implied. It is important though to know what implications this use has in the physical dimensions of the quantities involved in the field theory and moreover what happens with the dimensions of these quantities on the lattice, when the lattice spacing $a$ does not appear explicitly in the action.

Starting from Einstein' s famous formula $E = mc^2$ and setting the speed of light equal to 1, this formula reduces to $E = m$. Therefore, mass dimensions also describe energy $[E] = [M]$. The next step is to use Planck' s law $E = \hbar f$. Setting $\hbar = 1$, energy dimension become also equal to the inverse time $[E] = [T]^{-1}$. Obviously, then, $[T] = [M]^{-1}$. Considering, the relation between a wavelength and its frequency $\lambda = cf^{-1}$, then the length dimension in natural units is $[L] = [T] = [M]^{-1}$. As a result, every physical quantity has mass or equivalently length dimension in an appropriate power, when the natural units system is in use.

As one can see from Equation 4.9 the action appears in a exponent, which means that it must be a dimensionless quantity. The action is given as an integral of the Lagrangian density.

$$\begin{aligned} [d^4x] &= [L]^4 = [M]^{-4} \\ [\mathcal{L}] &= [L]^{-4} = [M]^4 \end{aligned}$$

The Lagrangian density of a pure Yang-Mills theory is:

$$\mathcal{L} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} = -\frac{1}{4} (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^a_\mu A^a_\nu)^2$$

A dimensional analysis on the first term gives:

$$
\begin{aligned}
[\partial_\mu]^2[A^a_\nu]^2 &= \Big([M]^4\Big)^2 \\
\left(\frac{1}{[L]}\right)^2 [A]^2 &= M]^4 \\
[M]^2[A]^2 &= [M]^4 \\
[A] &= [M]
\end{aligned}
$$

Finally, the interaction term gives the dimension of the coupling constant $[g]$.

$$
\begin{aligned}
\Big([g][A]^2\Big)^2 &= [M]^4 \\
[g]^2[A]^4 &= [M]^4 \\
[g]^2[M]^4 &= [M]^4 \\
[g] &= 1
\end{aligned}
$$

This is a very important result for Yang-Mills theory. The only external parameter of the theory is the coupling constant $g$, which is dimensionless. In terms of renormalization this means that the theory is exactly renormalizable.

On the lattice the position vector $x^\mu$ is discrete. It can only take values on the lattice sites, so it can be written as $x^\mu = \hat{n}a$, where a is the lattice spacing and $\hat{n}$ is a 4-tuple of integers $\hat{n} = (n_1, n_2, n_3, n_4)$, that encodes how many lattice spacings in each of the four dimensions a particle must hop to arrive at a lattice site. Therefore, the physical dimensions are hidden on the lattice spacing, which is $[a] = [L] = [M]^{-1}$.

As one can see form Equation 4.7 the lattice spacing does not appear explicitly in the action. The action is calculated only by summing the values of elementary plaquettes $W_\square$, which are group elements, therefore dimensionless quantities. When measuring observables $H$ on the lattice, they appear as dimensionless quantities. As a result, to extract a physical value for these quantities their physical dimensionality must be fixed as:

$$H = \hat{H}a^{-d} \tag{4.10}$$

where $d$ is the exponent of the mass dimension (in natural units) of the observable and $\hat{H}$ is the dimensionless quantity measured from the lattice.

## 4.3 Group Integration

According to the transformation rule for the parallel transporter $W$, presented in Equation 3.72, in order to define a gauge invariant path integral, the integration measure $\mathcal{D}W$ must invariant under the left and right action of the group elements $U \in G$.

The purpose of this section is to define an integration measure over the group manifold that is invariant under gauge transformations. Only compact Lie Groups will be considered. As an ordinary integral, this should also be bilinear and positive definite. Rigorously, consider two functions $f, h : G \to G$ and two complex numbers $a, b \in \mathbb{C}$. The basic properties of the integral over the Lie Group $G$ are:

$$\int dg\,(af(g) + bh(g))) = a\int dgf(g) + b\int dgh(g), \quad \forall g \in G$$
$$\int dgf(g) > 0, \text{ if } \quad f(g) > 0 \ \ \forall g \in G$$

An additional property corresponds to the shift invariance of an one dimensional real integral $\int dxf(x+a) = \int dxf(x), \ \ \forall a \in \mathbb{R}$. In the group this shift will be $g' \odot g$, where as usual, " $\odot$ " denotes the binary operation between the group elements. The group element $g'$ is just an arbitrary but fixed element of the group $G$ and, of course $g' \odot g \neq g \odot g'$ for a general group.

So, as an analogue to the shift of regular integrals one can take the invariant measure to be left or right invariant under the action of the arbitrary element $g'$. Here, left invariance will be imposed as a condition and but it can be proven that for compact groups the measure is also right-invariant.

$$\int dgf(g) = \int dgf(g' \odot g) \tag{4.11}$$

The integral over the group is actually a volume over the group manifold, which in our case is a compact group. This means that this integral can be finite. Therefore, the following condition for the normalization of the measure has a reasoning.

$$\int dg = 1 \tag{4.12}$$

The first task is to find an expression for the measure $dg$. The manifold of the Lie Group can be parametrized as surface of $\mathbb{R}^N$, where $N$ is the dimension of the group manifold[2]

$$G = \{g(a) | a \in I \subseteq \mathbb{R}^N\}$$

The volume $I$ in the above equation is the largest possible volume that can be chosen such that when the parameter $a$ runs once over the volume the group elements $g(a)$ are produced one time. This measure exists and is known as the Haar measure. The way to transform the integration from the group manifold to the set $I \subseteq \mathbb{R}^N$ is known from the Chapter 2.1.3.

$$\int dgf(g) = \int_I da^N \sqrt{\det\{M\}} f(g(a)) \tag{4.13}$$

where M denotes the metric on the group manifold given by the following formula for a general compact Lie Group.

[2]see Table 2.3 for the dimensions of common Lie Groups used in theoretical physics.

$$M_{ij} = \mathrm{Tr}\left\{ \frac{\partial g(a)}{\partial a^i} \frac{\partial g^{-1}(a)}{\partial a^j} \right\} \tag{4.14}$$

By differentiating over $a$ the relation $gg^{-1} = \mathbb{1}$ a different form of Equation 4.14 can be derived. Note that the identity element $\mathbb{1}$ can be obtained from the parametrization function $g(a)$ for a specific value of $a$.

$$\begin{aligned} g(a)g^{-1}(a) &= \mathbb{1} \\ \frac{\partial g(a)}{\partial a} g^{-1}(a) + g(a) \frac{\partial g^{-1}}{\partial a} &= 0 \\ \frac{\partial g^{-1}(a)}{\partial a} &= -g^{-1}(a) \frac{\partial g(a)}{\partial a} g^{-1}(a) \end{aligned}$$

Using the cyclic property of the trace operator the metric tensor $M_{ij}$ can be written as:

$$M_{ij} = \mathrm{Tr}\left\{ -g^{-1}(a) \frac{\partial g(a)}{\partial a^i} g^{-1}(a) \frac{\partial g(a)}{\partial a^j} \right\} \tag{4.15}$$

The existence of the Haar measure has a very important result on gauge theories. That is the integral over the group of every gauge dependent element will always be zero. Therefore, for an observable $H(g)$ all its gauge dependent parts vanish. An intuitive example is the integral of $f(g) = g$. By appropriate elements $U, U' \in G$ this function can transformed to $-f(g) = -g$. Then by left and right invariance the integral equals to zero.

$$\begin{aligned} \int_G dgg &= \int_G dgUgU' \\ \int_G dgg &= \int_G dg(-g) \\ \int_G dgg &= -\int_G dgg \\ \int_G dgg &= 0 \end{aligned}$$

The Haar measure has a really important consequence for gauge theories, which is known as Elitzur' s theorem. The theorem states that only gauge invariant functions can have a non-vanishing expectation value. This can be seen from Equation 4.9, where the path integral measure is actually a product of Haar measures on the group manifold $\mathcal{D}W = \prod dW$. More specifically, an observable acquires a non zero value only if every variable $W_{ij}$ is multiplied by its inverse $W_{ij}^{-1}$. So, in order to measure an observable on the lattice one needs to construct a gauge invariant operator (in the continuum) and then the exponential of the action will make this operator also gauge invariant on the lattice.

Last but not least, some important integrals, that will be used on the lattice are presented bellow. The notation $W_{ij}$ means that $W$ is a Wilson line (parallel transporter) from point $i$

to point $j$, or from site $i$ to site $j$, for a spacetime lattice. It is mentioned once more that $W$ takes values on the group.

$$\int_G dW W_{ij} = 0 \tag{4.16}$$

$$\int_G dW W_{ij} W^{-1}_{kl} = \frac{1}{N}\delta_{ik}\delta_{jl} \tag{4.17}$$

## 4.4 Order Parameters and Phase Transitions

Phase transitions are a quite known subject from statistical mechanics. A famous phase transition occurs in ferromagnetic materials. Considering the Ising model where spins are considered as a scalar quantity on the sites of a lattice and can only take a value from the group $Z_2 = \{+1, -1\}$. Spin-up is a spin that has the value $+1$, and correspondingly spin-down is the spin with the value $-1$. The basic postulate of the Ising model is that a particular spin can only interact with its nearest neighboring spins and no other.

Above a specific temperature, which is called critical temperature $T_c$ the spins are arranged randomly on the lattice, which results in $\langle\sigma\rangle = 0$. Bellow the critical temperature spins can be aligned in a certain direction causing $\langle\sigma\rangle = 1$ or $\langle\sigma\rangle = -1$. This is phenomenon is known as spontaneous symmetry breaking.

The importance of this preliminary from statistical mechanics is that $\langle\sigma\rangle$ can be used to determine whether a system is in the ferromagnetic state or not. That is the definition of an order parameter. An order parameter is a number that becomes trivial in one phase and remains always non-trivial for the other phase of the system.

