RENORMALIZABILITY AND FINITENESS OF NONLOCAL QUANTUM GRAVITY
Quantum mechanics and general relativity are the two main heritages of the last century physics. Both theories have solid experimental foundations and are equally valid in their own fields of application. On the one hand, general relativity has explained phenomena like the precession of Mercury and predicted others long before they were experimentally seen, from black holes to the very recent verification of the existence of the gravitational waves. In sum, it explained the structure of the universe as a whole and the rules among macroscopic objects. On the other hand, the realm of quantum mechanics is the infinitesimally small universe: it explains the interactions among particles through quantum field theory and helped to predict the existence of new particles (the Higgs boson is one of the last examples). In the microscopic world Einstein’s theory does not play any relevant role, since its effects are negligible, allowing to consider flat space as an acceptable approximation.

When quantum field theory was developed in the second half of the twentieth century, it was natural to promote spacetime to a curved entity, in order to include general relativity in the quantum framework. It was suddenly clear that such an approach was unsatisfactory, since, due to many serious problems, quantum field theory was spoiled of one of its important features, the renormalizability. During the last decades, many different solutions have been proposed: some of them, like string theory and its further developments, introduce radical physical changes; other approaches are softer in the sense that they try to be as coherent with quantum field theory as possible. In this work we will explore one of the second possibilities, the nonlocal theory of quantum gravity that has recently aroused interest among some groups of theoretical physicists.

Our aim is to correctly incorporate gravity in quantum field theory, in order to handle it with the powerful tools that are extensively and successfully used in particle physics. The guiding principles designing the nonlocal theory are the following:

- **Perturbative renormalizability**: interactions can be depicted as an infinite series of Feynman diagrams, some of which can generate divergences, conveniently parametrized. The renormalization tells us how to control the infinities of the theory: in order to eliminate them, we must add counterterms to the bare Lagrangian, that have to be of a finite number of types to achieve renormalizability.

- **Lorentz invariance**: it is a well tested symmetry of nature and it should be a symmetry of quantum gravity as well.

- **Unitarity**: that is, calling $S$ the scattering matrix of some process, the condition $SS^\dagger = 1$ must be satisfied. Otherwise, the theory cannot be considered a faithful description of reality, since some pieces would actually be missing.

The only principle we decide to break is locality: we assume that the Lagrangian can contain terms that are nonpolynomial in the derivatives.

This work is organized into two parts: the first two chapters draw a path
that leads from Einstein's gravity to nonlocal gravity, while the last three chapters constitute a study of the main aspects of nonlocal gravity; in detail:

1. in the first chapter we start with a brief summary of Einstein’s theory of gravitation, after which we proceed with the usual quantization of the classical theory: Einstein’s gravitational Lagrangian is the starting point of the quantum theory, with a flat background and a particle - the graviton - describing the properties of the curved spacetime. What we obtain is a unitary theory which is not renormalizable, due to the presence of a coupling constant with negative dimension in units of mass;

2. in the second chapter we look at the solutions to the renormalization problem of quantum gravity. A first attempt was made in 1977 by Stelle [19], who presented a higher derivative version of Einstein’s theory: the renormalizability is restored, with the price of losing unitarity. We are naturally directed towards a nonlocal theory of gravity by introducing nonpolynomial expressions in the bare gravitational Lagrangian. We examine Kuz’min’s nonlocal theory of gravity, that has been amended with some additional hypotheses in the last few years;

3. in the third chapter we systematically study the principles of unitarity and causality as basic ingredients to build a field theory. We see in which sense local gauge theories are unitary and causal and we would like these features to be present in the nonlocal theory as well. We see that it is not simple to show that nonlocality is compatible with them and we do not give a final answer, since the research is still ongoing;

4. in the fourth chapter, we review the observations made in [24] and [8], in order to show a simple way to rewrite the nonlocal vertices. Then we present a unique and original complete demonstration of the super-renormalizability of the nonlocal theory of quantum gravity. We arrive at the conclusion that actually only four parameters are subjected to renormalization and, with a suitable choice of the nonlocal functions, only one loop diagrams diverge and the renormalization procedure does not even produce nonlocal counterterms;

5. finally, in the last chapter, in connection with some recent developments in [13] and [14], we add another requirement to Kuz’min’s theory: the finiteness, that is achieved by adding a potential at least cubic in Riemann tensor to the nonlocal gravitational Lagrangian. We compute explicitly the divergences for a minimal correction that is quartic in the graviton field; then, we propose an original, general method of dealing with corrections that are quartic in Riemann tensor.
1 FROM CLASSICAL TO QUANTUM GRAVITY 1
1.1 Describing a curved spacetime 1
1.2 The Einstein action 3
1.3 The graviton 4
1.4 Gauge fixing in the canonical formalism 7
1.4.1 Canonical formalism 8
1.4.2 Canonical transformations 10
1.4.3 Gauge fixing 11
1.5 Quantum gravity: gravity as a gauge theory 13
1.6 Divergences of quantum gravity 16

2 THE PROBLEM OF RENORMALIZABILITY 21
2.1 Stelle’s theory: higher derivative quantum gravity 21
2.2 The loss of unitarity in higher derivative models 22
2.3 Kuz’mín’s theory: nonlocal quantum gravity 24
2.4 Some properties of entire functions 25
2.5 Hypotheses for the nonlocal function 26
2.6 An explicit example of nonlocal function 28
2.7 The exponential function 31

3 SYSTEMATICS OF UNITARITY AND CAUSALITY 33
3.1 Unitarity and causality as basic requirements 33
3.1.1 Unitarity condition 33
3.1.2 Bogoliubov’s causality condition 34
3.2 Building up a new formalism 36
3.3 Cutting rules 40
3.4 The largest time equation and its consequences 40
3.4.1 Unitarity 42
3.4.2 Causality 43
3.5 Summing over intermediate states 45
3.6 The case of nonlocal theories 48

4 VERTICES AND POWER COUNTING 51
4.1 Vertices of nonlocal gravity 51
4.2 Properties of $\delta_1(z)$ 54
4.3 UV behavior of nonlocal vertices 58
4.4 UV behavior of diagrams 59
4.5 Structure of counterterms 61
4.5.1 Locality of counterterms 61
4.5.2 Correspondence between nonlocal and higher derivative theories 63
4.5.3 The renormalized Lagrangian of nonlocal quantum gravity 63

5 FINITE NONLOCAL GRAVITY 65
5.1 In search for a finite theory of quantum gravity 65
5.2 Influences of the new terms on renormalization 66
5.2.1 Renormalization of the curvature and the cosmological constant 67
5.2.2 Renormalization of the quadratic operators 68
5.3 An explicit example 70
5.4 The general case 73
A NOTATION AND USEFUL EXPANSIONS  79
A.1 Notation and definitions  79
A.2 Useful expansions  80
B RENORMALIZATION  81
B.1 Γ function and solid angle in D dimensions  81
B.2 Convergence  82
B.3 Feynman parameters and other useful integrals  82

BIBLIOGRAPHY  83
1.1 Describe a curved spacetime

Spacetime is a four dimensional continuum, described by a set of coordinates \( x^\mu \), where the 0-th component stands for time and the other three, that we will denote as \( x^i \), are the space components. General relativity broke down the Newtonian prejudice to think of spacetime as a flat entity, introducing the idea that is better explained by a manifold, which, in general, is curved and resembles flat spacetime only locally.

Explicitly, if we stick with a spacetime point \( P \) of the manifold, we can define the tangent space \( V_P \) to the manifold in \( P \), which is a four dimensional vector space as well. We can define a basis \( \{ e_\mu \} \), so that a vector \( v \) in \( V_P \) can be written as:

\[
v = v^\mu e_\mu \tag{1.1}\]

where \( v^\mu \) are the vector components and with the caveat that \( \{ e_\mu \} \) has only a local meaning.

Choosing different coordinates \( x'^\mu \), the new vector components are related to the old ones by the transformation rule:

\[
v'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} v^\nu \tag{1.2}\]

The vector defined in Eq. 1.1 is however a coordinate independent object and the basis vectors transform as:

\[
e'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} e_\nu \tag{1.3}\]

For instance, the partial derivatives \( \partial_\mu \) represent a good basis for \( V_P \).

The cotangent space \( V^*_P \) is defined as the space of linear maps \( f : V_P \to \mathbb{R} \); its basis vectors \( \omega^\mu \) are defined by their action over \( e_\mu \):

\[
\omega^\mu (e_\nu) \equiv \delta^\mu_\nu \tag{1.4}\]

A covector \( \alpha \) in \( V^*_P \) can then be represented as:

\[
\alpha = \alpha_\mu \omega^\mu \tag{1.5}\]

The components of a covector \( \alpha_\mu \) and the basis \( \omega^\mu \) transform according to:

\[
\alpha'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} \alpha_\nu, \quad \omega'^\nu = \frac{\partial x'^\nu}{\partial x^\mu} \omega^\mu \tag{1.6}\]

The differentials \( dx_\mu \) are an example of basis covectors. In general, we can define a \((l, k)\) type tensor \( T \):

\[
T = T^{\mu_1 \cdots \mu_l}_{\nu_1 \cdots \nu_k} e_{\mu_1} \otimes \cdots \otimes e_{\mu_l} \otimes \omega^{\nu_1} \otimes \cdots \otimes \omega^{\nu_k} \tag{1.7}\]
whose components transform as:

\[ T_{\nu_1 \mu_1}^{\mu_1 \ldots \mu_k} = \prod_{i=1}^{k} \frac{\partial x^{\nu_i}}{\partial x^{\mu_i}} \prod_{j=1}^{k} \frac{\partial x^{\nu_j}}{\partial x^{\mu_j}} T_{\nu_1 \ldots \nu_k}^{\mu_1 \ldots \mu_k} \]  

(1.8)

The metric

\[ g = g_{\mu \nu} \, dx^\mu \otimes dx^\nu \]  

(1.9)

is a (0, 2) type tensor, that is:

- symmetric: \( \forall v_1, v_2 \in V_p \) \( g(v_1, v_2) = g(v_2, v_1) \), that implies that \( g_{\mu \nu} \) is also symmetric;
- nondegenerate: if \( g(v, v_0) = 0 \) \( \forall v \), then \( v_0 = 0 \).

Equipped with these hypotheses, \( g \) can express the concept of infinitesimal squared distance. Furthermore, we can always find an orthonormal basis \( \{e_\mu\} \) so as to diagonalize \( g_{\mu \nu} \). We choose the following convention:

\[ g(e_\mu, e_\nu) = 0 \text{ if } \mu \neq \nu, \quad g(e_\mu, e_\mu) = +1, \quad g(e_\mu, e_\nu) = -1 \]  

(1.10)

and we will call \( \eta_{\mu \nu} \) its components.

From a physical point of view, Eq. 1.10 means that we can always find a reference system such that the metric reduces to the one of flat space; but this is true only locally, for a chosen point \( P \) of the manifold. On all the manifold, it is in general not possible to reduce the metric \( g_{\mu \nu}(x) \), defined differently in all the tangent spaces, to a unique diagonal form.

The nondegeneracy of \( g \) can allow us to interpret the metric as a linear map \( \nu \rightarrow g(\cdot, \cdot) \) from \( V_p \) into \( V_p^* \):

\[ \nu_\mu = g_{\mu \nu} v^\nu \]  

(1.11)

and, since it is an one-to-one map, the inverse also exists, defining a procedure to raise or lower the indices, according to convenience.

Introducing the notions of the differential calculus in curved spacetime is not trivial. A first difficulty is encountered in the definitions of derivatives: for instance, it does not make sense to consider quantities like \( \Delta A^\mu = A^\mu(P) - A^\mu(Q) \), given a vector field \( A^\mu \) and two points \( P \) and \( Q \) of the manifold, because \( A^\mu(P) \) and \( A^\mu(Q) \) belong to different vector spaces. Thus, we first need an operation that transports \( A^\mu(Q) \) into the vector space of \( A^\mu(P) \), such that \( \Delta A^\mu \) transforms simply according to the rules of the tangent space \( V_p \).

The solution is to use the covariant derivatives \( D_\mu \):

\[ D_\mu A^\gamma \equiv \partial_\mu A^\gamma + \Gamma^\gamma_{\mu \rho} A^\rho \]  

(1.12)

where, besides the usual differential \( \partial_\mu A^\gamma \), there is an additional piece that allows \( D_\mu A^\gamma \) to transform correctly; \( \Gamma^\gamma_{\mu \rho} \) is the connection and it is not a tensor (its explicit expression is given in Appendix A).