Pursuing this analogy, one should look for different phases in a lattice gauge theory, where the interaction are not between spins, but between the link variables $W_\mu(n)$. The parameter $\beta$ plays the role of the Boltzmann factor. Therefore, when one talks about phase transitions in lattice gauge theories, what it means is that a certain number vanishes beyond (or bellow) a critical value $\beta_c$ and remains non-zero in the other region of $\beta_c$.

This behaviour was expected in a theory that pursues confinement. More specifically, in a confined phase the coupling constant $g$ is large, while in the unconfined phase of the theory $g$ vanishes asymptotically. In terms of $\beta$, which relates to $g$ via Equation 4.8, the confined phase appears in small $\beta$, while the asymptotically free/unconfined phase makes its appearance in large values of $\beta$.

Having established that a phase transition between confined and unconfined phases might appear in lattice gauge theories, an order parameter must be found, in order to distinguish these two phases. In analogy with the Ising model of ferromagnets, one can think that the expectation value $W_{ij}$ is an appropriate parameter. This expectation value is calculated analytically using the path integral formulation as follows:

$$\langle W_\mu(n)\rangle = \frac{1}{Z}\int \mathcal{D}W W_\mu(n) e^{-S[W]} = 0$$

where the measure of the integration is $\mathcal{D}W = \prod W_\mu(n)$. As explained in the previous section all integrals on a compact group manifold are zero, when integrating gauge non-invariant quantities as $W_\mu(n)$ (see Equation 4.16). As a result this expectation is always $\langle W_\mu(n) \rangle = 0$ and cannot serve as an order parameter.

The previous result is also really important because it encodes the fact that no gauge fixing is needed in lattice gauge theory. When the measure of the path integral is over the gauge fields, then integration is carried out through all the functions $A_\mu$, including those that do not respect gauge symmetry. That is why gauge fixing is needed. On the lattice, the integration measure is the Haar measure as stated in the previous section, that sends to zero all the non gauge invariant functions.

The simplest non-trivial gauge invariant operator that can be constructed from the links $W_{ij}$ is the plaquette $W_\square$ used to define the action, or more specifically the trace of this function $\mathrm{Tr}\{W_\square\}$. On the lattice a plaquette connecting the sites $ijkl$ transforms as follows:

$$\begin{aligned}
\mathrm{Tr}\{W'_\square\} &= \mathrm{Tr}\Big\{W'_{ij}W'_{jk}W'^{-1}_{lk}W'^{-1}_{il}\Big\} \\
&= \mathrm{Tr}\Big\{g_i W_{ij} g_j^{-1} g_j W_{jk} g_k^{-1} g_k W_{lk}^{-1} g_l^{-1} g_l W_{il}^{-1} g_i^{-1}\Big\} \\
&= \mathrm{Tr}\{W_\square\}
\end{aligned}$$

As a result, the first parameter that can be measured on the lattice, without giving a non-zero expectation value, is the average plaquette, which represents the internal energy of the system.

$$P = \left\langle \frac{1}{N} \mathrm{Tr}\{W_\square\} \right\rangle = \frac{\frac{1}{N}\int \mathcal{D}W \, \mathrm{Tr}\{W_\square\} e^{-S[W]}}{\int \mathcal{D}W e^{-S[W]}} \tag{4.18}$$

The average plaquette does not have the vanishing properties of magnetization, but as one can see from simulations it can exhibit singularities when plotted as a function of $\beta$, in the region where the phase transition is expected. In the case of the groups $SU(2)$ and $SU(3)$, Monte Carlo results are presented in Figures 4.3, 4.4. The average plaquette seems to change behavior when $\beta$ is in the interval $(2, 2.5)$ for $SU(2)$ and $(5, 6)$ for $SU(3)$, which correspond to $g \in (1, 2)$. This change of behavior is not singular, therefore no phase transition can be seen in non-Abelian $SU(2)$ and $SU(3)$ gauge theories. The system continuously pass from the unconfined phase to the confined one. The fitted curves appearing in these figures will be explained in the next section. Briefly, they correspond to the theoretically predicted curves expected in the weak and the strong coupling regimes.

The next order parameter that is usually measured on the lattice is the expectation value of the Wilson loop. As mentioned in Section 3.3 the Wilson loop is expected to follow an area law in the strong coupling regime and a perimeter law for weak couplings. It does not have the vanishing properties, but it does behave differently. The expectation value for a Wilson loop around a path $C$ is:

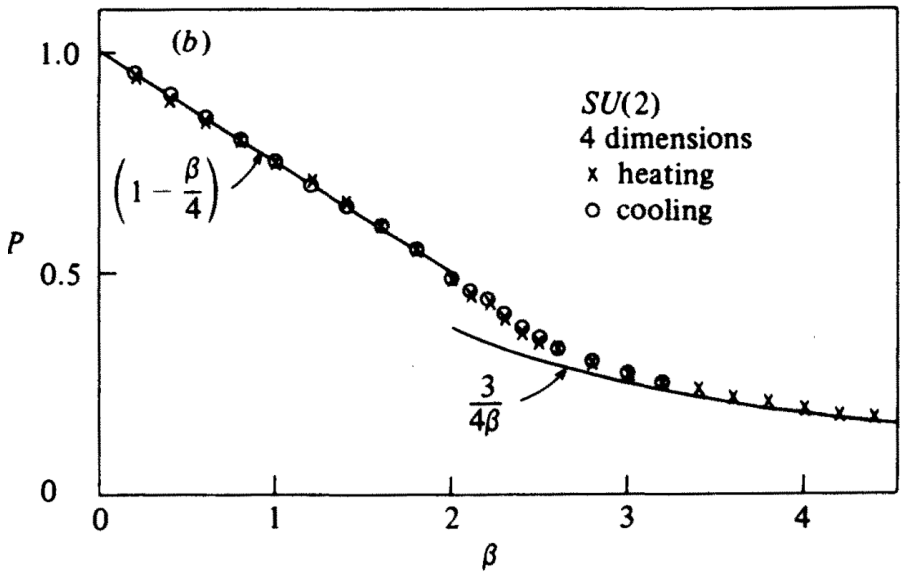



**Figure 4.3:** Monte Carlo results for the average plaquette $P$ as a function of the parameter $\beta$ for the gauge group $SU(2)$ on a $8^4$ lattice [5].

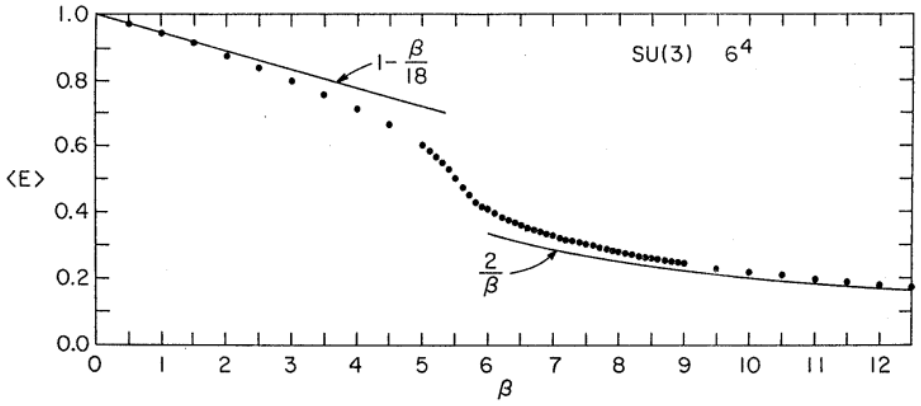



**Figure 4.4:** Monte Carlo results for the average plaquette $P$ as a function of the parameter $\beta$ for the gauge group $SU(3)$ on a $8^4$ lattice [2].

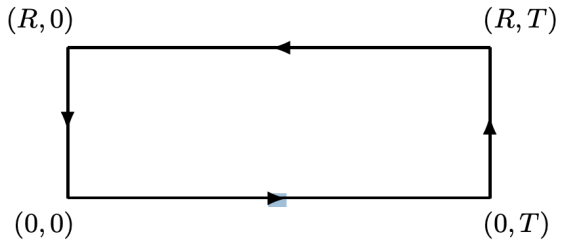



**Figure 4.5:** A timelike rectangular Wilson loop on the lattice.

$$\langle \mathrm{Tr}\{W_{\mathrm{loop}}\}\rangle = \frac{\int \mathcal{D}W \,\mathrm{Tr}\Big\{\prod_{ij\in C} W_{ij}\Big\} e^{-S[W]}}{\int \mathcal{D}W e^{-S[W]}} \tag{4.19}$$

In order to give some physical intuition for the Wilson loop one can take a timelike rectangular loop on the lattice, as the one presented in Figure 4.5. According to Wilson' s string model for confinement this loop can be interpreted as a test pair $q\bar{q}$ (meson) that is created on the lattice separated by distance $R$. The pair is steadily displaced for time $T$, when it annihilates. [3] For large times the Wilson loop must give the energy of separation of the pair.

$$\langle W(R,T)\rangle \xrightarrow{T\to\infty} e^{-E(R)T} \tag{4.20}$$

If the confinement picture the energy of the pair is considered linear when the pair is largely separated:

$$E(R) \xrightarrow{R\to\infty} KR \tag{4.21}$$

So, for a large enough Wilson loop on the lattice an expectation value that respects the above physical arguments is expected. The following area law will be confirmed on the lattice with the strong coupling expansion in the next section.

$$\langle W(R,T)\rangle \xrightarrow[R\to\infty]{T\to\infty} e^{-KRT} = e^{-KA(C)} \tag{4.22}$$

Gluons carry color charge. This property make possible for gluons to self-interact. This self-interaction can lead to producing massive colorless particles constituting only from gluons. These particles are called glueballs. The lowest possible mass of such a particle is referred to as mass gap.[4] The lattice has provided proof that this mass gap exist for non-Abelian gauge theories.