Let us consider a curve \( \mathcal{C} \), connecting two points, \( P \) and \( Q \), of the manifold. A vector \( A^\mu \) is said to be parallel transported along \( \mathcal{C} \) if

\[ T^\mu D_\mu A^\gamma = 0 \]  

(1.13)

is satisfied in each point of \( \mathcal{C} \), being \( T^\mu \) the vector tangent to \( \mathcal{C} \) in that point. This equation tells us the correct way to transport a vector from \( P \) to \( Q \).
If \( \mathcal{C} \) is a closed line, it is in general not true that a parallel transported vector comes back to its original orientation, after turning around. We can define the \( (0,2) \) type operator \( D_\mu D_\nu - D_\mu D_\nu \) that acts on covectors, producing a \( (0,3) \) tensor; it then can be rewritten as the action of a \( (1,3) \) tensor on covectors as:

\[
D_\mu D_\nu \omega_\rho - D_\nu D_\mu \omega_\rho = R_{\mu\nu\rho\sigma}^\sigma \omega_\sigma
\]  

(1.14)

\( R_{\mu\nu\rho\sigma}^\sigma \) is the Riemann tensor, whose explicit expression and symmetries are enlisted in Appendix A. Contracting the first and the third indices, we get the Ricci tensor \( R_{\mu\nu} \):

\[
R_{\mu\nu} = g^{\alpha\beta} R_{\alpha\mu\beta\nu}
\]  

(1.15)

whose trace is the scalar curvature \( R \):

\[
R = g^{\alpha\beta} R_{\alpha\beta}
\]  

(1.16)

If \( R_{\mu\nu\rho\sigma} \) is zero, then the Ricci tensor is zero and if \( R_{\mu\nu} \) is zero, then the curvature is also zero; the vice versas are not true.

### 1.2 The Einstein Action

The starting point of each quantum field theory is a classical action. In natural units, it is a dimensionless scalar quantity, that in curved spacetime can be cast in the form:

\[
S_{\text{cl}} = \int d^4 x \sqrt{-g} \mathcal{L}(x)
\]  

(1.17)

where \( \mathcal{L}(x) \), the Lagrangian density, is a local scalar function of the metric \( g_{\mu\nu} \) and \( \sqrt{-g} \) is necessary to make the integral measure invariant under changes of coordinates.

\( \mathcal{L}(x) \) has to be chosen carefully and some conditions have to be satisfied. Besides being a scalar built only with the covariant objects listed in Sec. 1.1, it should not contain second or higher order derivatives of the metric: this means that, in the equations of motion, derivatives of third or higher orders do not appear. It is essential, as we will discuss further in the following chapter, if we want to obtain stable equations of motion. Moreover, if we consider a pointlike mass as the source of the field, in the weak field limit, we would like to get the classical Newtonian equations of motions.

It is not possible to build a scalar only with the metric \( g_{\mu\nu} \) and \( \Gamma^\rho_{\mu\nu} \) that satisfies all these requisites, due to the nontensorial behavior of the connection \( \Gamma^\rho_{\mu\nu} \). The only scalar enlisted in Sec. 1.1, that is the scalar curvature \( R \), has the awkward feature that does contain second order derivatives, and the situation gets worse if we include contractions such as \( R_{\mu\nu} R^{\mu\nu} \). Thus, since we have no other ingredients, let us set:

\[
S_{\text{cl}} = \int d^4 x \sqrt{-g} \mathcal{L}(x) = -\frac{2}{\chi^2} \int d^4 x \sqrt{-g} R
\]  

(1.18)

where \( \chi \) is a constant with dimensions \(-1\). We will set:

\[
\chi = \sqrt{32\pi G}
\]  

(1.19)
where $G \approx 6.67 \times 10^{-8} \text{cm}^3 \text{g}^{-1} \text{s}^{-2}$ is Newton’s gravitational constant.

The equations of motions are now derived by the principle of least action:

$$\delta S_{\text{cl}} = 0$$

for arbitrary variations $\delta g^{\mu \nu}$ of the metric tensor.

It is immediately clear that, with respect to the field theories defined in flat spaces, one more variation - the one of the invariant measure - appears:

$$\delta S_{\text{cl}} = -\frac{2}{\chi^2} \int d^4x \left( (\delta \sqrt{-g}) R + \sqrt{-g} g^{\mu \nu} (\delta R_{\mu \nu}) + \sqrt{-g} (\delta g^{\mu \nu}) R_{\mu \nu} \right)$$

(1.21)

Let us look at the second term in 1.21. In a locally flat system, i.e. where $\Gamma_{\mu \nu}^\rho = 0$ in a certain point $P$:

$$g^{\mu \nu} \delta R_{\mu \nu} = g^{\mu \nu} (\partial_{\alpha} \delta \Gamma_{\mu \nu}^{\alpha} - \partial_{\nu} \delta \Gamma_{\mu \alpha}^{\mu})$$

(1.22)

The quantities $\delta \Gamma_{\mu \nu}^{\alpha}$ do transform like tensors: in fact, let us consider another point $P'$, infinitesimally close to $P$ and a vector $A_\mu$ defined in $P'$. Let us now parallel transport it to $P$ through two different displacements, one varying the connection, the other leaving it invariant. The difference of these two parallel transported vectors is $\delta \Gamma_{\mu \nu}^{\alpha} A_{\alpha} dx_{\nu}$ and transforms like a vector in $V_P$. Thus, we can set:

$$g^{\mu \nu} \delta R_{\mu \nu} \equiv \partial_{\alpha} w^{\alpha}$$

(1.23)

where $w^{\mu}$ is a four vector. This is of course not true for a generic reference system, for Eq. 1.23 is not covariant; indeed we can guess the general result by simply substituting $\partial_{\alpha} \rightarrow D_{\alpha}$ in the previous result:

$$g^{\mu \nu} \delta R_{\mu \nu} = D_{\alpha} w^{\alpha} = \frac{1}{\sqrt{-g}} \partial_{\alpha} (\sqrt{-g} w^{\alpha})$$

(1.24)

Therefore, when integrated in $\int d^4x \sqrt{-g}$, by the Gauss theorem, it reduces into an integral of $w^{\alpha}$ over the surface surrounding the whole volume, where the variations of the field are zero, and it does not contribute to the action: hence, in the equations of motion, derivatives of third or higher order do not appear.

Finally we get:

$$\delta S_{\text{cl}} = \int d^4x \delta L = -\frac{2}{\chi^2} \int d^4x \sqrt{-g} \left( R_{\mu \nu} - \frac{1}{2} g_{\mu \nu} R \right) \delta g^{\mu \nu} = 0$$

(1.25)

or, since the variations are arbitrary:

$$R_{\mu \nu} - \frac{1}{2} g_{\mu \nu} R = 0$$

(1.26)

These are the well known Einstein equations - without matter - and are clearly nonlinear.

### 1.3 THE GRAVITON

In quantum field theory, the classical interactions are interpreted as exchanges of virtual particles. The classical gravitational field, when promoted
to a quantum one, is not an exception: the influence of the spacetime curvature on the other particles can be explained as their interaction with a specific quantum field, that we call graviton.

Thus, let us see what properties we should expect for the graviton, starting from experimental facts.

**The graviton is massless** Let us suppose that the on-shell graviton has a mass $\mu > 0$; then, we should expect the static potential between two fermions to be Yukawa-like:

$$\varphi \propto e^{-\mu r} \quad (1.27)$$

with a range of order $1/\mu$. This can be useful to set an upper bound on the graviton mass. However, in this work, we shall directly make the assumption that the graviton is massless: it produces a Newtonian potential that falls off like $1/r$, without any exponential suppression.

**The graviton has integer, even spin** The existence of a static potential tells more about the spin of the graviton.

![Figure 1.1: Scattering between two spin-$\frac{1}{2}$ particles - $f$ and its antiparticle $\bar{f}$ - mediated by a virtual graviton.](image)

Recalling that the static potential is essentially the low energy limit of a scattering process like the one represented in Fig. 1.1, if its spin were half-integer, then the spins of the final particles would inevitably change and their matter structure would be different. Only integer spins can justify unaltered internal states. Moreover, an odd integer spin for the intermediate particles - just as in the case of photons - generates a repulsive potential if the two particles carry the same charges and an attractive one, if the charges are different. Then, we conclude that the spin has to be even.

**The graviton has spin 2** Now, let us suppose that the graviton is spinless, that is to say, it is described by a scalar field. The Fourier transform $G(k)$ of the propagator is then:

$$G(k) \propto \frac{1}{k^2} \quad (1.28)$$

When we compute scattering amplitudes, such as the one represented in Fig. 1.1, we should contract the indices of the propagator with the indices of the stress-energy tensors of the two particles and, since in this case it does not carry any Lorentz index, the only possibility is to contract separately the indices of the stress tensors:

$$T^\alpha_\alpha \frac{1}{k^2} T^\beta_\beta \quad (1.29)$$
Thus, a spin-0 particle can couple only with the trace of the stress-energy tensor. The Pound-Rebka experiment in 1959 or its more recent version, the Vessot experiment in 1980, showed that the photon *feels* the spacetime geometry where it is moving. Its energy - thus its frequency - changes according to the intensity of the gravitational potential. If we call $\varphi_E$ the potential in the point where the photon is emitted and $\varphi_R$ the one where it is detected, then its frequency $\omega$ changes as:

$$\Delta \omega = (\varphi_E - \varphi_R) \omega$$

(1.30)

In particular, if the photon moves towards regions of lower potential (i.e. $\varphi_E < \varphi_R$), it experiences a redshift. From the point of view of a quantized field theory, this is equivalent to say that the photon interacts with gravitons. But a spin-0 graviton would forbid such a possibility, for the electromagnetic-stress tensor in 1.29 is traceless. Therefore, we are left with the possibility that the graviton has integer, even spin, at least 2. There are no physical reasons to discard a spin-2 graviton: hence, we suppose the graviton to be a spin-2 particle.

But we would like to represent a spin-2 particle as a convenient covariant object, that is, through a representation of the Lorentz group. We shall consider the group $SU(2) \otimes SU(2)$, that is homomorphic to the restricted Lorentz group $SO(3,1)^+$ and classify the representations according to two spins as $(s_1, s_2)$. For example, a spin-1 particle, is described by a vector $A_\mu$, the $(1/2, 1/2)$ representation. However, a four-vector is the sum of two $SU(2)$ irreducible representations:

$$A_\mu \rightarrow \left(\begin{array}{c} 1/2 \\ 1/2 \end{array}\right) \otimes \left(\begin{array}{c} 1/2 \\ 1/2 \end{array}\right) = 0 \oplus 1$$

(1.31)

Thus, the spin-1 representation is encoded in $A_\mu$, but it is also a redundant representation, since it includes a spin-0 component. In addition, in the case of a massless spin-1 particle - the photon, for instance - there are only two physical degrees of freedom, against the four written in $A_\mu$. Following the same steps, in order to find a spin-2 representation, we can look at a rank 2 tensor $\tau_{\mu \nu}$, which can be thought of as a dyad $v_\mu \otimes w_\mu$ and corresponds to the sixteen-dimensional representation:

$$\tau_{\mu \nu} \rightarrow \left(\begin{array}{c} 1 \\ 2/2, 1/2 \end{array}\right) \otimes \left(\begin{array}{c} 1 \\ 2/2, 1/2 \end{array}\right) = 0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 2$$

(1.32)

or, in terms of irreducible representations:

$$\tau_{\mu \nu} \rightarrow 0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 2$$

(1.33)

But the tensor $\tau_{\mu \nu}$ can be written as a sum of a symmetric part $\phi_{\mu \nu}$ and an antisymmetric one $\psi_{\mu \nu}$:

$$\tau_{\mu \nu} = \frac{1}{2} (\tau_{\mu \nu} + \tau_{\nu \mu}) + \frac{1}{2} (\tau_{\mu \nu} - \tau_{\nu \mu})$$

(1.34)

An antisymmetric tensor can always be reduced to two spin-1 representations: $\psi_{0 \ell}$ and $\varepsilon^{ijk} \psi_{jk}$ - in the case of the electromagnetic tensor, they would correspond to the electric and the magnetic field. Hence:

$$\psi_{\mu \nu} \rightarrow 1_e \oplus 1_b$$

(1.35)
where we have added the subscripts e and b to distinguish the two subspaces. Thus, from \(1.33\) the ten-dimensional symmetric tensor representation is:

\[
\phi_{\mu\nu} \rightarrow o_w \oplus o_s \oplus t_m \oplus 2
\]

where \(o_w\) denotes the spin-0 subspace \(\phi_{00}\) (the energy), \(o_s\) is the stress scalar \(\phi_{0i}\), \(t_m\) stands for the momentum vector \(\phi_{0i}\). Having assumed the graviton massless, it has only two degrees of freedom and there are eight unphysical components written in \(\phi_{\mu\nu}\).