[3] According to Feynman-Stuckelberg interpretation antiparticles travel backwards in time.

[4] The theoretical proof of the existence of the aforementioned mass gap is a Millenium Prize Problem.

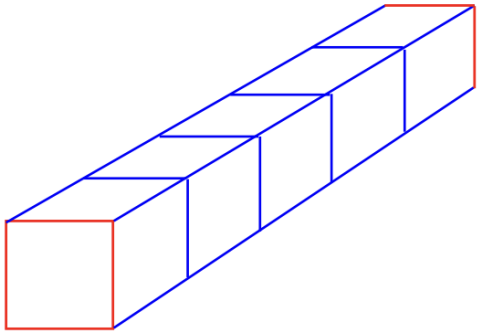


**Figure 4.6:** Two separated plaquettes (red) on the lattice.

It is not in the purposes of the present thesis to measure the mass gap on the lattice with Monte Carlo simulations, that is why the following discussion appears only for completeness purposes.

The mass gap is the best order parameter, because it is the only one so far that does indeed vanish in the weak coupling regime and takes a specific value in the confined phase. On the lattice the mass gap can be extracted by studying the correlation function between two spatially separated plaquettes. This function decays with a decay rate depending on the mass gap when separated by distance $R$.

$$\langle \mathrm{Tr}\{W_{p_1}\}\,\mathrm{Tr}\{W_{p_2}\}\rangle = Ce^{-m_g R} \tag{4.23}$$

## 4.5 Strong Coupling Expansion

When the coupling constant $g$ is small enough the action can be expanded around it. But, that is what is known already from perturbation theory with Feynman' s diagrams. One of the main purposes of using the lattice though, is to produce results in a region, where perturbation theory cannot be applied, meaning in large separations (confined phase), where the coupling constant takes large values.

In the confined region $g$ is large, but the parameter $\beta$ appearing in the Wilson action is proportional to $g^{-2}$, which makes it rather small. An expansion around $\beta$ is known as the strong coupling expansion.

This expansion is needed in order to predict the behavior of physical values in the strong coupling limit. The physical quantities are expressed as gauge invariant functionals of $W_{ij}$. For example, the expectation value of a Wilson loop is given by Equation 4.19. Expanding the action 4.7 around $\beta$ one gets:

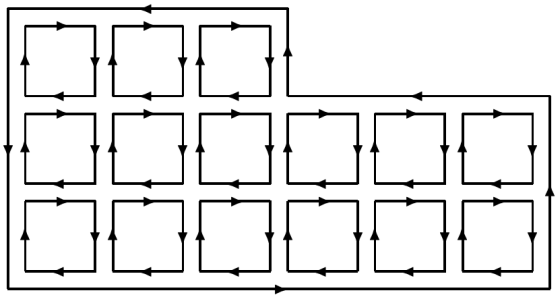


**Figure 4.7:** A Wilson loop on the lattice tiled with plaquettes.

$$e^{-\beta S[W]} = 1 + \beta \sum_{\square} \Big(1 - \frac{1}{N} \operatorname{Tr}\{\operatorname{Re}(W_{\square})\}\Big) + \mathcal{O}\Big(\beta^2\Big) \tag{4.24}$$

As a result, when $\beta \to 0^+$, the Wilson loop will "decrease" to an infinite integration with integrals of the form 4.16, which are zero. The simplest way for the Wilson loop to give a non-vanishing integral is by multiplying every $W_{ij}$ with its inverse $W_{ij}^{-1}$. Then the integral 4.17 can be used. The inverse link variables for each link in the functional will be provided by the action.

The simplest way to achieve a non-vanishing expectation value of the Wilson loop in the strong coupling expansion is to tile the loop with plaquettes. The plaquettes must provide the inverse link variable for every link variable of the Wilson loop and of course, the appropriate link variable for the other plaquettes. The method can be seen visually in Figure 4.7. Of course, the appropriate link variables can be provided to the loop and to the other plaquettes by a lot different ways, for example by using plaquettes belonging to another plane. These are higher order expansions.

For a rectangular Wilson loop with dimensions $R, T$, as the one shown in Figure 4.8, the expectation value using the expanded action is:

$$\langle W(C) \rangle = Z^{-1} \int \mathcal{D}W \left[1 + \beta \sum_{\square} \Big(1 - \frac{1}{N} \operatorname{Tr}\{\operatorname{Re}(W_{\square})\}\Big)\right] \frac{1}{N} Tr \left[\prod_{i<j \in C} W_{ij}\right] \tag{4.25}$$

After the loop is tiled with plaquettes the order of this expectation value can be calulated as follows.

- A factor $\beta/(2N)$ will come from each plaquette.
- A factor $1/N$ will come from each link that is integrated according to the integral 4.17.
- A factor of $N$ will come for each site of the lattice, because of the $\delta$ matrices of the integral 4.17.

The final result is:

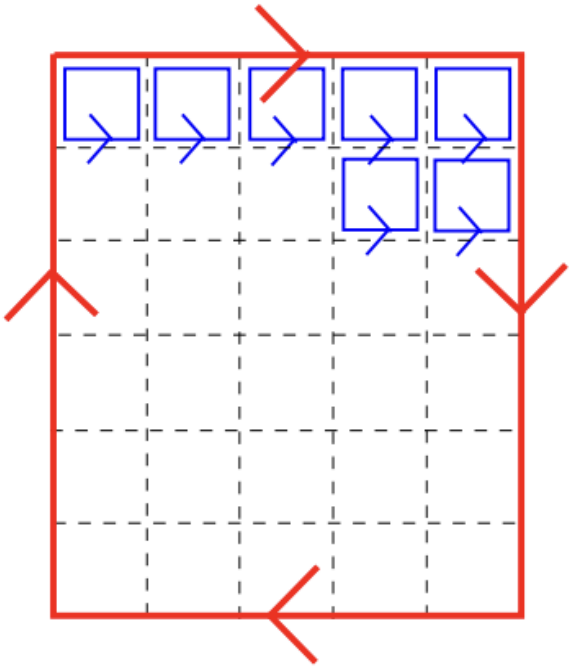


**Figure 4.8:** A rectangular Wilson loop tiled with plaquettes. These are the Wilson loops calculated from the Monte Carlo simulations.

$$\begin{aligned}\langle W(C)\rangle &\propto \left(\frac{\beta}{2N}\right)^{\#plaquettes}\left(\frac{1}{N}\right)^{\#links} N^{\#sites}\\ &= \left(\frac{\beta}{2N}\right)^{RT}\left(\frac{1}{N}\right)^{T(R+1)+R(T+1)} N^{(R+1)(T+1)}\\ &= \left(\frac{\beta}{2N^2}\right)^{RT}\end{aligned} \tag{4.26}$$

This result already reproduces the area law that is expected for the Wilson loop in the strong coupling limit. The same procedure can be performed for the mass gap, where the "tube" that connects the plaquettes (see Figure 4.6 must be tiled with elementary plaquettes on every plane.

## 4.6 Renormalization Using the Lattice

The term renormalization is used in quantum field theories to describe the process, in which divergent integrals describing physical quantities are turned into finite integrals. The program to renormalize such a theory consists of two steps.

The first step is to introduce an external parameter in the theory, such that all the divergent integrations are carried out in a finite domain. The results of the integrals become dependent

on this external parameter, that is why is known in the bibliography as a cutoff. There are many available regularization procedures, such as the Pauli-Villars regularization or the dimensional regularization. Imposing a lattice structure on spacetime is just another regularization scheme. Note that the physical results should be independent of the chosen regularization scheme.

After the cutoff is introduced the observable, in which the divergent integrals appeared, has acquired a dependency on the cutoff. The next step is to absorb this dependency by some of the free parameters of the theory. At this point, it should be stated that, as far as quantum filed theories concern, masses, charges and coupling constants of particles are free parameters. Stated differently, every Lagrangian describing a quantum field theory contains masses, charges and coupling constants that are not predicted by the theory itself. This makes them the perfect candidates to absorb the dependence on the cutoff and rescale to a new value without altering the form of the Lagrangian. This step is the renormalization. A theory where this program can run succefully is called renormalizable.

As mentioned the lattice spacing $a$ plays the role of the cutoff. Real physics must appear in the continuum limit $a \to 0^+$. In section 4.1, Wilson' s action was introduced and it was mentioned that when $a \to 0^+$ then this action reduces to the classical Yang-Mills action[5]. This is not enough when one considers the quantum theory. Independently of this succession, the Wilson' s action must be able to reproduce quantum amplitudes, where divergences are present. Hence, when removing the cutoff $a$, this divergences must disappear and give their place to finite quantities.

Consider a general observable $H$. The first step is to define the dependencies of this function. As stated, the only free parameter of the theory is the coupling constant $g_0$. The general idea of renormalization is to make the observables acquire an explicit dependency on the free parameters in a certain scale. The index 0 on $g_0$ is used to emphasize that $g$ depends on the scale. As explained in Section 4.2 every observable has dimensions $[H] = [M]^d$, where the dimensions in lattice units are understood as dependency on the lattice spacing $a$. As a result, $H$ depends on $g_0$ and $a$ and can be decomposed as:

$$H(a, g_0) = \left(\frac{1}{a}\right)^d \hat{H}(g_0) \tag{4.27}$$

In the continuum limit, the observable $H$ must remain finite in order to describe real physics.

$$\lim_{a \to 0^+} H(a, g_0) = H_{\text{phys}} \in \mathbb{R} \tag{4.28}$$

But, the factor including the $a$ diverges:

$$\lim_{a \to 0^+} \left(\frac{1}{a}\right)^d = \infty \tag{4.29}$$

[5]This can be seen from the Taylor expansion of Equation 4.5 around the lattice spacing $a$

These equations imply that the function $\hat{H}$, whose values is what are actually measured from the lattice must diverge. If one takes the observable to be a mass $m$, then:

$$\begin{aligned} m(a, g_0) &= \frac{1}{a\xi(g_0)} \\ \xi(g_0) &= (ma)^{-1} \end{aligned} \tag{4.30}$$

where $\xi$ is known from statistical mechanics as the correlation length. According to all the above $\xi$ must diverge. In the language of statistical mechanics this is the definition of a second order phase transition. In statistical mechanics phase transitions appear bellow a critical temperature. The role of temperature is taken by the free parameter $\beta$, so one expects a that there exists a certain critical value of $\beta$. Bellow this value the system must behave differently, or in the language of the strong interaction, the system confines.