Therefore a symmetric rank two tensor \(\phi_{\mu\nu}\) is a good candidate to describe the graviton field. Actually, we already have a symmetric tensor that describes the curvature of the spacetime, that is the metric \(g_{\mu\nu}\), but it will be more convenient to set:

\[
g_{\mu\nu} \equiv \eta_{\mu\nu} + \chi \phi_{\mu\nu}
\]

Being \(g_{\mu\nu}\) dimensionless, with this definition, \(\phi_{\mu\nu}\) gets the dimension of an energy (we will explain the reason of this choice in Sec. 1.5). However, just as in the case of photon, the redundancy forces us to treat carefully a quantized version of gravity. And the clear nonlinearity of the field equations \(1.26\) gives also a first insight of the theory we are going to deal with: quantum gravity is a gauge theory, like QED and QCD.

### 1.4 Gauge Fixing in the Canonical Formalism

Gauge invariance is responsible of a degeneracy that makes the propagator impossible to be computed. The solution is to correct the theory by adding a gauge fixing term to the Langrangian, that breaks gauge invariance and allows us to compute the propagator. At a second stage, we must show that the physical results cannot depend on this arbitrary choice.

There are many available methods to handle gauge theories and give a systematic procedure of gauge fixing; we will use a very powerful and elegant formalism developed by Batalin and Vilkovisky during the ’80s ([1], [28] and the original paper [2]), called canonical formalism, that resembles the one used in classical mechanics.

First, let us consider a theory described by some classical fields, that we collect into a single row:

\[
\Phi_{\text{cl}}^i = (\phi_{\mu\nu}, A_{\mu}^a, \bar{\psi}, \psi, \phi)
\]

\(\phi_{\mu\nu}\) represents a spin-2 field - the graviton, for instance \(A_{\mu}^a\) are spin-1 fields - that is to say, photons in QED or gluons in QCD \(\bar{\psi}\) and \(\psi\) are (sets of) spinor fields and \(\phi\) are scalar fields. In all the applications of this work, we will consider only the graviton field, and the row \(1.38\) actually reduces to the single \(\phi_{\mu\nu}\).

Let us consider a transformation of these fields, parametrized by a set of parameters \(\Lambda\):

\[
\delta_{\Lambda} \Phi_{\text{cl}}^i = R_{\text{cl}}^i(\Phi_{\text{cl}}, \Lambda)
\]

under which the action is invariant, that is:

\[
0 = \delta_{\Lambda} S_{\text{cl}} = \int \delta_{\Lambda} \Phi_{\text{cl}}^i \frac{\delta S_{\text{cl}}}{\delta \Phi_{\text{cl}}^i} = \int R_{\text{cl}}^i \frac{\delta S_{\text{cl}}}{\delta \Phi_{\text{cl}}^i}
\]
The transformation in Eq. 1.39 satisfies the closure relation:

\[ [\delta_{\Lambda}, \delta_{\Sigma}] = \delta_{\Delta(\Lambda, \Sigma)} \]  

(1.41)

We will see that Batalin-Vilkovisky’s canonical formalism can gather all this information into a single equation.

1.4.1 Canonical formalism

The canonical formalism requires three basic ingredients. First, once we quantize the theory, the classical fields \( \Phi_{\alpha}^{cl} \) are not enough and we must enlarge them to include the ghosts, that is to say fields with fermionic statistics, one for each generator of the group of the transformation 1.39. In our case, with the gravitational field only:

\[ \Phi_{\alpha} = (\phi_{\mu\nu}, C_{\mu}) \]  

(1.42)

and \( C_{\mu} \) are the ghosts associated with a local coordinate transformation that has four degrees of freedom. Now, let us define the conjugate sources:

\[ K_{\alpha} = (K_{\mu\nu}^{0}, K_{C}^{\mu}) \]  

(1.43)

such that their statistics are opposite to those of their respective fields: calling \( \varepsilon_{\Phi_{\alpha}} \) the statistics of the fields (equal to 0 if the field is bosonic, 1 if fermionic), the statistics of the sources \( \varepsilon_{K_{\alpha}} \) are defined as:

\[ \varepsilon_{K_{\alpha}} = \varepsilon_{\Phi_{\alpha}} + 1 \mod 2 \]  

(1.44)

The last ingredient that we need is a set of functions of the fields \( R_{\alpha} \). \( R_{\mu} \) are defined starting from 1.39, substituting the parameter \( \Lambda \) with \( \theta C \) (where \( \theta \) is a constant anticommuting parameter, while \( C \), the ghosts, are also anticommuting but point dependent), moving \( \theta \) to the far left and finally dropping it:

\[ \theta R_{\mu}(\Phi_{\alpha}^{cl}, C) \equiv R_{\mu}^{cl}(\Phi_{\alpha}^{cl}, \theta C) \]  

(1.45)

\( R_{\mu}^{cl}(\Phi_{\alpha}^{cl}, C) \) can differ from \( R_{\mu}^{cl}(\Phi_{\alpha}^{cl}, \theta C) \) only by sign. For the ghost index, the functions \( R_{\alpha} \) are defined starting from the closure relation 1.41, by:

\[ R_{C}^{\mu}(\theta\Lambda + \theta'\Sigma) = -\theta\theta' \Delta(\Lambda, \Sigma) \]  

(1.46)

with \( \theta \) and \( \theta' \) anticommuting parameters.

Given two functionals \( X \) and \( Y \), let us define their antiparenthesis as:

\[ (X,Y) = \int d^{4}x \left( \frac{\delta_{r}X}{\delta\Phi^{\alpha}} \frac{\delta_{l}Y}{\delta K_{\alpha}} - \frac{\delta_{r}X}{\delta K_{\alpha}} \frac{\delta_{l}Y}{\delta\Phi^{\alpha}} \right) \]  

(1.47)

where the fields are all evaluated in the same spacetime point; \( \delta_{r} \) and \( \delta_{l} \) denote respectively right and left derivatives, necessary when dealing with fermionic objects.

They satisfy the following properties:

\[ (Y,X) = -(-1)^{(\varepsilon_{X}+1)(\varepsilon_{Y}+1)}(X,Y) \]  

(1.48)

\[ -(-1)^{(\varepsilon_{X}+1)(\varepsilon_{Y}+1)}(X,(Y,Z)) + \text{cyclic permutations} = 0 \]  

(1.49)

If we consider a purely fermionic functional \( F \), we trivially get:

\[ (F,F) = 0 \]  

(1.50)
And for a purely bosonic functional $B$:

$$(B, B) = 2 \int d^4 x \frac{\delta B}{\delta \Phi^\alpha} \frac{\delta B}{\delta K_\alpha} = -2 \int d^4 x \frac{\delta B}{\delta K_\alpha} \frac{\delta B}{\delta \Phi^\alpha} \tag{1.51}$$

The nontriviality of the antiparentheses for bosons allows us to give a new definition for the quantized action.

The action $S(\Phi, K)$ is defined as solution of the master equation:

$$(S, S) = 0 \tag{1.52}$$

with the boundary conditions:

$$S(\Phi, 0) = S_{cl}(\Phi_{cl}), \quad \left. -\frac{\delta S(\Phi, K)}{\delta K_1} \right|_{K=0} = R^i(\Phi_{cl}, C) \tag{1.53}$$

The minimal solution of the master equation is:

$$S(\Phi, K) = S_{cl}(\Phi_{cl}) + S_K(\Phi, K) \tag{1.54}$$

where $S_{cl}$ is the classical action and $S_K$ is defined as:

$$S_K(\Phi, K) = - \int R^\alpha(\Phi) K_\alpha = - \int R^i(\Phi_{cl}, C) K_i - \int R^C(\Phi) K^\mu_C \tag{1.55}$$

A direct consequence of the master equation and Eq. 1.49 is that, for each functional $X$:

$$(S, (S, X)) = 0 \tag{1.56}$$

Moreover, the fields $R^\alpha$ can be also seen as the antiparenthesis of the action and the fields:

$$R^\alpha(\Phi) = (S, \Phi^\alpha) \tag{1.57}$$

The set of fields $R^x$ defines the BRST transformations of the fields (after the names of their discoverers Becchi, Rouet, Stora [3] and Tyutin [25]):

$$\delta_{BRST} \Phi^\alpha = \theta R^\alpha(\Phi) \tag{1.58}$$

where $\theta$ is anticommuting infinitesimal parameter and the nilpotent operator $(S, \cdot)$ that generates the infinitesimal transformations is sometimes called BRST operator. We will see that they play a crucial role in the gauge fixing procedure, since they constitute the residual symmetry of the theory after applying the gauge fixing and can be used (following, in reverse, the steps that brought us to the definition of $R^\alpha$) to recover the symmetry of the original theory.

Using Eq. 1.51, the master equation can be rewritten as:

$$\int d^4 x R^\alpha(\Phi) \frac{\delta S}{\delta \Phi^\alpha} = \int d^4 x \left( R^i(\Phi) \frac{\delta S}{\delta \Phi^i} + R^C(\Phi) \frac{\delta S}{\delta C} \right) = 0 \tag{1.59}$$

Let us check explicitly that it condenses all the information we need.

**Symmetry** The order zero in $K$ of Eq. 1.59 gives:

$$\int R^i(\Phi) \frac{\delta S_{cl}}{\delta \Phi^i} = 0 \quad \text{(1.60)}$$
that expresses exactly the invariance of the classical action under the local transformation of Eq. 1.39.

**Closure** The first order in the sources of Eq. 1.59 gives:

\[
\int R^\alpha(\Phi) \frac{\delta l}{\delta \Phi^\alpha} R^\beta = 0
\]

\[
\Rightarrow \int R^\alpha(\Phi) \frac{\delta l R^\beta}{\delta \Phi^\alpha} = 0 \quad \forall \beta
\]  

(1.61)

Now let us take \(\beta = i\):

\[
\int R^j(\Phi) \frac{\delta l}{\delta \Phi^{jcl}} R^i(\Phi) + \int R_C \frac{\delta l}{\delta C} R^i(\Phi) = 0
\]

(1.62)

We shall now set \(C = \theta \Lambda + \theta' \Sigma\) in order to come back to a relation over the classical fields. Recalling 1.39 and using the linearity of \(R\) in \(C\), the first term in Eq. 1.62 gives:

\[
\int \theta R^j_{cl}(\Phi, \Lambda) \frac{\delta l}{\delta \Phi^{jcl}} \theta R^i_{cl}(\Phi, \Sigma) + \int \theta' R^j_{cl}(\Phi, \Sigma) \frac{\delta l}{\delta \Phi^{jcl}} \theta R^i_{cl}(\Phi, \Lambda)
\]

(1.63)

or

\[
\theta \theta'[\delta \Lambda, \delta \Sigma] \Phi^i_{cl}
\]

(1.64)

Since \(R^i\) are linear in \(C\), the second term of Eq. 1.62 can be written as:

\[
R^i_{cl}(\Phi, R_C(\theta \Lambda + \theta' \Sigma)) = -\theta \theta'[\delta \Lambda, \delta \Sigma] \Phi^i_{cl}
\]

(1.65)

Thus, since the choice of \(\theta\) and \(\theta'\) is arbitrary, Eq. 1.62 directly gives the closure relation 1.41.

**Jacobi identity** If now, in Eq. 1.61, we choose \(R^\beta \rightarrow R_C\), we get:

\[
\int R_C(\delta l) R_C(\delta C)
\]

(1.66)

that, following the same steps as before, indeed gives the closure of the closure, i.e. the Jacobi identity of the algebra.

### 1.4.2 Canonical transformations

A transformation of the fields

\[
\Phi^\alpha'(\Phi, K), \quad K'_\alpha(\Phi, K)
\]

(1.67)

is said to be *canonical* if it preserves the antiparentheses, that is, for every functional \(X\) and \(Y\):

\[
(X', Y') = (X, Y)
\]

(1.68)

where the transformed functionals \(X'\) and \(Y'\) are defined as:

\[
X'(\Phi', K') = X(\Phi(\Phi', K'), K(\Phi', K'))
\]

\[
Y'(\Phi', K') = Y(\Phi(\Phi', K'), K(\Phi', K'))
\]
And exactly in the same way as Classical Mechanics, the canonical transformations are generated by a functional $\mathcal{F}(\Phi, K')$, such that:

$$\begin{align*}
\Phi^{\alpha'} &= \frac{\delta \mathcal{F}}{\delta K'_\alpha} \\
K_\alpha &= \frac{\delta \mathcal{F}}{\delta \Phi^\alpha}
\end{align*}$$

In particular, the identity transformation is generated by the functional:

$$\mathcal{F}(\Phi, K) = \int d^4 x \Phi^\alpha(x) K'_\alpha(x)$$

It is also very useful to write a generic generating functional as the sum of a functional $\Psi(\Phi)$ dependent only on the fields $\Phi^\alpha$ plus all the rest:

$$\mathcal{F}(\Phi, K') = \Psi(\Phi) + \int d^4 x U^{\alpha}(\Phi, K') K'_\alpha$$

### 1.4.3 Gauge fixing

Now we have all the necessary tools to gauge fix the theory; let us enlarge the field row of Eq. 1.42 so as to include antighost fields $\bar{C}^\mu$ - that are fields with fermionic statistics, just like $C^\mu$, but with opposite ghost number with respect to them - and auxiliary bosonic fields $B^\mu$:

$$\Phi^\alpha = (\phi_{\mu \nu}, C^\mu, \bar{C}^\mu, B^\mu)$$

adding consequently their conjugate sources $\bar{R}_\mu$ and $K^R_\mu$ in 1.43.

The solution to the master equation can be easily obtained starting from the minimal solution to the master equation of Eq. 1.54 - that we now call $S_{\text{min}}$ - which was obtained without $\bar{C}^\mu$ and $B^\mu$; it is given by:

$$S(\Phi, K) = S_{\text{min}}(\Phi, K) - \int B^\mu \bar{R}_\mu$$

In the Batalin-Vilkovisky formalism the gauge fixing is a field redefinition of the kind of Eq. 1.67, that is generated by a functional

$$\mathcal{F}(\Phi, K') = \int d^4 x \Phi^\alpha K'_\alpha + \Psi(\Phi)$$

where

$$\Psi(\Phi) = \int d^4 x \bar{C}^\mu \left( -\frac{\lambda}{2} B_\mu + \mathcal{G}_\mu(\Phi) \right)$$

that, due to its statistical nature, is called gauge fermion and $\mathcal{G}_\mu$ is the gauge fixing function. For instance, in the harmonic gauge (the gravitational analogue of the Lorentz gauge of the electromagnetism) $\mathcal{G}_\mu(\phi) = \partial^\gamma \phi_{\mu \gamma}$. The transformation of the field is:

$$\begin{align*}
\Phi'^\alpha &= \Phi^\alpha \\
K'_\alpha &= K_\alpha - \frac{\delta \Psi(\Phi)}{\delta \Phi^\alpha}
\end{align*}$$

And the new, gauge fixed action $S_\Psi(\Phi, K)$ is:

$$\begin{align*}
S_\Psi(\Phi, K) &= S(\Phi', K') = S(\Phi, K) + \int R^{\alpha}(\Phi) \frac{\delta \Psi(\Phi)}{\delta \Phi^\alpha} \\
&= S(\Phi, K) + \{S, \Psi\}
\end{align*}$$
that can also be written as
\[ S_\Psi (\Phi, K) = S_{cl}(\Phi_{cl}) + S_{gf}(\Phi) - \int R^\alpha K_\alpha \] (1.77)

to make the gauge fixing term evident.
The fields $B^\mu$ appear only as Lagrange multipliers: they can be integrated out or, equivalently, substituted by the solution of their own field equations. The following proposition holds:

**Proposition 1.1:** If the action $S_\Psi (\Phi, K)$ satisfies the master equation, it continues to satisfy it even after integrating the auxiliary fields $B^\mu$ out.

And, since the gauge fixing is just a particular canonical transformation:

**Proposition 1.2:** If the action $S(\Phi, K)$ satisfies the master equation, then every $S_\Psi = S + \langle S, \Psi(\Phi) \rangle$ satisfies the master equation.

Let us now consider a generic gauge invariant functional $Q(\Phi)$; its expectation value is
\[ \int [d\Phi] Q(\Phi) \exp \left( i S(\Phi, K) + i \int L^I(\Phi_{cl}) \right) \] (1.78)

where $L^I$ are other additional sources coupled to gauge invariant composite fields $O^I$, made with the classical fields. Under the canonical transformation
\[
\Phi'^\alpha = \Phi^\alpha + \theta R^{\alpha} \\
K'_\alpha = K_\alpha - \int \frac{\delta I^\beta}{\delta \Phi^\alpha} K_\beta \theta
\] (1.79)

with $\theta$ an anticommuting constant, the functional $Q(\Phi)$ and the action become
\[
Q(\Phi') = Q(\Phi) + \theta \int R^\alpha \frac{\delta Q(\Phi)}{\delta \Phi^\alpha} = Q(\Phi) + \theta \langle S, Q \rangle \\
S(\Phi', K') = S(\Phi, K) + \left( \theta \langle S, S \rangle + \int (S, (S, \Phi^\alpha)) K_\alpha \theta \right)
\]

that are exact expansions. If $S$ satisfies the master equation, only $Q(\Phi)$ is altered by the transformation 1.79; actually, Eq.1.79 is just a change of variables in the functional integral 1.78. In fact, the $K$ transformation in the second line does not affect the action because $S$ depends on $K$ only via the combination $-\int R^\alpha K_\alpha$ and its contribution vanishes:
\[
\int R^\alpha \frac{\delta I^\beta}{\delta \Phi^\alpha} K_\beta \theta = \int (S, (S, \Phi^\alpha)) K_\alpha \theta = 0 \] (1.80)

Therefore the extra contribution that we get from Eq. 1.78, after applying 1.79:
\[
\langle \int R^\alpha \frac{\delta Q}{\delta \Phi^\alpha} \rangle_0 \] (1.81)

(the subscript 0 recalls that the external sources coupled to the elementary fields are set to zero) has to be zero. It can be rewritten as:
\[
\langle (S, Q) \rangle_0 = 0 \] (1.82)
and it is called Ward identity. In particular, if we choose $\Omega(\Phi)$ to be the gauge fermion $\Psi(\Phi)$, we get the equivalence of the two functional measures:

$$\int [d\Phi] \exp\left( iS(\Phi, K) + i \int L_I(0^I) \right) = \int [d\Phi] \exp\left( iS(\Phi, K) + \int L_I(0^I) \right)$$

(1.83)

Therefore, more in general, the following two important theorems hold [1]:

**Theorem 1.1:** The correlation functions

$$\langle O^{I_1}(x_1) \cdots O^{I_n}(x_n) \rangle$$

of gauge invariant composite fields $O^{I_i}(x_i)$

- are invariant under the canonical transformation 1.73, with arbitrary $\Psi(\Phi)$;
- are gauge independent, i.e. their values are independent of the choice of the gauge function $G(\phi)$.

**Theorem 1.2:** The physical quantities are invariant under the most general canonical transformation.

### 1.5 Quantum Gravity: Gravity as a Gauge Theory

We want now to address an answer to the following issue: how can gravity be interpreted as a gauge theory? Or, equivalently, what is the local transformation that characterizes the theory of gravitation?

For instance, the photon free Lagrangian of QED is invariant under the transformation of the photon field

$$A_\mu \rightarrow A_\mu - \partial_\mu \Lambda$$

where $\Lambda$ is a scalar function; or, in QCD, the SU(3) local invariance makes the Lagrangian invariant under:

$$A_{\mu}^{a} \rightarrow A_{\mu}^{a} - D_{\mu} \Lambda^{a}$$

where here $D_{\mu}$ is the covariant derivative associated with SU(3).

If we look back at the classical gravitational action 1.18, we find that it is invariant under:

$$\phi_{\mu\nu} \rightarrow \phi_{\mu\nu} - \frac{1}{\chi} (D_\mu \xi_\nu + D_\nu \xi_\mu)$$

(1.84)

where $\xi_\mu(x)$ is a point dependent vector. The group that generates this symmetry is the one of the local translations:

$$x'^\mu = x^\mu + \xi_\mu(x)$$

(1.85)

In fact, under such a transformation, the metric tensor transforms as:

$$g'_{\mu\nu}(x') = g_{\mu\nu}(x) - g_{\alpha\nu} \partial_\mu \xi^\alpha - g_{\mu\alpha} \partial_\nu \xi^\alpha + O(\xi^2)$$

(1.86)

But we are interested in the difference $\delta g_{\mu\nu} = g'_{\mu\nu}(x) - g_{\mu\nu}(x)$ of the metric tensors computed at the same spacetime point; then, expanding $g'_{\mu\nu}(x')$ around $x$ as well, we get:

$$g'_{\mu\nu}(x) = g_{\mu\nu}(x) - \xi^\alpha \partial_\alpha g_{\mu\nu} - g_{\alpha\nu} \partial_\mu \xi^\alpha - g_{\mu\alpha} \partial_\nu \xi^\alpha + O(\xi^2)$$

(1.87)
that, using Eq. 1.37, can be written as a difference between the fields $\phi_{\mu \nu}$:

$$
\delta \phi_{\mu \nu} = -\xi^\alpha \partial_\alpha \phi_{\mu \nu} - \phi_{\mu \alpha} \partial_\nu \xi^\alpha - \phi_{\alpha \nu} \partial_\mu \xi^\alpha - \frac{1}{\chi} (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) + \mathcal{O}(\xi^2) 
$$

(1.88)

that is exactly 1.84. To make the notation more compact, we can define the operator $D^\alpha_{\mu \nu}$ such that:

$$
\delta \phi_{\mu \nu} = D^\alpha_{\mu \nu} \xi^\alpha 
$$

(1.89)

neglecting all the terms $\mathcal{O}(\xi^2)$.

The BRST transformations for gravity are:

$$
\begin{align*}
\delta_{\text{BRST}} \phi_{\mu \nu} &= \theta R_{\mu \nu} (\Phi) = \theta \chi^2 (D^\alpha_{\mu \nu} C_\alpha) \\
\delta_{\text{BRST}} C_\mu &= \theta R^C_\mu (\Phi) = -\theta \chi^2 \nabla^\gamma C_\mu \\
\delta_{\text{BRST}} \bar{C}_\mu &= \theta R_{\mu} (\Phi) = \theta B_\mu \\
\delta_{\text{BRST}} B_\mu &= 0
\end{align*}
$$

(1.90)

The first line comes directly from 1.89; the second line expresses the closure relation

$$
[\delta_\Lambda, \delta_\Sigma] \phi_{\mu \nu} = \chi^4 (\Sigma^\alpha (\partial_\alpha \Lambda^\beta) - \Lambda^\alpha (\partial_\alpha \Sigma^\beta)) \partial_\beta \phi_{\mu \nu} \\
\equiv \delta_{\Delta(\Lambda, \Sigma)} \phi_{\mu \nu}
$$

(1.91)

that corresponds to:

$$
\Delta^\mu (\Lambda, \Sigma) = \chi^2 (\Lambda^\gamma \nabla_\gamma \Sigma^\mu - \Sigma^\gamma \nabla_\gamma \Lambda^\mu)
$$

(1.92)

To gauge fix the theory, we choose the gauge fermion:

$$
\Psi (\Phi) = -\frac{2}{\chi^2} \int d^4 x \bar{C}_\alpha \left( -\frac{\xi}{2} B_\alpha + \chi \partial^\beta \phi_{\beta \alpha} \right)
$$

(1.93)

The gauge fixed action is:

$$
S_g = S_{\text{cl}} + S_{\text{gf}}
$$

(1.94)

where:

$$
S_{\text{gf}} = \frac{2}{\chi^2} \left[ d^4 x \left( -\frac{\xi}{2} B_\mu B^\mu + \chi B_\mu \partial_\lambda \phi^{\lambda \mu} - \chi^2 C^\mu \partial^\gamma D^\rho_{\mu \nu} C_\rho \right) \right]
$$

(1.95)

and with $S_{\text{cl}}$ defined in Eq. 1.18. But the field $B_\mu$ can be integrated out, substituting it with the solution of its field equation:

$$
B_\mu = \frac{\chi}{\xi^2} \partial^\gamma \phi_{\mu \nu}
$$

(1.96)

and 1.95 becomes:

$$
S_{\text{gf}} = \frac{2}{\chi^2} \left[ d^4 x \left( \frac{\chi}{2 \xi^2} \partial_\lambda \phi^{\lambda \alpha} \partial^\sigma \phi_{\sigma \alpha} - \chi^2 \bar{C}_\mu \partial^\gamma D^\rho_{\mu \nu} C_\rho \right) \right]
$$

(1.97)

In order to treat gravity perturbatively, we need the Feynman rules of the theory. The first feature that makes gravity peculiar is the presence of an infinite number of vertices. In fact, expanding both the curvature $R$ and $\sqrt{-g}$ in 1.18 using Eq. 1.37, terms with an arbitrary number of gravitational
fields are generated.

The graviton propagator can be built by extracting, from the gauge fixed action, the part quadratic in the graviton field:

$$S^{(2)}_g = \frac{1}{2} \int d^4x \, \phi^{\mu \nu} Q_{\mu \nu, \rho \sigma} \phi^{\rho \sigma}$$  \hspace{1cm} (1.98)

The operator $Q_{\mu \nu, \rho \sigma}$ has to be symmetric for exchange of the couples of indices $\mu \nu \leftrightarrow \rho \sigma$ and symmetric for exchanges between indices in the same couple $\mu \leftrightarrow \nu$, $\rho \leftrightarrow \sigma$. In momentum space, calling $k$ the momentum, the propagator $P_{\mu \nu, \rho \sigma}(k)$ is defined as:

$$Q_{\mu \nu, \alpha \beta}(k) P^{\alpha \beta, \rho \sigma}(k) = \frac{i}{2} \left( \delta_{\mu}^{\rho} \delta_{\nu}^{\sigma} + \delta_{\mu}^{\sigma} \delta_{\nu}^{\rho} \right)$$  \hspace{1cm} (1.99)

where on the right side there is the identity for symmetric tensors.