Using the general idea of renormalization the free parameter $g_0$ must obtain a dependency on the cutoff $a$. Therefore, in all the above equations $g_0 = g_0(a)$. The important result comes from Equation 4.28. According to this limit, in order for $\hat{H}(g_0(a)) \in \mathbb{R}$ when $a \to 0^+$ this can happen only if $g_0$ acquires a specific value $g_{cr}$ at this limit.

$$\lim_{a \to 0} g_0(a) = g_{cr}$$

An important remark is that this value $g_{cr}$ must be chosen in a way that all observables lose their divergences simultaneously and in the continuum limit they acquire their physical value.

Therefore the first question that arises is to calculate this critical value of $g_0$. This can be calculated from the renormalization group equation for the pure Yang-Mills theory if one identifies the lattice spacing $a$ as a cutoff. The beta function is[6] [17]:

$$\gamma(g_0) = a\frac{\mathrm{d}g_0(a)}{\mathrm{d}a} = \gamma_0 g_0^3 + \gamma_1 g_0^5 + \mathcal{O}\left(g_0^7\right) \tag{4.31}$$

where the coefficients $\gamma_0, \gamma_1$ are independent of the renormalization scheme and for a theory without fermions as considered here are:

$$\gamma_0 = \frac{11N}{48\pi^2} \qquad\qquad \gamma_1 = \frac{17N^2}{384\pi^4} \tag{4.32}$$

Equation 4.31 is a simple first order differential equation that can be solved by separating the variables and partial fractions decomposition. For $g_0 \neq 0$

$$\frac{1}{\gamma_0 g_0^3 + \gamma_1 g_0^5} = -\frac{\gamma_1}{\gamma_0^2}\frac{1}{g_0} + \frac{1}{\gamma_0}\frac{1}{g_0^3} + \left(\frac{\gamma_1}{\gamma_0}\right)^2 \frac{g_0}{\gamma_1 g_0^2 + \gamma_0}$$

[6]The letter $\gamma$ is preferred for the beta function in order to avoid any confusion with the parameter $\beta$ that appears in the action.

The solution of the differential equation can be expressed after integration as:

$$\ln(\Lambda a) = -\frac{\gamma_1}{\gamma_0^2}\ln(g_0) - \frac{1}{2\gamma_0}\frac{1}{g_0} + \frac{\gamma_1}{2\gamma_0^2}\ln\Big(\gamma_1 g_0^2 + \gamma_0\Big)$$

where $\Lambda$ is the integration constant. In physical terms $\Lambda$ is there to preserve the non-existence of dimensions in the second part. Therefore $[\Lambda] = [L]^{-1} = [M]$. The $\Lambda$ term can be interpreted as the scale of the theory where $g_0$ takes a specific value, always near $g_0 \to 0$, because these are perturbative results.

With some simple algebra the solution can be expressed as:

$$\Lambda a(g_0) = \left(\gamma_1 + \frac{1}{\gamma_0 g_0^2}\right)^{\frac{\gamma_1}{2\gamma_0^2}} \exp\left\{-\frac{1}{2\gamma_0 g_0^2}\right\}$$

where the leading term near $g_0 \to 0^+$ is $1/g_0^2$:

$$a(g_0) = \Lambda^{-1}\left(\frac{1}{\gamma_0 g_0^2}\right)^{\frac{\gamma_1}{2\gamma_0^2}} \exp\left\{-\frac{1}{2\gamma_0 g_0^2}\right\} = \Lambda^{-1} f(g_0) \tag{4.33}$$

The last function has a singular point as $g_0 \to 0^+$. This is the desired divergence of the lattice spacing in order for the limit of 4.28 to be finite. So, the perturbative arguments above lead to the conclusion that $g_0 = 0$ is a possible critical point in order for observables to acquire a physical value in the continuum limit.

The important remark comes from Equation 4.27. In order for the continuum limit of this observable to exist and be finite the dimensionless function $\hat{H}$ must diverge as $\hat{H} \propto f(g)^d$. As a result the observable will acquire a physical value in the continuum.

$$H(a, g_0) = a^{-d} C f^d(g) = C\Lambda^d \tag{4.34}$$

The importance of the scale $\Lambda$ can be clearly seen. Every observable in the continuum limit can be expressed as a power of this scale. This is named dimensional transmutation, because after renormalization every observable depends on the parameter $\Lambda$, which has mass dimensions, rather than the dimensionless parameter $g_0$.

The really important result here stands on the remark that in order to take the continuum limit of a lattice gauge theory to describe real physics, perturbative results are needed. If the results from a simulation validate the divergence of the form $f^d(g)$ in the weak coupling regime then results in the strong-coupling regime can be reliable. More details on how the renormalization is performed using results from the lattice will be given in Section 4.10.

## 4.7 Transfer Matrix and Hamiltonian

Having defined the Wilson action on the discrete spacetime lattice it is time to retrieve the

transfer matrix formalism known from the path integral formulation of Quantum Mechanics [11, 9]. According, to Wilson' s idea, spacetime is Euclidean and constitutes of $N^4$ sites. There are $N$ in each of the 4 dimensions. One of them can be chosen arbitrarily to be the time dimension, because the spacetime is considered Euclidean, therefore there is no sign difference in the diagonal matrix elements of the metric tensor.

In Wilson' s lattice formalism the parallel transporters $W_{ij}$ between two neighboring sites $ij$ can be thought of as group elements that "lie" on the link that connects the sites. In path integral formulation the partition function of a quantum mechanical system can be expressed in terms of its transfer matrix operator as:

$$Z = \int \mathcal{D}[x(t)] e^{-S} = \mathrm{Tr}\Big\{\hat{T}^N\Big\} \tag{4.35}$$

where the operator $\hat{T}$ is the transfer matrix operator, acting on the Hilbert space of states.

$$\langle x'| \hat{T} |x\rangle = T_{x',x} \tag{4.36}$$

and $N$ is the number of timelike lattice sites. The states $|x\rangle, \langle x'|$ refer to two different time at time $t$ and $t'$ respectively with $t' > t$.

The purpose of this section is to define the transfer matrix operator in Yang-Mills theory and to retrieve a similar result for the partition function. Specifically, the form of the transfer matrix will be defined such that the following equation holds.

$$Z = \int \prod_{x,\mu} dW_{x\mu} e^{S[W]} = \mathrm{Tr}\Big\{\hat{T}^N\Big\} \tag{4.37}$$

where the product is performed in each lattice site $x$, in all the four dimensions $\mu \in \{1, 2, 3, 4\}$.

The first step is to define the Hilbert space of states. The action is defined as a functional of the group variables $W$, which means that these variables should be "promoted" to operators that act on a Hilbert space. In the Schrodinger picture the states are time dependent, which means that on the lattice the states must only depend on the spacelike links and their time evolution is expressed by spacelike links on the next time-layer.

From every site of the lattice 4 links (therefore 4 links variables $W$) emerge.[7] Three of them are spacelike and one is timelike. The three spacelike link variables at every site $W_{xm}, m \in 2, 3, 4$ take the role of the coordinate space of quantum mechanics. As a result, they form a basis for the Hilbert space with the following properties:

$$(\hat{W}_{xm})^{ab} |W\rangle = (W_{xm})^{ab} |W\rangle \tag{4.38}$$

where the indices $a, b$ are the explicit matrix indices of every group element $W_{xm}$. So, the basis $|W\rangle$ is the set of eigenvectors of the operators $(\hat{W}_{xm})^{ab}$. The basis is orthonormal and

[7] Of course, there are other 4 links in the opposite direction.

complete.

$$\langle W'|W\rangle = \prod_{x,m} \delta(W'_{xm}, W_{xm}) \tag{4.39}$$

$$\mathbb{1} = \int \prod_{x,m} dW_{xm} \, |W\rangle \, \langle W| \tag{4.40}$$

The first step is to rewrite the r.h.s of Equation 4.37, starting from the definition of the trace for an operator.

$$\mathrm{Tr}\Big\{\hat{T}^N\Big\} = \int \prod_{x,m} dW_{xmN} \, \langle W_{xmN}| \, \hat{T}^N \, |W_{xmN}\rangle = \int \prod_{x,m} dW_{xmN} \, \langle W_{xmN}| \, \hat{T} \cdot \hat{T} \ldots \hat{T} \, |W_{xmN}\rangle$$

The next step is to insert the completeness relation of Equation 4.40 between each pair of $\hat{T} \cdot \hat{T}$, which results in the following formula.

$$\mathrm{Tr}\Big\{T^N\Big\} = \int \prod_{n=1}^{N-1} \prod_{x,m} dW_{xmn} \, \langle W_{n+1}| \, \hat{T} \, |W_n\rangle \tag{4.41}$$

The number $n$ in the above equation is the number of time slice. Therefore the integration measure is not the same with the one appearing in the partition function. In Equation 4.41 the link variables are integrated over a whole time slice. The timelike link variables that connect one time slice with its next do not appear in the measure. So, this integration must occur from the matrix elements of the operator $\hat{T}$.

Because the transfer matrix is in principle a Green's function that propagates the system from a Hilbert space to the another Hilbert space in the next time slice, the action on the lattice must be re-expressed in a form where time and space are separated.