In principle, $Q_{\mu \nu, \rho \sigma}$ can be computed by expanding the curvature using our field definition 1.37. At this order we need:

$$S^{(2)}_g = \int d^4x \left( \sqrt{-g^{(1)}} R^{(1)} + \sqrt{-g^{(2)}} R^{(2)} \right)$$  \hspace{1cm} (1.100)

The expansions are enlisted in Appendix A, but here we do not report the result directly as it is, since it gives no particular physical insight. Indeed, it is much more useful to express $Q_{\mu \nu, \rho \sigma}$ using the Barnes-Rivers spin operators [17]: looking at the group decomposition 1.36, it is clear that we need four projectors on the states of definite spin. If we define the transverse and longitudinal projectors in momentum space as:

$$\theta_{\mu \nu} = \eta_{\mu \nu} - \frac{k_{\mu} k_{\nu}}{k^2}$$  \hspace{1cm} (1.101)

$$\omega_{\mu \nu} = \frac{k_{\mu} k_{\nu}}{k^2}$$  \hspace{1cm} (1.102)

the spin projectors are:

$$p^{(2)}_{\mu \nu, \rho \sigma} = \frac{1}{2} \left( \theta_{\mu \rho} \theta_{\nu \sigma} + \theta_{\mu \sigma} \theta_{\nu \rho} - \frac{1}{3} \theta_{\mu \rho} \theta_{\nu \sigma} \right)$$  \hspace{1cm} (1.103)

$$p^{(1)}_{\mu \nu, \rho \sigma} = \frac{1}{2} \left( \theta_{\mu \rho} \omega_{\nu \sigma} + \theta_{\mu \sigma} \omega_{\nu \rho} + \theta_{\nu \rho} \omega_{\mu \sigma} + \theta_{\nu \sigma} \omega_{\mu \rho} \right)$$  \hspace{1cm} (1.104)

$$p^{(0-s)}_{\mu \nu, \rho \sigma} = \frac{1}{3} \theta_{\mu \rho} \theta_{\nu \sigma}$$  \hspace{1cm} (1.105)

$$p^{(0-w)}_{\mu \nu, \rho \sigma} = \omega_{\mu \nu} \omega_{\rho \sigma}$$  \hspace{1cm} (1.106)

where the dependence on the momentum is understood.

They are orthonormal in the sense

$$p^{(i-a)}_{\alpha \beta, \rho \sigma} p^{(j-b)}_{\alpha \beta, \rho \sigma} = \delta_{ij} \delta_{ab} P^{(i-b)}_{\mu \nu, \rho \sigma}$$  \hspace{1cm} (1.107)

$$p^{(2)}_{\mu \nu, \rho \sigma} + p^{(1)}_{\mu \nu, \rho \sigma} + p^{(0-s)}_{\mu \nu, \rho \sigma} + p^{(0-w)}_{\mu \nu, \rho \sigma} = \frac{1}{2} \left( \eta_{\mu \rho} \eta_{\nu \sigma} + \eta_{\mu \sigma} \eta_{\nu \rho} \right)$$  \hspace{1cm} (1.108)

With these projectors, the quadratic part can be expressed as:

$$Q_{\mu \nu, \rho \sigma} = k^2 \left[ p^{(2)}_{\mu \nu, \rho \sigma} - 2 p^{(0-s)}_{\mu \nu, \rho \sigma} - \frac{1}{8} \left( p^{(1)}_{\mu \nu, \rho \sigma} + 2 p^{(0-w)}_{\mu \nu, \rho \sigma} \right) \right]$$  \hspace{1cm} (1.109)
whence the graviton propagator:

\[ P_{\mu\nu,\rho\sigma} = \frac{i}{k^2} \left[ P_{\mu\nu,\rho\sigma}^{(2)} - \frac{P_{\mu\nu,\rho\sigma}^{(0-s)}}{2} - \Xi \left( P_{\mu\nu,\rho\sigma}^{(1)} + \frac{P_{\mu\nu,\rho\sigma}^{(0-w)}}{2} \right) \right] \]  

(1.110)

Then the propagator falls off like \( \sim k^{-2} \). According to the power counting theorem (for instance, in [27]), if the propagator falls off like \( k^{-(2+2s)} \), then the respective fields must have the dimensions of an energy to the power of \( 1+s \), in order to make the study of divergences with power counting work properly. That is the reason of our field definition 1.37.

### 1.6 Divergences of Quantum Gravity

In 1974, when t’Hooft and Veltman presented their archetypal version of quantum gravity in [23], they computed the one loop divergences of the theory. They found that the counterterm Lagrangian is:

\[ \mathcal{L}_g^{\text{count}}(g_{\mu\nu}) = \frac{1}{16\pi^2} \left( \frac{1}{120} R^2 + \frac{7}{20} R^\mu{}^\nu{}^\rho{}^\sigma R_{\mu\nu\rho\sigma} \right) \]  

(1.111)

where \( \epsilon = 4 - D \) parametrizes the divergences in the dimensional regularization framework and \( D \) is the continued dimension.

But they also noted that the theory is one-loop finite, that means that the counterterms of Eq. 1.111 can be eliminated by an appropriate field redefinition. Let us consider a new field \( g'_{\mu\nu} \), that differs from the field \( g_{\mu\nu} \) in 1.111 by \( \delta g_{\mu\nu} \):

\[ g'_{\mu\nu} = g_{\mu\nu} + \delta g_{\mu\nu} \]  

(1.112)

Expanding the gravitational Lagrangian around the field \( g_{\mu\nu} \), we get:

\[ \mathcal{L}_g(g'_{\mu\nu}) = \mathcal{L}_g(g_{\mu\nu}) + \frac{\delta \mathcal{L}_g(g_{\mu\nu})}{\delta g_{\mu\nu}} \bigg|_{g_{\mu\nu}} \delta g_{\mu\nu} \]  

(1.113)

and, in order to cancel out the divergences, we must set:

\[ \frac{\delta \mathcal{L}_g(g_{\mu\nu})}{\delta g_{\mu\nu}} \bigg|_{g_{\mu\nu}} \delta g_{\mu\nu} \equiv \mathcal{L}_g^{\text{count}}(g_{\mu\nu}) \]  

(1.114)

Let us consider a more general version of 1.111:

\[ \mathcal{L}_g^{\text{count}}(g_{\mu\nu}) = A R^2 + B R^\mu{}^\nu{}^\alpha{}^\beta R_{\mu\nu}{}^\alpha{}^\beta \]  

(1.115)

We can set:

\[ \delta g_{\mu\nu} = a R_{\mu\nu} + b R g_{\mu\nu} \]  

(1.116)

which, substituted in 1.114, gives:

\[ a = -\frac{\chi^2}{2} A, \quad b = \frac{\chi^2}{2} \left( B + A \right) \]  

(1.117)

Hence, the theory can be made finite at one loop. It was then reasonable to inquire whether the theory is finite at a higher number of loops. In 1985, Goroff and Sagnotti [9] examined the two loops behavior of quantum gravity; they found a counterterm of the form:

\[ \mathcal{L}_g^{\text{count}, 2\text{loop}}(g_{\mu\nu}) \propto R_{\alpha\beta\gamma\delta} R^\gamma{}^\delta{}^\mu{}^\nu R_{\mu\nu}{}^\alpha{}^\beta \]  

(1.118)
that could not be reabsorbed by a field redefinition, making all the hopes of a finite quantum gravity fade away.

All the difficulties in renormalizing the theory come from the definition \ref{def:gravitational_field} of the gravitational field. In detail, let us now consider a generic diagram \( D \) built using some of graviton vertices and propagators. It can depend on some external momenta \( p_i \) and the corresponding integral \( I_D \) over the loop momenta \( k_i \) assumes the general form:

\[
I_D \sim \int \! d^4 k_1 \ldots d^4 k_L \prod_{i=1}^{V} \left( V^{(N_i)}_i([k_i, p_i]) \right) \prod_{j=1}^{1} \left( P_j(\{k_j, p_j\}) \right) \tag{1.119}
\]

up to a constant symmetry factor and where we have called:

- \( I \) the number of internal propagators;
- \( L \) the total number of loops (i.e. the momenta to integrate);
- \( V \) the number of vertices.

They satisfy the topological relation:

\[
L - I + V = 1 \tag{1.120}
\]

Moreover, \( V^{(N_i)}_i \) is a generic \( N_i \)-leg vertex and \( P_j \) is the propagator \ref{prop:graviton_propagator}, both in momentum space.

Let us call \( \Lambda \) the high energy scale: that is, each internal line carries a momentum \( k_i^2 \to \Lambda^2 \) in such a limit. Then, from Eq. \ref{prop:graviton_propagator}, it is clear that, in high energy regime, the graviton propagator falls off as:

\[
P_j \sim \frac{1}{\Lambda^2} \tag{1.121}
\]

Since all the vertices are born from the curvature that contains two derivatives, going to the momentum space they behave at most like:

\[
V^{(N_i)}_i \sim \Lambda^2 \tag{1.122}
\]

that is, we are considering the worst case, where the legs of the vertices that carry the momenta are internal to the diagram. Thus, the integrand in \ref{1.119} falls off as \( \Lambda^{2(V-I)} \) and the superficial degree of divergence is equal to:

\[
\omega_D = 4L - 2I + 2V = 2(L + 1) \tag{1.123}
\]

A diagram can diverge if

\[
\omega_D \geq 0 \tag{1.124}
\]

Now the nonrenormalizability of quantum gravity is also evident by power counting; in fact, \( \omega_D \) in Eq. \ref{1.123} increases with the number of loops, making the renormalization process generate more and more terms to correct the divergences at each order of the perturbative expansion.

A useful theorem, stated in [26], tells what kind of counterterms we should expect at each order:

**Theorem 1.3:** The leading \( L \)-loop divergences of the action \ref{action:quantum_gravity} in four dimensions are of the form:

\[
S_{L}^{\text{div}} = \frac{\lambda^{2(L-1)}}{\epsilon^L} \int \! d^4 x \sqrt{-g} A(x) \tag{1.125}
\]

where \( A(x) \) is a functional such that:
(i) it is a local scalar constructed from the metric tensors $g_{\mu\nu}(x)$;
(ii) it does not depend on $\chi$;
(iii) it is constructed from the product of $L - k + 1$ Riemann tensors and $2k$ covariant derivatives $D_\mu$, with $k$ integer and $0 \leq k \leq L$.

Proof. Let us consider a generic $L$-loop diagram, built only with the graviton propagators 1.110 and vertices that stem from the action 1.18. Expanding the action 1.18, using the definition of the field $\phi_{\mu\nu}$ 1.37, we see that an $n$-leg vertex carries $\chi^{n-2}$. Thus, the total powers of $\chi$ that an $L$-loop diagram carries is:

$$\sum_{i=1}^{V} (n_i - 2) = \sum_{i=1}^{V} n_i - 2V$$

(1.126)

Since a propagator is connected to two internal vertex legs, calling $E$ the number of the external, we also have the relation:

$$2I + E = \sum_{i=1}^{V} n_i$$

(1.127)

And 1.126 becomes:

$$\sum_{i=1}^{V} (n_i - 2) = 2L - 2 + E$$

(1.128)

Comparing this result to the exponent of $\chi$ in Eq. 1.125, we see that here we have an additional $E$. But the constants $\chi^{E}$ are reabsorbed when we express the diagrams through the metric tensor $g_{\mu\nu}$, by means of Eq.1.37. This proves that the $\chi$ factor in Eq. 1.125 is correct.

The divergences can only be Lorentz scalars and the ingredients to build them are only three: the metric tensor $g_{\mu\nu}$, the Riemann tensor $R_{\mu\nu\rho\sigma}$ and the covariant derivative $D_\mu$ that can only appear in pair. Moreover, the dimension of the scalar $A(x)$ has to be four, in order to make $S^\text{div}_{L}$ dimensionless. Finally, the only counterterms allowed are the ones expressed in the condition (iii).