To re-express the action one can start from Equation 4.6 and separate the term to the spacelike and timelike plaquettes. Timelike are considered the plaquettes that have two links in the time direction and two links on a space direction. In contrary, spacelike plaquettes have all four links lying on space directions.

For timelike plaquettes denoted by $\square_t$: [8]

$$\mathrm{Tr}\{W_{\square_t}\} = N - \frac{g^2 a_s^2 a_t^2}{2} \mathrm{Tr}\Big\{F_{tx}^2\Big\} = N - V_c \frac{g^2 \frac{a_t}{a_s}}{2} \mathrm{Tr}\Big\{F_{tx}^2\Big\}$$

and for spacelike plaquettes denoted by $\square_s$:

$$\mathrm{Tr}\{W_{\square_s}\} = N - \frac{g^2 a_s^4}{2} \mathrm{Tr}\Big\{F_{xx}^2\Big\} = N - V_c \frac{g^2 \frac{a_s}{a_t}}{2} \mathrm{Tr}\Big\{F_{xx}^2\Big\}$$

[8] $N$ in the following formula is the dimension of gauge matrices, not the number of sites.

$a_s$ is the lattice spacing between two sites lying on a space direction, and accordingly $a_t$ is the lattice spacing in the temporal direction. $V_c = a_s^3 a_t$ is the volume of an elementary hypercube of the lattice. The action on this anisotropic lattice shifted by $N$ is:

$$\frac{1}{V_c} S = -\beta_s \sum_s \mathrm{Tr}\{W_{\square_s}\} - \beta_t \sum_t \mathrm{Tr}\{W_{\square_t}\} \tag{4.42}$$

The timelike and spacelike terms in the above action are multiplied by a different coupling constant defined as follows with the aid of the anistropy parameter $\xi$.

$$\xi = \frac{a_s}{a_t} \qquad \beta_s = \frac{2N}{g^2}\frac{1}{\xi} \qquad \beta_t = \frac{2N}{g^2}\xi \tag{4.43}$$

Therefore, the matrix elements $\langle W_{n+1}| \hat{T} |W_n\rangle$ appearing in Equation 4.41, must reproduce the exponential of the action. As mentioned before, an integration over the time like is missing from this formula in order to agree with the partition function. This integration must come from the matrix elements of the timelike plaquettes.[9]

$$\langle W_{n+1}| \hat{T} |W_n\rangle = \exp\left\{\beta_s \sum_s \mathrm{Tr}\{W_{\square_s}\}\right\} \prod_{x,m} \int dW_{1x} \exp\left\{\beta_t \sum_t \mathrm{Tr}\left\{W_{xm} W_{x+\hat{m}1} W'^{-1}_{mx} W^{-1}_{1x}\right\}\right\} \tag{4.44}$$

The final result is that the transfer matrix can be written as an exponential of the operator which is equivalent to the system' s Hamiltonian.

$$\hat{T} = K e^{-a_t \hat{H}} \tag{4.45}$$

$$Z = \mathrm{Tr}\left\{K e^{-N a_t \hat{H}}\right\} \tag{4.46}$$

This result can one more time be seen as an equivalence between statistical quantum system and lattice gauge theories. $\hat{H}$ is the systems Hamiltonian and $Na_t$ can be seen as the inverse temperature of the system.

## 4.8 Thermodynamics on the Lattice

In Section 4.7 a formula that connects the partition function of gauge theories with the transfer matrix was presented. According to this formula the Euclidean path integral can be interpreted as a quantum statistical system in three space dimensions and finite temperature $T$. The time dimension becomes the temperature of the statistical system.

[9]The number 1 denotes the time direction and $W'$ is the link variables in the next time slice.

In order to study the properties of pure Yang-Mills theory the expectation values of statistical variables must be defined. The energy density given in Equation 4.47 is such a variable. As mentioned in section 4.3 this expectation value must be expressed as a functional of gauge invariant quantities in order to achieve a non-zero expectation value.

$$\langle \epsilon \rangle = -\frac{1}{V_s} \frac{\partial(\ln(Z))_{V_s}}{\partial(1/T)} \tag{4.47}$$

where $V_s = n_s^3 a_s^3$ is the spatial volume of the lattice. The temperature depends on the time direction of the lattice, as mentioned in section 4.7, with $T = (n_t a_t)^{-1}$. Because the above relation demands the spatial volume to be held fixed, then to perform the differentiation the action on an anisotropic lattice is needed, in order to be able to change $a_s$ and $a_t$ independently. This action is given in Equation 4.42 with its parameter from Equation 4.43. The temperature can be then expressed as a function of the anisotropy parameter $\xi$.

$$T = \frac{1}{n_t a_t} = \frac{\xi}{N_t a_s} \tag{4.48}$$

$$d\left(\frac{1}{T}\right) = -\frac{N_t a_s}{\xi^2} d\xi \tag{4.49}$$

The partition function as well as the temperature are expressed as functions of $\xi$. Therefore, the differentiation to obtain the energy density can be expressed as:

$$\epsilon = \frac{1}{N_t N_s a_s^2 a_t^2} Z^{-1} \frac{\mathrm{d}Z}{\mathrm{d}\xi} \tag{4.50}$$

The dependence of the partition function on the anisotropy parameter $\xi$ comes from the coefficients $\beta_s$ and $\beta_t$ of Equation 4.42.

$$\epsilon \propto Z^{-1} \int \prod_{ij} W_{ij} \left[ -\frac{\mathrm{d}\beta_s}{\mathrm{d}\xi} \sum_s \mathrm{Tr}\{W_{\square_s}\} - \frac{\mathrm{d}\beta_t}{\mathrm{d}\xi} \mathrm{Tr}\{W_{\square_t}\} \right] e^{-S} = -\frac{\mathrm{d}\beta_s}{\mathrm{d}\xi} \langle W_{\square_s} \rangle - \frac{\mathrm{d}\beta_t}{\mathrm{d}\xi} \langle W_{\square_t} \rangle$$

where the derivatives are:

$$\frac{\mathrm{d}\beta_s}{\mathrm{d}\xi} = 2N \frac{d}{d\xi} \left\{ \frac{1}{\xi g^2} \right\}$$

$$\frac{\mathrm{d}\beta_t}{\mathrm{d}\xi} = 2N \frac{d}{d\xi} \left\{ \frac{\xi}{g^2} \right\}$$

The coupling constant $g$ depends on the anisotropy parameter, because according to the idea of renormalization it must depend on the lattice spacing. As a result its dependence can be expressed as follows.

$$\frac{1}{g^2(a_s,\xi)} = \frac{1}{g^2(a_s)} + C_s(\xi) \tag{4.51}$$

$$\frac{1}{g^2(a_t,\xi)} = \frac{1}{g^2(a_t)} + C_t(\xi) \tag{4.52}$$

$$g(a) = g(a_s,1) = g(a_t,1) \tag{4.53}$$

The derivatives become:

$$\frac{\mathrm{d}\beta_s}{\mathrm{d}\xi} = 2N\left(-\frac{1}{\xi^2 g^2(a)} + \frac{C_s(\xi)}{\xi^2} + \frac{C_s'(\xi)}{\xi}\right) \tag{4.54}$$

$$\frac{\mathrm{d}\beta_t}{\mathrm{d}\xi} = 2N\left(-\frac{1}{g^2(a)} + C_t(\xi) + \xi C_t'\xi\right) \tag{4.55}$$

The isotropic lattice can be reestablished by taking $\xi = 1$ after any differentiation. The above formulas allow the calculation of the energy density of a system put in a heatbath of finite temperature. A similar procedure can be followed to measure different thermodynamic observables.

## 4.9 Simulation Algorithms

Observables on the lattice are calculated by Equation 4.9. The path integral measure is a multi-dimensional group integral over all link variables. Therefore the dimension of this integral is the number of links contained in the 4-dimensional lattice used. Computers offer a much more affordable method to calculate those integrals via methods using the idea of important sampling. The idea is the same with statistical physics, where one does not need the data for all molecules contained in a glass of water to study its microscopical properties.

The multidimensional integral of Equation 4.9 can be interpreted in terms of probability theory as the mean value of the functional $H[W]$ with probability distribution $\exp\{-S[W]\}$. The basic idea of the Monte Carlo method is to find all the sets $\{W_{x\mu}\}$, which from now on will be called configurations $C$, that contribute more in the averaging of the functional $H[W]$. To be more specific, every link contains a link variable $W_{x\mu}$ the set $\{W_{x\mu}\}$, which contains all the link variables is a configuration of the lattice system. Even one different value for a link variable cahnges the configuration.

After finding the important configurations, the integral can be approximated according to the following formula, where it was assumed that the number of the important configurations is $n$:

$$\langle H\rangle = \frac{1}{n}\sum_{i=1}^{n} H(C_i) \tag{4.56}$$

### 4.9.1 Markov Chains

One of the most famous methods to generate the important configurations is the Markov Process or Markov Chain. According to this method, configurations are generated randomly. The probability to generate a new configuration $C'$ depends only on the previous configuration. More specifically, every configuration $C$ can have access to every other configuration $C'$ with probability $P(C, C')$. This can be thought of as matrix, where every element $P_{ij}$ denotes the probability for the system to pass from the configuration $C_i$ to the configuration $C_j$ with probability $P_{ij}$. The function $P(C, C')$ is called "transition probability" for obvious reasons.

Every element of $P_{ij}$ are probabilities, meaning that the usual two conditions must hold:

$$0 \leq P(C, C') \leq 1, \ \forall C, C' \tag{4.57}$$

$$\sum_{C'} P(C, C') = 1 \tag{4.58}$$

On the lattice the idea is that with subsequent changes of configurations new configurations with higher values of the Boltzmann factor $\exp\{-S[W]\}$ will be produced. Beyond a configuration the system will not be able to produce a configuration with higher Boltzmann factor. At this point the system has achieved its equilibrium configuration $C_{eq}$. It can be mathematically proven that Markov Chains can always converge to an equilibrium configuration $C_{eq}$, independently of the starting configuration when the probability matrix obeys certain conditions. When simulating lattice gauge systems the conditions are obeyed. For further details see [22].

Let $V^{(n)}(C)$ denote the so called state/configuration vector of the Markov Chain. This can be thought of as a column matrix, whose entries are the probability for the system to be in the each configuration $C$ after $n$ steps. As soon as the system has reached its equilibrium state it must remain in that state. This condition is expressed as:

$$V^{eq}(C) = \sum_{C'} P(C, C') V^{eq}(C') \tag{4.59}$$

Therefore the state vector of the equilibrium state must be an eigenvector of the transition probability. For lattice gauge theories Markov Chains must be able to sample the probability $\exp\{-S[W]\}$. A condition that gives this distribution and simultaneously satisfies 4.59 is the detailed balance condition:

$$e^{-S(C)} P(C, C') = e^{-S(C')} P(C', C) \tag{4.60}$$

The detailed balance condition is enough for the system to converge to the equilibrium configuration. Although Markov Chains give a condition for convergence they do not specify the transition probabilities $P(C, C')$, for each step in order to achieve the equilibrium. These probabilities are specified by the chosen algorithm.

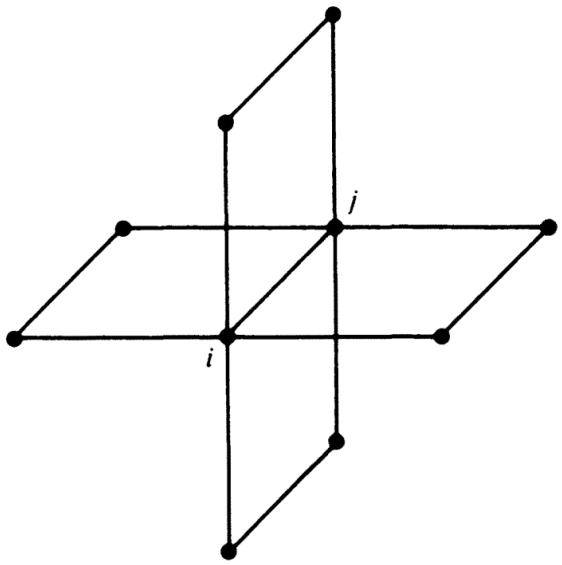



**Figure 4.9:** Four of the six plaquettes containing the link $ij$.The remaining two utilize the fourth dimension.[7]

### 4.9.2 Metropolis Algorithm

The Metropolis algorithm describes a method to advance a system from a configuration $C$ to a new configuration $C'$, that will have less distance from the equilibrium configuration.

The first step is to choose an initial configuration for the system. There are two such configurations that are commonly used. The first is referred to as cold or ordered start. All the variables $W_{x\mu}$ are chosen equal to the identity element of the group, which makes the Wilson' s action minimal, that is why is called cold start. The second commonly used initial configuration is the opposite. The action must be maximal, which can be achieved by giving to the link variables $W_{x\mu}$ completely random values.

The next step is to alter the configuration. As mentioned, a configuration $C$ consists of all the link variables of the lattice. This configuration is changed when even one link variable is changed. That is how the Metropolis algorithm changes the configuration. A candidate $W_{ij}$ is chosen and is altered by being multiplied by an appropriate random matrix $T \in SU(N)$ which has a weight towards the identity element $\mathbb{1} \in SU(N)$.

The change of a single link variable $W_{ij}$ is then accepted or rejected. First, it has to be mentioned that the change of $W_{ij}$ alters the action of the system, therefore the Boltzmann factor and the purpose of the algorithm is to minimize the action. The action is the sum of all elementary plaquettes. As a result, in a $4D$ lattice the change of one single link will affect only 6 plaquettes. A $3D$ sketch of this fact is given in Figure 4.9.

Considering the above the change of the action when only a single link is altered can be computed only by changing the link variable and multiplying it with its 6 surrounding

staples.[10]

$$\Delta S = \beta \,\mathrm{Tr}\Big\{\mathrm{Re}\Big\{\Delta W_{ij} \sum_{i=1}^{6} V_i\Big\}\Big\} \tag{4.61}$$

where $\Delta W_{ij} = W'_{ij} - W_{ij}$ and $V_i$ is the product of link variables for the six surrounding staples.

After calculating $\Delta S$ the change of the link variable $W_{ij}$ is accepted if $\Delta S \leq 0$. Otherwise, a random number $r \in [0, 1]$ is generated and the change is accepted only if $r \leq \exp\{-\Delta S\}$. This step gives the chance for new configurations with higher action to be accepted. This is desirable, because in that way the algorithm can escape for loops of metastable states.

The whole idea behind the Metropolis algorithm is to maximize the rate of accept for changes that minimize the action. In order to further maximize this rate one can repeat the change of a particular link several times, or as commonly said "hits". Physically, this correspond to submerging every link in a heatbath.

After the change of a single link is accepted or rejected another link variable is chosen. The choice of the new link can be made randomly or systematically. The change of a single variable is often called Monte Carlo step. After all the link variables of the lattice are changed [11] a Monte Carlo step is finished and the whole procedure is repeated.

When the configurations of several Monte Carlo steps do not appear to change significantly the system is considered to be in equilibrium state. The expectation value of an observable can be calculated by retrieving from the computer memory several configurations, after the equilibrium is reached, and applying Equation 4.56.

### 4.9.3 Heatbath Algorithm

The heatbath algorithm is another method to choose the new candidate for every link variable $W_{ij}$. In the case of Metropolis the new value $W'_{ij}$ is calculated by multiplication with an arbitrarily/empirically chosen matrix. This can lead to $W'_{ij}$ which does not belong in the desired Boltzmann distribution.

The heatbath algorithm considers new candidates $W'_{ij}$ that are selected among all possible values the follow the desired probability distribution $e^{-S(W'_{ij})}$. As a result, this algorithm makes no selection that will be rejected. In term of the variable $W_{ij}$ that is about to be changed in each step the probability distribution is:[12]

$$P(W) = e^{-S(W)} = e^{-\frac{\beta}{N}\,\mathrm{Tr}\{\mathrm{Re}\{WV\}\}}$$
$$dP(W) \propto e^{-\frac{\beta}{N}\,\mathrm{Tr}\{\mathrm{Re}\{WV\}\}} dW$$

[10] Staple is the term used in the bibliography for a plaquette without one link.
[11] Independently of whether the change was accepted or rejected.
[12] The indices $ij$ of the sites are omitted for simplicity

where $V$ is the sum over the staples containing the link $ij$ and $dW$ is the Haar measure. In order to generate group elements $W$ that follow the above probability distribution an explicit formula for the Haar measure must be known. This is the most serious disadvantage of this algorithm. In order to have a formula for the Haar measure a parametrization for the group manifold must be known, which is really difficult for complex manifolds. For example, a parametrization is known for $SU(2)$ but not for $SU(3)$, which is the group of interest for the Standard Model.

The following calculation is for the group $SU(2)$. The manifold of this group can be parametrize as a surface of a $4D$ of unit radius.

$$SU(2) = \{a_0 + i\vec{a}\cdot\vec{\sigma} | a_0^2 + \vec{a}^2 = 1\},\ a = (a_0, \vec{a}) \in \mathbb{R}^4 \tag{4.62}$$

As a result the Haar measure of $SU(2)$ can be expressed as:

$$dW = \frac{1}{2\pi^2}\delta(a^2 - 1)d^4a \tag{4.63}$$

The factor $(2\pi^2)^{-1}$ comes from the normalization condition 4.12 after taking into consideration that the surface area of a $4D$ sphere of unit radius is $2\pi^2$.

$SU(2)$ has another important property. The sum of any two of its elements is proportional to another element of the group. Therefore, the sum over the 6 staples can be written as:

$$V = \sum_{i=1}^{6} V_i = k\bar{U},\ \bar{U} \in SU(2) \tag{4.64}$$

$$k = \sqrt{\det\{V\}} \tag{4.65}$$

For SU(2) ($N = 2$) and the trace of $SU(2)$ matrices is a real number. The probability distribution is:

$$dP(W) \propto e^{-\frac{k\beta}{2}\mathrm{Tr}\{W\bar{U}\}}dW$$

The measure is right invariant, so one can redefine is through a matrix $X = W\bar{U}$.

$$dP(X) \propto e^{-\frac{k\beta}{2}\mathrm{Tr}\{X\}}dW$$

After substituting the Haar measure for the matrix $X$ from Equation 4.63 the probability distribution becomes:

$$dP(X) \propto \frac{1}{2\pi^2}e^{-k\beta x_0}\delta(a^2 - 1)d^4a$$

The measure on the $4D$ sphere can be expressed with the $3D$ solid angle as:

$$d^4a = \|a\|^2 d\Omega d\|a\| da_0 = \|a\|^2 \sin(\theta) d\theta d\phi d\|a\| da_0,\ \theta \in [0, 2\pi), \phi \in [0, \pi) \tag{4.66}$$

The $\delta$ function has the following useful property:

$$\delta(a^2 - 1) = \delta(\|a\|^2 - (1 - a_0^2)) = \delta\left(\|a\| - \sqrt{1 - a_0^2}\right) + \delta\left(\|a\| + \sqrt{1 - a_0^2}\right) \tag{4.67}$$

Substituting these equations to the probability distribution and using the delta function to "kill" the measure $d\|x\|$ one gets:

$$dP(X) \propto \frac{1}{2\pi^2}\sqrt{1 - a_0^2} e^{-k\beta x_0} \sin(\theta) d\theta d\phi da_0 \tag{4.68}$$

In the above distribution the variables factorize, therefore the distributions can be generated independently. First one generates random variables that follow the Gaussian distribution $\exp\{-k\beta x_0\}$ and accept or reject with probability $\sqrt{1 - a_0^2}$. The variables $\theta, \phi$ show the direction of the vector $\vec{a}$ and can be chosen at arbitrarily.