For example, at one-loop, there are only three possible counterterms:

$$R^2, \ R^{\mu\nu} R_{\mu\nu}, \ R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}$$

(1.129)

A counterterm of the form $R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}$ cannot be reabsorbed by means of a field redefinition of the kind of Eq. 1.112. But in four dimensions the Gauss-Bonnet theorem states that:

$$\sqrt{-g} \left( R^2 - 4R^{\mu\nu} R_{\mu\nu} + R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} \right) = \text{total derivative}$$

(1.130)

allowing to reparametrize a divergence proportional to $R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}$ as a linear combination of divergences proportional to $R^2$ and $R^{\mu\nu} R_{\mu\nu}$: it is this crucial identity that makes quantum gravity one-loop finite in four dimensions. At two loops, more counterterms can be generated:

$$(D_\alpha R)(D_\beta R),$$

$$(D_\alpha R_{\mu\nu})(D^\alpha R^{\mu\nu}),$$

$$R R_{\mu\nu} R^{\mu\nu},$$

$$R_{\mu\rho} R_{\nu\sigma} R^{\mu\nu\rho\sigma},$$

$$R R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma},$$

$$R^{\mu\nu\rho\sigma} R^{\rho\tau\sigma\tau} R_{\mu\nu\rho\sigma},$$

$$R^{\mu\nu\rho\sigma} R^{\mu\tau\rho\sigma\tau} R^{\nu\tau\rho\sigma\tau}$$

(1.131)
All these terms can be reabsorbed by a field redefinition, but the two in the last line. Moreover, \( R^{\mu \nu \rho \sigma} \) and \( R^{\mu \tau \rho \sigma} \) are not independent and one of them can be expressed in terms of the other; hence all the two loop divergences are proportional to \( R^{\mu \nu \rho \sigma} \). Thus, it is impossible to get just a renormalizable theory with a trivial quantization of the classical action 1.18. In the next chapter, we will examine how slight modifications of this action can solve these problems partially or totally.
THE PROBLEM OF RENORMALIZABILITY

2.1 STELLE’S THEORY: HIGHER DERIVATIVE QUANTUM GRAVITY

We saw that quantum gravity is nonrenormalizable due to the presence of the coupling constant $\chi$ with negative energy dimension, that is born from the gravitational field definition $1.37$, which is strictly connected with the behavior of the propagator $1.110$. But if the propagator falls off faster for high values of the momentum (at least as $1/k^4$), then we can drop the coupling constant $\chi$ in the change of variables of $1.37$, so as to deal with a dimensionless field that makes the theory renormalizable by power counting.

In 1977 Stelle proposed $[19]$ to enlarge the gravitational action $1.18$ with higher derivative terms that provide the correct UV behavior. However, the new terms that we add have to satisfy some basic requirements, such as Lorentz covariance, and we must add terms that contain parts quadratic in the graviton fields: for instance, adding terms like $R^3$ and $R^4$ alone are useless for our purpose, since they produce terms with at least three or four graviton fields; instead, terms like $R^2$ or $R\Box R$ start quadratically in the graviton field and modify the original propagator $1.110$. But Stelle proposed a minimal solution: inspired by the one loop counterterms that appear in quantum gravity $1.111$, he added their functional form directly to the original action:

$$S_{\text{HD}} = -\frac{2}{\chi^2} \int d^4x \sqrt{-g} \left( R + \alpha \chi^2 R^2 + \beta \chi^2 R_{\mu\nu} R^{\mu\nu} \right)$$ (2.1)

We do not need to add a term proportional to $R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$, because, in four dimension and for spaces topologically equivalent to the Euclidean one, the Gauss-Bonnet theorem $1.130$ holds.

The gravitational field $\phi_{\mu\nu}$ is now defined by:

$$g_{\mu\nu} \equiv \eta_{\mu\nu} + \phi_{\mu\nu}$$ (2.2)

that is dimensionally correct a posteriori; in fact, let us choose the gauge fermion

$$\Psi = -\frac{2}{\chi^2} \int d^4x \, C^\alpha \omega \left( -\frac{\Box}{\mu^2} - \xi \partial_\alpha + \partial_\beta \phi_{\beta\alpha} \right)$$ (2.3)

where $\omega \left( -\frac{\Box}{\mu^2} - \mu$ is a constant with the dimensions of an energy - is a polynomial function that, in the momentum representation, grows at least as $\sim k^3$. It provides the gauge fixing action:

$$S_{\text{HD,gf}} = -\frac{2}{\chi^2} \int d^4x \, \partial_\lambda \phi^{\lambda\alpha} \omega \left( -\frac{\Box}{\mu^2} - \partial_\alpha \phi_{\alpha\beta} \right)$$

$$- \int d^4x \, C^\mu \omega \left( -\frac{\Box}{\mu^2} - \partial_\alpha \partial_\beta C_{\alpha\beta} \right)$$ (2.4)
and the quadratic operator to be inverted is

\[ Q_{\mu\nu,\rho\sigma}^{\text{HD}} = \frac{k^2}{\chi^2} \left[ (1 - \chi^2 \beta k^2) p^{(2)}_{\mu\nu,\rho\sigma} - 2 \left( 1 + 2 \chi^2 k^2 (3\alpha + \beta) \right) p^{(0-s)}_{\mu\nu,\rho\sigma} - \frac{\omega(k^2/\mu^2)}{\xi} \left( p^{(1)}_{\mu\nu,\rho\sigma} + 2p^{(0-w)}_{\mu\nu,\rho\sigma} \right) \right] \]

whence the propagator:

\[ p_{\mu\nu,\rho\sigma}^{\text{HD}} = i\chi^2 \frac{k^2}{\chi^2} \left[ \frac{p^{(2)}_{\mu\nu,\rho\sigma}}{1 - \beta \chi^2 k^2} - \frac{p^{(0-s)}_{\mu\nu,\rho\sigma}}{2 \left( 1 + 2 \chi^2 k^2 (3\alpha + \beta) \right)} - \frac{\xi}{\omega(k^2/\mu^2)} \left( p^{(1)}_{\mu\nu,\rho\sigma} + 2p^{(0-w)}_{\mu\nu,\rho\sigma} \right) \right] \]

It falls off like \( \sim 1/k^4 \) in the high energy regime, justifying our definition 2.2 of the field.

We also note that, not to have tachions (i.e. to have positive definite mass squared at the denominators), we have to introduce the constraints:

\[ \begin{align*}
\beta &> 0 \\
3\alpha + \beta &< 0
\end{align*} \tag{2.7} \]

And there are some dangerous choices of the parameters: for example, the limit \( \beta \to 0 \) cannot be taken, or there will be a part of the propagator 2.6 that falls off like \( \sim 1/k^2 \). Instead, the limit \( \alpha \to 0 \) is safe.

Power counting is now very simple: using the same notation as in Sec. 1.5 and noting that the vertices behave like \( \sim k^4 \) because of the newly added higher derivative terms, the superficial degree of divergence is:

\[ \omega_D = 4L - 4I + 4V = 4 \tag{2.8} \]

In contrast with Eq. 1.123, it does not increase with the number of loops; moreover, all the coupling constants \( \alpha, \beta \) and \(-2/\chi^2\) have positive or null dimension: in sum, the theory is renormalizable by power counting.

A more complete demonstration of renormalizability at each order of the perturbative series can be found in Ref. [19].

### 2.2 The Loss of Unitarity in Higher Derivative Models

Although Stelle’s theory of gravitation achieves renormalizability, it shares the problems of higher derivative theories [18]. From a classical point of view, higher derivatives introduce instabilities in the solution of the equations of motion. There are some classical theories where higher derivatives naturally occur, typically seen as corrections to a lower order derivative theory. An example is the Abraham-Lorentz model for a radiating charged particle: for null external forces the equation of motion reads:

\[ \ddot{v} - \tau \dot{v} = 0 \tag{2.9} \]

where \( \tau = \frac{2e^2}{3mc^2} \sim 10^{-23} \text{s} \), and has two possible solutions:

\[ \dot{v}(t) = \begin{cases} 
0 & \\
\frac{ae^{t/\tau}}{\tau} & \end{cases} \tag{2.10} \]
The same happens in its relativistic version, the Dirac equation:

\[ \dot{v}_\mu = \tau \left( \frac{\dot{v}_\mu + \dot{v}_\nu \dot{v}^\nu}{c^2} - v_\mu \right) \]  \hspace{1cm} (2.11)

where now, beside the obvious solution \( v_\mu = 0 \), we also have:

\[
\begin{cases}
    v_t(s) = \cosh \left( e^{\hat{\tau}} + a \right) \\
    v_x(s) = \sinh \left( e^{\hat{\tau}} + a \right)
\end{cases}
\]  \hspace{1cm} (2.12)

Both the second solution of 2.9 and 2.12 are signals of an acausal behavior. Instead, from a quantum perspective, higher derivative terms generate ghosts. In fact, let us rewrite the propagator 2.6 in the following way, in the Landau gauge \( \xi = 0 \):

\[
P_{HD}^{\mu\nu,\rho\sigma} = i\chi^2 \left[ P^{(2)}_{\mu\nu,\rho\sigma} - \frac{P^{(0-s)}_{\mu\nu,\rho\sigma}}{2k^2} + \frac{P^{(0-s)}_{\mu\nu,\rho\sigma}}{2(2k^2 + 1/(3\alpha + \beta)\chi^2)} - \frac{P^{(2)}_{\mu\nu,\rho\sigma}}{k^2 - 1/\beta\chi^2} \right]  \hspace{1cm} (2.13)
\]

According to Kallen-Lehmann spectral representation (for instance, in [15, Chapter VII]), we can read the particle content of the theory examining the poles in \( k^2 \) of the propagators. Then, we see that the first line of 2.13 corresponds to a massless spin-2 particle, i.e. the graviton itself, that was already present in quantum gravity 1.110 and it brings two degrees of freedom; the first term in the second line is associated with a spin-0 particle, with mass \( m_0^2 \equiv -1/(2(3\alpha + \beta)\chi^2) \) and one degree of freedom, and the last one with a spin-2 particle, whose mass is \( m_2^2 \equiv 1/\beta\chi^2 \) and has five degrees of freedom. However, the last term of the propagator has the wrong sign, that is the residue at \( m_2^2 \) is negative. But, since the residue is essentially a measure of the probability to create a particle (i.e. \( \langle 0 | \Phi(0) | n \rangle \) \( \chi^2 \), where \( \Phi(0) \) is the Heisenberg operator and \( |n\rangle \) the eigenstate of the particle to create [15, Chapter VII]), in these conditions unitarity cannot be achieved. This problem is not due to our particular choice of higher derivative terms. We could have added other terms with higher derivatives, such as \( R \Box R \), \( R \Box^2 R \), \( \ldots \), \( R \Box^{n-1} R \) (where \( R \) is any contraction of the Riemann tensor), to make the propagator fall off like \( \sim 1/(k^2 p_n(k^2)) \), where \( p_n(k^2) \) is a polynomial of degree \( n \). But applying the fundamental theorem of algebra, we could have written it as:

\[
\frac{1}{k^2(1 + p_n(k^2))} = \frac{c_0}{k^2} + \sum_{i=1}^{n} \frac{c_i}{k^2 - m_i^2}  \hspace{1cm} (2.14)
\]

and multiplying both the sides by \( k^2 \) and taking the limit \( k^2 \to \infty \), we would have got:

\[
0 = c_0 + \sum_{i=1}^{n} c_i  \hspace{1cm} (2.15)
\]

that is to say, at least one residue would always be negative. Therefore, even if higher derivative theories of gravitation are satisfactory from the point of view of renormalization, the price to be paid, the loss of unitarity, is too high and it does not allow to give them a direct physical interpretation.
2.3 Kuz'min’s Theory: Nonlocal Quantum Gravity

From the structure of the propagator of Stelle’s theory 2.6, it is clear that unitarity can be restored (at least at the tree level) if the terms that we add to the original action 1.18 do not give rise to new poles in the propagator. This can be achieved only if we abandon the locality of the Lagrangian, allowing for the presence of a transcendental function along with the usual local derivative terms. This approach was first presented by Kuz’min in a short article in 1989 [10]; it then appeared eight years later in an article by Tomboulis [24] and recently a new interest in nonlocal field theories has grown (for example, in [13], [14] or [5], [6], [20]).

The nonlocal action is defined by:

\[
S_{NL} = \int d^4 x \sqrt{-g} L_{NL} = -\frac{2}{\chi^2} \int d^4 x \sqrt{-g} \left( R + \alpha \chi^2 R_1 \left( -\frac{\Box \chi}{\mu^2} \right) R + \beta \chi^2 R \mu \nu h_2 \left( -\frac{\Box \chi}{\mu^2} \right) R_{\mu \nu} \right) \tag{2.16}
\]

where \( \mu \) is a constant with the dimension of a mass and specifies the energy scale. The functions \( h_1(z) \) and \( h_2(z) \) are introduced to get a natural generalization of Stelle’s action 2.1: they can be chosen to be analytical, transcendental functions, that is, they can be represented as infinite series of derivatives.

The gauge fixing action is the same as Eq. 2.4, but now \( \omega(z) \) is a function that, in the high energy regime, grows faster than both \( zh_1(z) \) and \( zh_2(z) \).

Then, the quadratic operator is:

\[
Q_{k_{\mu \nu}, \rho \sigma}^{NL} = \frac{k^2}{\chi^2} \left[ \left( 1 - \chi^2 \beta k^2 h_2(k^2/\mu^2) \right) P_{k_{\mu \nu}, \rho \sigma}^{(2)} - 2 \left( 1 + 2 \chi^2 k^2 (3 \alpha h_1(k^2/\mu^2) + \beta h_2(k^2/\mu^2)) \right) P_{k_{\mu \nu}, \rho \sigma}^{(-s)} - \frac{\omega(k^2/\mu^2)}{\xi} \left( P_{k_{\mu \nu}, \rho \sigma}^{(1)} + 2 P_{k_{\mu \nu}, \rho \sigma}^{(-w)} \right) \right] \tag{2.17}
\]

from which the propagator becomes

\[
p_{NL}^{k_{\mu \nu}, \rho \sigma} = \frac{i \chi^2}{k^2} \left[ \frac{P_{k_{\mu \nu}, \rho \sigma}^{(2)}}{1 - \beta \chi^2 k^2 h_2(k^2/\mu^2)} - \frac{P_{k_{\mu \nu}, \rho \sigma}^{(-s)}}{2 \left( 1 + 2 \chi^2 k^2 (3 \alpha h_1(k^2/\mu^2) + \beta h_2(k^2/\mu^2)) \right)} - \frac{\xi}{\omega(k^2/\mu^2)} \left( P_{k_{\mu \nu}, \rho \sigma}^{(1)} + \frac{P_{k_{\mu \nu}, \rho \sigma}^{(-w)}}{2} \right) \right] \tag{2.18}
\]

With a proper choice of the functions \( h_1(z) \) and \( h_2(z) \) we can get rid of the unphysical poles.

However, interpreting this theory as an extension of quantum gravity 1.18, we would also like to obtain general relativity as a limit of the nonlocal action 2.16 and, in the meantime, we would also like to find the positive features of higher derivatives theory - that is, the renormalizability - in the nonlocal theories as well.

Thus, the following sections will be devoted to the hypotheses that we must
formulate over the nonlocal functions $h_1(z)$ and $h_2(z)$ in order to make the theory consistent with the previously examined ones.

### 2.4 Some Properties of Entire Functions

Let us recall some properties of analytic functions, that we need in the following sections. We follow the exposition in [12].

An entire function:

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

(2.19)

has an infinite radius of convergence and can be classified according to its behavior at the point $z = \infty$:

- if $f(z)$ is not singular at the point $z = \infty$, then, as a consequence of Liouville’s theorem, it is a constant;
- if $z = \infty$ is a simple pole, $f(z)$ is a polynomial, i.e $a_k = 0 \forall k \geq k_0$;
- if $z = \infty$ is an essential singularity, $f(z)$ is said to be transcendental.

Considering the results of the previous section, we will focus only on the last kind of functions.

Let us call $r = |z|$. Given $r$, we can define the maximum modulus function as:

$$M(r) = \max_{|z|=r} |f(z)|$$

(2.20)

Clearly, the modulus of any transcendental function grows as $r \to \infty$:

$$\lim_{r \to +\infty} M(r) = +\infty$$

(2.21)

Thus, it is not particularly illuminating to classify such functions through the limit of their modulus at $z = \infty$. Indeed, it is useful to compare the growth of the modulus with an exponential; that is, let us suppose that, for some $\mu \in \mathbb{R}$, we have:

$$M(r) < e^{r^\mu}$$

(2.22)

we define the order $\rho$ of the function $f(z)$ as:

$$\rho \equiv \inf \mu$$

(2.23)

A practical way of computing the order is given by the following formula:

$$\rho = \lim_{r \to +\infty} \frac{\ln \ln M(r)}{\ln r}$$

(2.24)

We can also subdivide the set of functions of order $0 < \rho < \infty$; let us suppose that exists $K > 0$ such that

$$M(r) < e^{Kr^{\rho}}$$

(2.25)

then, we can define the type $\sigma$ of the entire function $f(z)$ as:

$$\sigma \equiv \inf K$$

(2.26)
and it can be computed through the formula:

\[ \sigma = \lim_{r \to +\infty} \frac{\ln M(r)}{r^p} \] (2.27)

For example, \( e^z \), \( \cos z \) and \( \sin z \) are all functions of order 1 and type 1, while \( e^{e^z} \) is of infinite order.

In the following sections, we will often have to study functions inside cone-like sectors of the complex plane. There is a useful result, a simple consequence of Liouville's theorem, that allows us to study the behavior of an entire function inside a conical region by simply knowing its behavior on the boundaries:

**Theorem 2.1:** Let \( \Gamma \) be the interior of an angle of \( \pi/\rho \) radians

\[ \begin{align*}
\Gamma &= \left\{ z \mid \alpha - \frac{\pi}{2\rho} < \arg z < \alpha + \frac{\pi}{2\rho} \right\} \\
\end{align*} \] (2.28)

with \( \alpha \in \mathbb{R} \) and \( \partial \Gamma \) its boundary and let \( f(z) \) be an entire function of order \( \rho_f < \rho \), such that:

\[ |f(z)| \leq C < \infty \quad \forall z \in \partial \Gamma \] (2.29)

Then:

\[ |f(z)| \leq C < \infty \quad \forall z \in \Gamma \] (2.30)

Two corollaries, that give further information about the order of the function, can be derived:

**Corollary 2.1:** Let us consider a family of rays emanating from the origin, such that the angle between every pair of adjacent rays does not exceed \( \pi/\rho \), where \( \rho \geq 1/2 \). Then, every entire function \( f(z) \), that is not constant, of order \( \rho_f < \rho \) is unbounded on at least one ray of the family.

**Corollary 2.2:** Let \( f(z) \neq \text{constant} \) be an entire function of order \( \rho_f < 1/2 \) or of order \( \rho_f = 1/2 \) and minimum type. Then \( f(z) \) is unbounded on every ray emanating from the origin.

### 2.5 Hypotheses for the Nonlocal Function

Let us now discuss the properties that we must provide the nonlocal functions with in order to achieve not only unitarity, but also a theory consistent with general relativity and renormalizability. We shall therefore analyze the functions appearing at the denominators of the nonlocal propagators 2.18:

\[
\begin{align*}
  f_1(k^2) &= 1 - \beta \chi^2 k^2 h_2(k^2/\mu^2) \\
  f_2(k^2) &= 1 + 2\chi^2 k^2 (3\alpha h_1(k^2/\mu^2) + \beta h_2(k^2/\mu^2)) 
\end{align*}
\]

However, throughout this work, we will uniform the discussion, concentrating only on their essential functional structure:

\[ f(z) = 1 - zh(z) \] (2.31)
Since $h(z)$ is an entire function, so is $f(z)$.
Clearly, the hypotheses that we formulate for $f(z)$ have to be satisfied by both $f_1(z)$ and $f_2(z)$, as well as the hypotheses for $h(z)$ have to be valid for $h_2(z)$ and $3\zeta h_1(z) + \beta h_2(z)$.
We require that:

(i) $f(z)$ has no zeroes in the complex plane $|z| < \infty$;
(ii) $f(z)$ is real and positive on the real axis;
(iii) $|h(z)|$ has the same asymptotic behavior along the real axis at $\pm \infty$;
(iv) $f(0) = 1$;
(v) there exists $0 < \Theta < \pi/2$, such that, for the complex values of $z$ in the conical region $\mathcal{C}$ defined by:

$$
\mathcal{C} = \{ z \mid -\Theta < \arg z < \Theta, \pi - \Theta < \arg z < \pi + \Theta \}
$$

there exist $\gamma \in \mathbb{N}$ and a real constant $C$ such that

$$
\lim_{z \to \infty} \frac{|h(z)|}{z^\gamma} = C
$$

(vi) along the real axis

$$
\lim_{|z| \to \infty} \frac{h(z) - q_\gamma(z)}{q_\gamma(z)} z^m = 0, \quad \forall \ m \in \mathbb{N}
$$

where $q_\gamma(z)$ is a real polynomial of degree $\gamma$.

The first hypothesis is fundamental in order not to generate new poles, in contrast with the higher derivative approach. From the theory of entire functions, we know that such a function can only be the exponential of another entire function $g(z)$:

$$f(z) \equiv e^{g(z)} \quad (2.35)$$

Condition (ii) is also necessary: the reality of $f(z)$ assures the reality of the nonlocal action 2.16 and the positiveness ensures that the nonlocal function does not change the sign of the residues, letting the unitarity condition be satisfied.

The hypothesis (iii) was not present in the original Kuz’min’s paper [10], where an asymmetric behavior of the function was allowed. Here, like in more recent papers (for instance, [14]), we prefer to adopt a symmetric approach. We also note that in Euclidean spacetime, if (v) holds, this hypothesis follows straightforwardly.

The hypotheses (iv) and (v) make the nonlocal theory consistent with the other previously examined. In fact, the condition (iv) states that, in the limit $k^2 \to 0$ the propagator 2.18 becomes the propagator 1.110: that is, in the infrared limit, we recover general relativity, consistently with our interpretation of classical gravity as an effective theory of a more complex one. Indeed, with the fifth hypothesis, in the ultraviolet regime, the nonlocal theory of gravitation resembles a higher derivative theory and therefore it is renormalizable. Hence, the true nonlocality emerges only in a middle energy range, as depicted in Fig.2.1.
We have imposed the convergence not only on the real axis, but also in the conical region that surrounds it. Finally, we have also added the hypothesis (vi) (that is not present in any paper, but has been already stated in [16]): it is a subtlety regarding the UV divergences and we will discuss it in Sec. 4.5.

2.6 AN EXPLICIT EXAMPLE OF NONLOCAL FUNCTION

Obviously many functions that satisfy the hypotheses (i)-(vi) can be chosen. We will now give a very general form of functions consistent with them: we can choose \( g(z) \) defined in Eq. 2.35 such that:

\[
g(z) \equiv \int_0^{p(z)} \frac{1 - \zeta(\omega)}{\omega} \, d\omega
\]  

(2.36)

\( \zeta(z) \) is an entire function and \( p(z) \) a real polynomial of degree \( \gamma + 1 \). The hypotheses (iii), (iv) and (v) can be stated as hypotheses over \( p(z) \) and \( \zeta(z) \) as follows:

1. \( p(0) = 0 \);
2. \( \zeta(z) \) is real and even on the real axis such that \( \zeta(0) = 1 \);
3. \( |\zeta(z)| \to 0 \) for \( |z| \to \infty \) in the conical region defined in Eq. 2.32.

The hypothesis according which \( \zeta(z) \) has to be even (not present in the original Kuz’min’s paper) is necessary in order to achieve the condition (iii). From corollaries 2.1 and 2.2, it is clear that, if we want to build a function unbounded only in some conical regions of complex plane and bounded, in particular, in the region surrounding the real axis, we have to look for functions of order \( \rho > 1/2 \). This also implies that the function \( f(z) \) of Eq. 2.31 is of infinite order.

Hence, we can choose, for example, \( \zeta(z) \) as a function of order 2:

\[
\zeta(z) \equiv e^{-z^2}
\]  

(2.37)

In order to study the convergence of \( f(z) \), consistently with theorem 2.1, let us divide the complex plane in four cone-like sectors \( S_j \), with a common vertex at the origin:

\[
S_j = \left\{ z \mid -\frac{\pi}{4} + \frac{j\pi}{2} < \arg z \leq \frac{\pi}{4} + \frac{j\pi}{2} \right\}, \quad j = 0, 1, 2, 3
\]  

(2.38)

Then, choosing \( 0 < \varepsilon < \pi/2 \), we define the closed angles \( \overline{S}_j \) as:

\[
\overline{S}_j = \left\{ z \mid -\frac{\pi}{4} + \frac{j\pi + \varepsilon}{2} \leq \arg z \leq \frac{\pi}{4} + \frac{j\pi - \varepsilon}{2} \right\}
\]  

(2.39)

and \( \overline{S}_j \subset S_j \forall j \).

Denoting \( \theta = \arg z \), \( \zeta(z) \) becomes:

\[
\zeta(z) = \exp\left(-|z|^2 e^{2i\theta}\right) = \exp(-|z|^2 \cos 2\theta) \exp(-i|z|^2 \sin 2\theta)
\]
whence its modulus:
\[ |\zeta(z)| = \exp(-|z|^2 \cos 2\theta) \]

On the boundaries of the even \( S_j \) sectors \( j = 0, 2 \), \( \cos 2\theta = \sin \epsilon \), hence:
\[ |\zeta(z)| = \exp(-|z|^2 \sin \epsilon) \quad |z| \to \infty \to 0 \]

Theorem 2.1 ensures the convergence inside the sectors as well. In the odd sectors \( j = 1, 3 \), we have \( \cos 2\theta < -\sin \epsilon \) and:
\[ |\zeta(z)| > \exp(|z|^2 \sin \epsilon) \quad |z| \to \infty \to \infty \]

The convergence is depicted in Fig. 2.2.