After generating the matrix $X$ through the vector $a$ following the precedent procedure, the matrix $W'$ is calculated by $X = W\bar{U}$.

Although this algorithm has an elegant formulation for the group $SU(2)$ there are no known extensions for groups of bigger matrices. The reason is that parametrizations of the group manifold are not as elegant as in the case of $SU(2)$. A way to generalize this algorithm for $SU(3)$ using its $SU(2)$ subgroups can be seen in [3].

## 4.10 Results from Simulations and Discussion

For the purposes of this thesis a Monte Carlo code was developed. The code is written in python and is based on open source code developed by Michael Creutz.[13] The code uses an isotropic lattice of $N^4$ sites and updates the link variables using the Metropolis algorithm described in Section 4.9.2 with 5 hits per link. Moreover the code can handle $SU(N)$ matrices up to $N = 20$. These matrices are saved using the parameters and the fundamental representation of the Lie Algebra. The purpose of the code is to measure Wilson loops in order to determine the parameter $\Lambda$ introduced in Section 4.6.

The plot of Figure presents the plaquette mean value as a function of $\beta$ for the group $SU(2)$. It is clear that the developed code is in well agreement with Creutz' s results in Figure 4.10. The agreement holds also for the groups $SU(3)$ (Figures 4.13 and 4.12), but for higher groups the results do not seem reliable as it is obvious from plots in Figures 4.14 and 4.15.

The last part of this thesis was to measure rectangular Wilson loops of higher dimensions on the lattice in order to calculate the $\Lambda$ parameter. To this end, the procedure introduced from

[13]The code can be downloaded for free from the official website https://latticeguy.net of Michael Creutz.

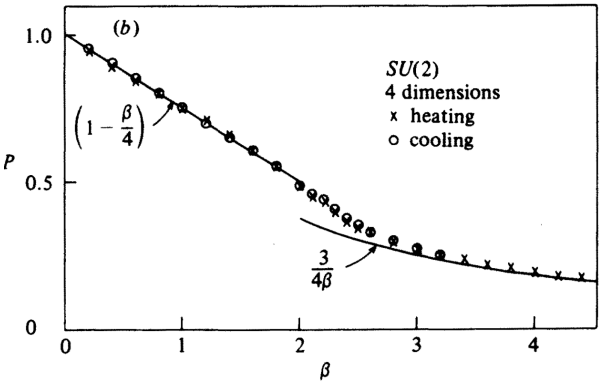



**Figure 4.10:** SU(2) simulation of the mean plaquette as a function of $\beta$ from [5]

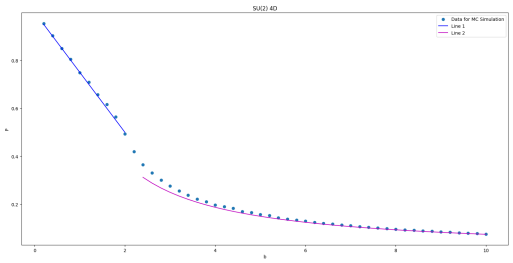



**Figure 4.11:** SU(2) simulation of the mean plaquette as a function of $\beta$ from the code developed for this thesis.

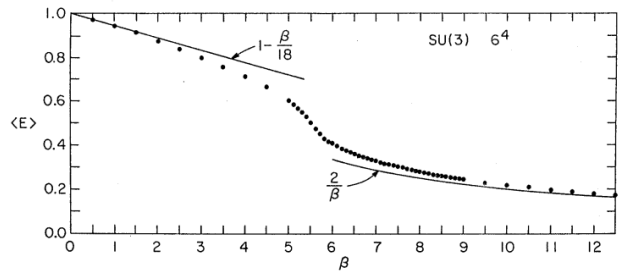



**Figure 4.12:** SU(3) simulation of the mean plaquette as a function of $\beta$ from [2]

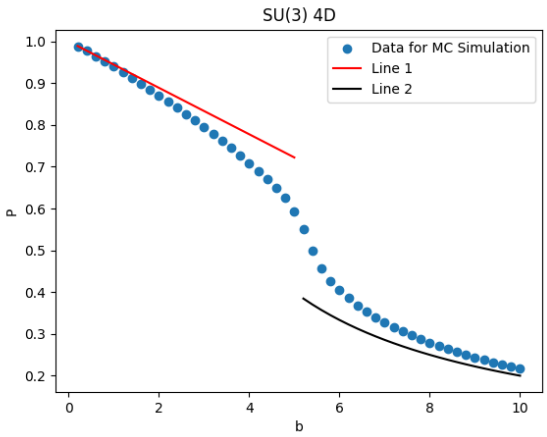



**Figure 4.13:** SU(3) simulation of the mean plaquette as a function of $\beta$ from the code developed for this thesis.

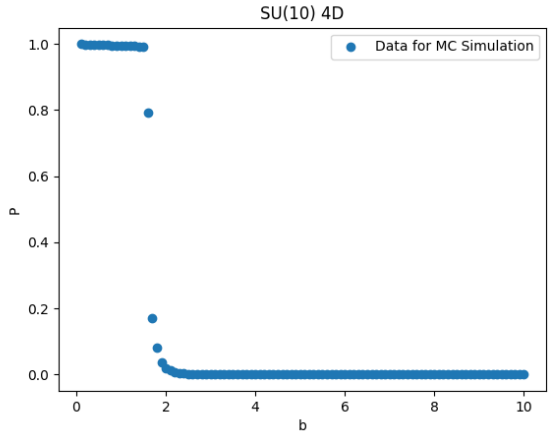



**Figure 4.14:** SU(10) simulation of the mean plaquette as a function of $\beta$ from the code developed for this thesis.

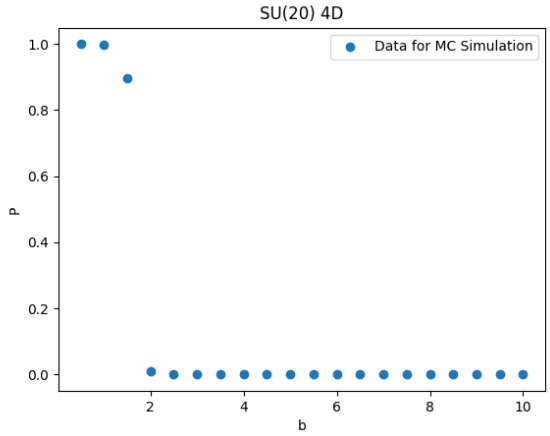



**Figure 4.15:** SU(10) simulation of the mean plaquette as a function of $\beta$ from the code developed for this thesis.

Creutz in [5] was followed. Creutz emphasized that the Wilson loop will always possess a perimeter and a Coulomb behavior. So, in order to minimize the error coming from these terms, he proposed the alternative observable $\chi(I, J)$ presented bellow, which is often called Creutz' s ratio. The advantage of this observable is that it is directly proportional to the string tension in physical units $\chi \propto a^2 K$.

$$\chi(I, J) = -\ln\left(\frac{W(I, J)W(I-1, J-1)}{W(I, J-1)W(I-1, J)}\right) \tag{4.69}$$

According to Section 4.6 $\chi(I, J)$ must behave as $a^2 K$, which in the weak coupling regime its behavior must follow that of Equation 4.33.

$$\chi \propto a^2 K = K\Lambda^{-2}\left(\frac{1}{\gamma_0 g_0^2}\right)^{\frac{\gamma_1}{\gamma_0^2}} e^{-\frac{1}{\gamma_0 g_0^2}} \tag{4.70}$$

Creutz plot this observable in a logarithmic scale as one can see in Figure 4.17. As mentioned in Section 4.2 every observable on the lattice is measured in dimensionless lattice units. Then to calculate real physical observables the dimensions are fixed by Equation 4.34. As a result, Creutz calculates the ratio $K\Lambda^{-2}$. After that, he ignores the asymptotic behavior of the curves and plots a band of straight lines to predict the interval where the observable $\chi$ would become zero. The parameter $\Lambda$ can be determined as the slope of the tangent lines. The plotted curves converge to a behavior in the strong coupling regime. As far as they agree with the expected weak coupling behavior and agree with each other for strong couplings, the strong coupling results are reliable.

Tries were made to reproduce this procedure with the code developed in python. The results are plotted in Figure 4.17. The result seems in good agreement with that of Figure **??** and

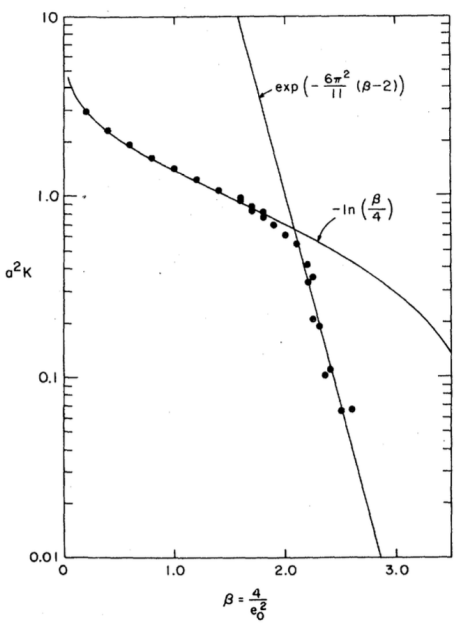



**Figure 4.16:** Calculation of the string tension $K$ from Creutz's ratios $\chi(I, J)$ [**?** ].

the string tension $K$ is determined from the slope of the tangent straight line appearing in the plot. The determined value is $\Lambda \approx 0.004\sqrt{K}$.

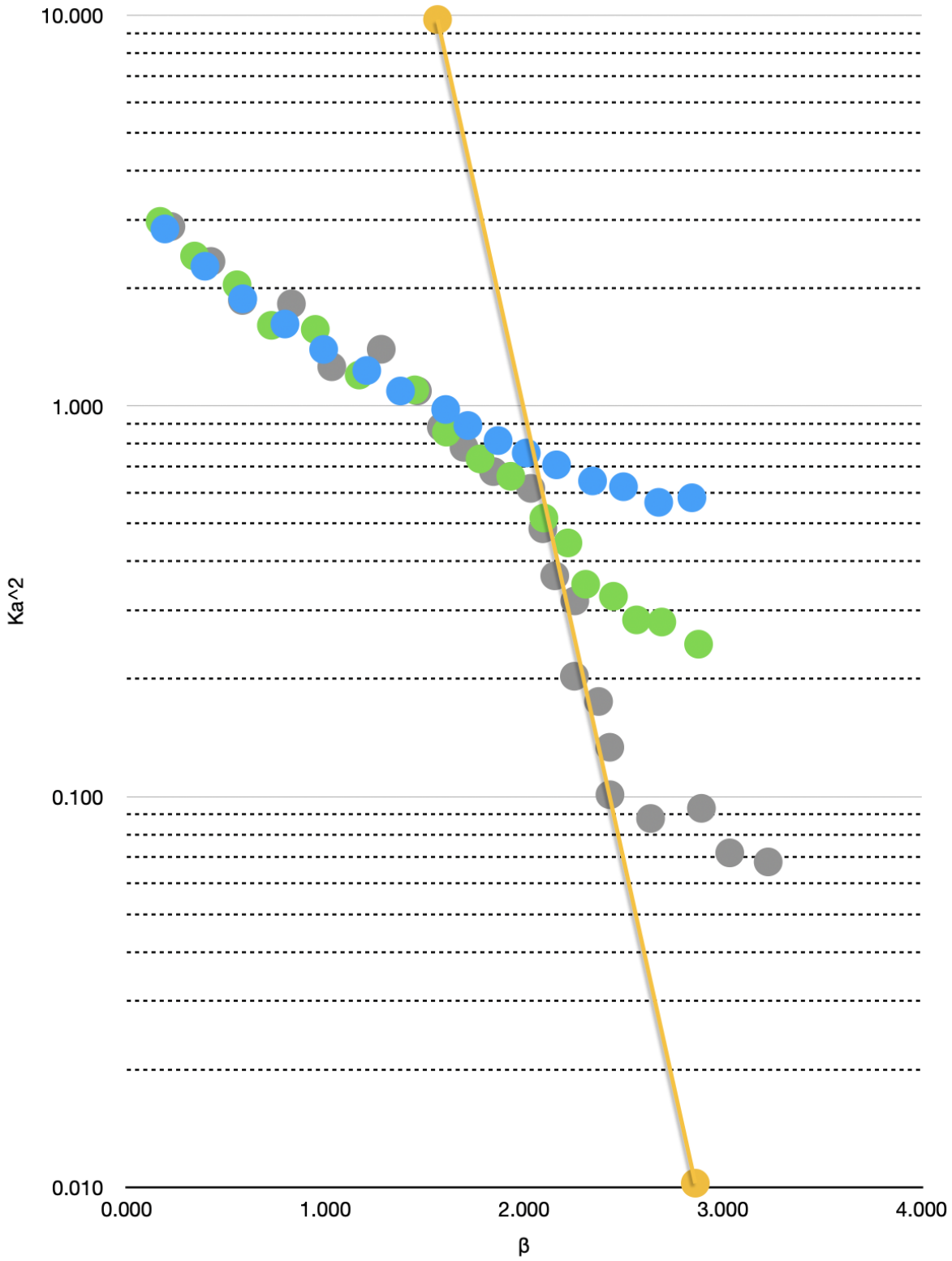



**Figure 4.17:** Data from simulating Creutz's ration. The behavior seems reliable for the plaquette but for bigger Wilson loops the behavior is not correct for weak couplings.

# Bibliography


[1] Baez, J., Muniain, J., 1994. Gauge Fields, Knots And Gravity. Series On Knots And Everything, World Scientific Publishing Company. URL: https://books.google.gr/books?id=qvw7DQAAQBAJ.

[2] Barkai, D., Creutz, M., Moriarty, K.J.M., 1984. MONTE CARLO STUDY OF SU(3) GAUGE THEORY ON A 12**4 LATTICE. Phys. Rev. D 29, 1207–1212. doi:10.1103/PhysRevD.29.1207.

[3] Cabibbo, N., Marinari, E., 1982. A New Method for Updating SU(N) Matrices in Computer Simulations of Gauge Theories. Phys. Lett. B 119, 387–390. doi:10.1016/0370-2693(82)90696-7.

[4] Creutz, M., 1978. ON INVARIANT INTEGRATION OVER SU(N). J. Math. Phys. 19, 2043. doi:10.1063/1.523581.

[5] Creutz, M., 1980. Monte Carlo Study of Quantized SU(2) Gauge Theory. Phys. Rev. D 21, 2308–2315. doi:10.1103/PhysRevD.21.2308.

[6] Creutz, M., 1981. Phase Transition in SU(5) Lattice Gauge Theory. Phys. Rev. Lett. 46, 1441. doi:10.1103/PhysRevLett.46.1441.

[7] Creutz, M., 1983. Quarks, Gluons and Lattices. Oxford University Press. doi:10.1017/9781009290395.

[8] Creutz, M., 2015. The lattice and quantized Yang-Mills theory, in: International Conference on 60 Years of Yang-Mills Gauge Field Theories. doi:10.1142/S021773231530027X, arXiv:1510.02059.

[9] Creutz, M., Freedman, B., 1981. A STATISTICAL APPROACH TO QUANTUM MECHANICS. Annals Phys. 132, 427. doi:10.1016/0003-4916(81)90074-9.

[10] Creutz, M., Jacobs, L., Rebbi, C., 1983. Monte Carlo Computations in Lattice Gauge Theories. Phys. Rept. 95, 201–282. doi:10.1016/0370-1573(83)90016-9.

[11] Feynman, R., Hibbs, A., Styer, D., 2010. Quantum Mechanics and Path Integrals. Dover Books on Physics, Dover Publications. URL: https://books.google.gr/books?id=JkMuDAAAQBAJ.

[12] Gattringer, C., Lang, C.B., 2010. Quantum chromodynamics on the lattice. volume 788. Springer, Berlin. doi:10.1007/978-3-642-01850-3.

[13] Griffiths, D.J., 2013. Introduction to electrodynamics. Pearson.

[14] Gross, D.J., Wilczek, F., 1973. Ultraviolet Behavior of Nonabelian Gauge Theories. Phys. Rev. Lett. 30, 1343–1346. doi:10.1103/PhysRevLett.30.1343.

[15] Hall, B.C., 2000. An elementary introduction to groups and representations. arXiv:math-ph/0005032.

[16] Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109. URL: https://doi.org/10.1093/biomet/57.1.97, doi:10.1093/biomet/57.1.97, arXiv:https://academic.oup.com/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf.

[17] 't Hooft, G., Veltman, M.J.G., 1972. Regularization and Renormalization of Gauge Fields. Nucl. Phys. B 44, 189–213. doi:10.1016/0550-3213(72)90279-9.

[18] Kogut, J.B., Susskind, L., 1975. Hamiltonian Formulation of Wilson's Lattice Gauge Theories. Phys. Rev. D 11, 395–408. doi:10.1103/PhysRevD.11.395.

[19] O'Neill, B., 1983. Semi-Riemannian Geometry With Applications to Relativity. ISSN, Elsevier Science. URL: https://books.google.gr/books?id=CGk1eRSjFIIC.

[20] Peskin, M.E., Schroeder, D.V., 1995. An Introduction to quantum field theory. Addison-Wesley, Reading, USA.

[21] Politzer, H.D., 1973. Reliable Perturbative Results for Strong Interactions? Phys. Rev. Lett. 30, 1346–1349. doi:10.1103/PhysRevLett.30.1346.

[22] Rothe, H.J., 2012. Lattice Gauge Theories : An Introduction (Fourth Edition). volume 43. World Scientific Publishing Company. doi:10.1142/8229.

[23] Savvidy, G., 2020. From heisenberg–euler lagrangian to the discovery of chromomagnetic gluon condensation. The European Physical Journal C 80, 165. URL: https://doi.org/10.1140/epjc/s10052-020-7711-6, doi:10.1140/epjc/s10052-020-7711-6.

[24] Schrödinger, E., 1985. Space-Time Structure. Cambridge Science Classics, Cambridge University Press.

[25] Schwartz, M.D., 2014. Quantum Field Theory and the Standard Model. Cambridge University Press.

[26] Smit, J., 2002. Introduction to Quantum Fields on a Lattice. Cambridge Lecture Notes in Physics, Cambridge University Press.

[27] Srednicki, M., 2007. Quantum field theory. Cambridge University Press. doi:10.1017/CBO9780511813917.

[28] Wilson, K.G., 1974. Confinement of Quarks. Phys. Rev. D 10, 2445–2459. doi:10.1103/PhysRevD.10.2445.

[29] Wilson, K.G., 1976. Quarks on a Lattice, or, the Colored String Model. Phys. Rept. 23, 331–347. doi:10.1016/0370-1573(76)90054-5.

[30] Wilson, K.G., 2005. The Origins of lattice gauge theory. Nucl. Phys. B Proc. Suppl. 140, 3–19. doi:10.1016/j.nuclphysbps.2004.11.271, arXiv:hep-lat/0412043.

[31] Yang, C.N., Mills, R.L., 1954. Conservation of isotopic spin and isotopic gauge invariance. Phys. Rev. 96, 191–195. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191, doi:10.1103/PhysRev.96.191.