Figure 2.2: The growth of the function \( |\zeta(z)| \). The dashed lines denote the sectors \( S_j \), the solid ones denote the sectors \( \bar{S}_j \).

Hence, we could be tempted to choose the angle defining the cone region of Eq. 2.32 as \( \Theta = \pi/4 \), but we recall from our choice of \( g(z) \) in Eq. 2.36 that \( \zeta(z) \) is then integrated in its variable and the right convergence has to be ensured over all the integration domain. Then, the correct choice of the angle is the one derived by our previous computation, divided by the degree of the polynomial \( p(z) \):
\[ \Theta \equiv \frac{\pi}{4(\gamma + 1)} \quad (2.40) \]

We could as well have chosen a more general definition of \( \zeta(z) \) as:
\[ \zeta(z) \equiv \exp \left( - \sum_{k=1}^{N} a_k z^{2k} \right) \quad (2.41) \]

with \( a_k \in \mathbb{R} \) and with the only restriction \( a_N > 0 \), but our analysis would have not been different. In fact, in this case, the sectors defined in Eqs. 2.38, 2.39 change into:
\[ S_j = \left\{ z \mid \frac{(2j-1)\pi}{4N} < \arg z \leq \frac{(2j+1)\pi}{4N} \right\}, \quad j = 0, \ldots, 2N-1 \quad (2.42) \]
\[
\tilde{g}_j = \left\{ z \left| \frac{(2j-1)\pi}{4N} + \frac{\epsilon}{2N} < \arg z \leq \frac{(2j+1)\pi}{4N} - \frac{\epsilon}{2N} \right. \right\} \tag{2.43}
\]

But we always have the convergence in half of the complex plane:

\[
|\tilde{z}(z)| \rightarrow 0 \quad j \text{ even}
\]

\[
|\tilde{z}(z)| \rightarrow \infty \quad j \text{ odd}
\]

and the angle of Eq. 2.40 now gets smaller by a factor \( N \):

\[
\Theta \equiv \frac{\pi}{4N(\gamma + 1)} \tag{2.44}
\]

Inserting 2.37 into our definition 2.36 of \( g(z) \), we get:

\[
g(z) = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{2m \, m!} p(z)^{2m} \tag{2.45}
\]

Recalling that the **incomplete gamma function**, defined by:

\[
\Gamma(0, x) = \int_x^{\infty} t^{-1} e^{-t} \, dt \tag{2.46}
\]

has the following Taylor expansion:

\[
\Gamma(0, x) = -\gamma_E - \ln x - \sum_{m=1}^{\infty} \frac{(-z)^m}{m \, m!} \tag{2.47}
\]

where \( \gamma_E \approx 0.577 \) is the Euler-Mascheroni constant, we can rewrite 2.45 as:

\[
g(z) = \frac{1}{2} \left[ \Gamma(0, p^2(z)) + \gamma_E + \ln p^2(z) \right] \tag{2.48}
\]

whence:

\[
f(z) = e^{g(z)} = |p(z)| e^{\frac{1}{2}[\Gamma(0, p^2(z)) + \gamma_E]} \]

It is now easy to check that the hypotheses (v) and (vi) over \( f(z) \) are satisfied. In fact, if we define:

\[
f_0(z) \equiv |p(z)| e^{\frac{\gamma_E}{2}} \tag{2.49}
\]

and rewrite:

\[
f(z) = f_0(z) + (f(z) - f_0(z))
\]

and use:

\[
\Gamma(0, x) = e^{-x} \left( \frac{1}{x} - \frac{1}{x^2} + o \left( \frac{1}{x^3} \right) \right) \tag{2.50}
\]

we get, on the real axis:

\[
f(z) - f_0(z) = \left[ e^{-p^2(z)} \left( \frac{1}{2p^2(z)} - \frac{1}{2p^4(z)} + o \left( \frac{1}{p^6(z)} \right) \right) + \right.
\]

\[
\left. + o \left( e^{-p^2(z)} \right) \right] |p(z)| e^{\frac{\gamma_E}{2}} \quad |z| \rightarrow \infty, 0
\]

Hence:

\[
f(z) \quad |z| \rightarrow \infty \rightarrow |p(z)| e^{\frac{\gamma_E}{2}} \tag{2.51}
\]

Because of the exponential suppression, the condition (vi) is also satisfied:

\[
(f(z) - f_0(z))z^m \quad |z| \rightarrow \infty \rightarrow 0 \quad \forall m \tag{2.52}
\]
2.7 THE EXPONENTIAL FUNCTION

As pointed out in Sec. 2.5, the transcendental functions of the nonlocal gravitational action 2.16 have to satisfy very strict hypotheses, but simpler alternatives have also been proposed. For instance, in [5] or [20], the function \( f(z) \) defined in 2.31 is an exponential, that is a function of finite order, in contrast with our choice of an infinite order one. With such a choice, the higher derivative theory cannot be restored in the UV limit: in fact, the hypotheses (v) and (vi) cannot hold.

However, let us see the implications of choosing an exponential nonlocal theory from the point of view of renormalizability, starting from a simple scalar model, whose action is:

\[
S_{\varphi} = \int d^4x \left\{ \frac{1}{2} \varphi (-\Box - m^2) \varphi - V_{\text{int}}(\varphi) \right\} \quad (2.53)
\]

where \( V_{\text{int}}(\varphi) \) is the interaction potential, that we suppose to be local (for example, it could be the \( \varphi^4 \) interaction). Let us suppose that we can correct the kinetic term such that the new action is:

\[
S_{\varphi,NL} = \int d^4x \left\{ \frac{1}{2} \varphi A \left( -\frac{\Box}{\mu^2} \right) (-\Box - m^2) \varphi - V_{\text{int}}(\varphi) \right\} \quad (2.54)
\]

with:

\[
A \left( -\frac{\Box}{\mu^2} \right) \equiv e^{-\Box/\mu^2} \quad (2.55)
\]

and \( \mu \), as usual, denotes the energy scale. The propagator of this theory is:

\[
G(k) = \frac{i}{\mu^2/\mu^2 (k^2 - m^2)} \quad (2.56)
\]

Since all the vertices that come from \( V_{\text{int}}(\varphi) \) contribute only with derivatives (or powers of momenta, in momentum space), in the ultraviolet regime all the diagrams are convergent: no divergences appear and the theory is finite.

Unfortunately, gravity is a gauge theory and we cannot modify the action as done before, because kinetic and interaction terms are strictly related to each other by the gauge symmetries and a modification of one of them implies the changes in the others. In [20], a scalar toy model that resembles gravity is proposed:

\[
S = S_{\text{free}} + S_{\text{int}} = \\
= \int d^4x \frac{1}{2} \varphi A \left( -\frac{\Box}{\mu^2} \right) (-\Box) \varphi + \\
+ \frac{1}{M_P} \int d^4x \left[ \frac{1}{4} \varphi \partial_\mu \varphi \partial^\mu \varphi + \frac{1}{4} \varphi \Box \varphi A \left( -\frac{\Box}{\mu^2} \right) \varphi - \\
- \frac{1}{4} \varphi \partial_\mu \varphi A \left( -\frac{\Box}{\mu^2} \right) \partial^\mu \varphi \right] \quad (2.57)
\]

where \( M_P \) is the Planck mass. And it is clear that, while the propagators are exponentially suppressed, the vertices are exponentially enhanced.

Let us consider a general 1-loop diagram. With the same notation as Sec. 1.6, it gives a contribution

\[
I_D \sim \int d^4k_1 \cdots d^4k_L \prod_{i=1}^V \left( V_i^{(N_i)}([k_i, p_i]) \right) \prod_{j=1}^1 \left( P_j([k_j, p_j]) \right) \quad (2.58)
\]

and, expanding the vertices and the propagators, more possibilities unfold:
• if the integrand does not contain exponentials, it can generate a divergence: in this case, it is computed with the usual rules of local theories;

• the terms coming from the expansion of the vertices that grow slower than the propagators give a convergent contribution;

• vertices also originate terms that contain exponentials of both external and internal momenta: they generate divergences that depend on exponentials of external momenta; thus, examining higher orders, we get subdivergent diagrams that also give exponential contributions, but now in terms of internal momenta.

For example, let us consider the two diagrams:

![Diagrams](image)

where the dark dot denotes in (a) the one-loop renormalized three-leg vertex (which we call $\Gamma_{3,\text{loop}}$) and in (b) the one-loop renormalized two-point function (which we call $\Gamma_{2,\text{loop}}$). The diagram (a) gives a contribution:

$$I_a \sim \int d^4k \frac{V^{(3)}(p, k, -p - k)\Gamma_{3,\text{loop}}(-p, -k, p + k)}{k^2(p + k)^2 e^{k^2/\mu^2} e^{(p+k)^2/\mu^2}}$$

while the diagram (b):

$$I_b \sim \int d^4k \frac{V^{(3)}(p, k, -p - k)\Gamma^{(3)}(-p, -k, p + k)\Gamma_{2,\text{loop}}(k)}{k^4(p + k)^2 e^{2k^2/\mu^2} e^{(p+k)^4/\mu^4}}$$

(2.59)

(2.60)

However, since both $\Gamma_{2,\text{loop}}$ and $\Gamma_{3,\text{loop}}$ contain exponentials, that - as noted in [20] - can also grow faster than the propagator, there are contributions that exponentially diverge. In sum, it is very difficult to take control of such subdivergences at every order of the perturbative expansion.

Therefore, up until now, there are no nonlocal quantum gravity theories that are as valid as Kuz’min’s theory.
SYSTEMATICS OF
UNITARITY AND
CAUSALITY

3.1 UNITARITY AND CAUSALITY AS BASIC REQUIREMENTS

In the previous chapters we stressed the importance of unitary, alongside renormalizability, to get a complete theory of quantum gravity. Moreover, we would also like a quantum theory of gravity to respect the causality principle. We now see what unitarity and causality mean in the framework of quantum field theory and we start with the general approach presented in [7]. Moreover, in this chapter, we will focus on *local* gauge theories, coming back to nonlocal theories only at the end.

Let us suppose that the interaction term of the Lagrangian defining the scattering matrix $S$ depends on a parameter $g(x)$ whose support is a space time manifold $M$. We suppose that $g(x)$ is normalized such that it assumes values only between 0 and 1: 0 means that the coupling is switched off, 1 that the interaction is completely active and the intermediate values describe a partial interaction.

If we consider a quantum state $\varphi$, its time evolution is determined by the action of the scattering matrix $S$ on $|\varphi\rangle$; in particular, if we denote with $|\varphi(-\infty)\rangle$ the state taken when the interaction is completely switched off at $t = -\infty$, then, at a later time $t$ and with coupling $g$, it will be:

$$|\varphi(t)\rangle = S(g)|\varphi(-\infty)\rangle$$  \hspace{1cm} (3.1)

We will now see how the unitarity and causality conditions can be seen as restrictions over the $S$-matrix.

3.1.1 Unitarity condition

The unitarity condition means that the norms of the states must be conserved; therefore the norm of the state $|\varphi(g)\rangle$, computed at a generic $g$, has to be the same as the norm at $t = -\infty$, when there is no coupling:

$$\langle \varphi(t) | \varphi(t) \rangle \equiv \langle \varphi(-\infty) | \varphi(-\infty) \rangle$$  \hspace{1cm} (3.2)

and this implies the unitarity of the operator $S(g)$:

$$S(g)S^\dagger(g) = S^\dagger(g)S(g) = 1$$  \hspace{1cm} (3.3)

We recall here that $S^\dagger$ can be defined in two equivalent ways. The first one is a classical matrix-based definition:

$$\langle \alpha | S^\dagger | \beta \rangle = \langle \beta | S | \alpha \rangle^*$$  \hspace{1cm} (3.4)

but we will also need a second definition, that is:

$$\langle \alpha | S^\dagger(L, i) | \beta \rangle = \langle \alpha | S(L^\dagger, -i) | \beta \rangle$$  \hspace{1cm} (3.5)
where $S^\dagger$ is defined starting from $S$, but computed from $\mathcal{L}^\dagger$, with all the $i$ factors in vertices and in Feynman prescriptions replaced by $-i$. In particular, if the Lagrangian is its own conjugate, $\mathcal{L}^\dagger = \mathcal{L}$, then:

$$\langle \alpha | S^\dagger (\mathcal{L}, i) | \beta \rangle = \langle \alpha | S(\mathcal{L}, -i) | \beta \rangle$$

(3.6)

that is, the elements of $S^\dagger$ are those of $S$, provided the $i$ factors of vertices and Feynman prescriptions are replaced with $-i$.

It is convenient to state the condition 3.3 in terms of the T-matrix, by isolating the identity in the scattering matrix:

$$S = 1 + iT$$

(3.7)

and Eq. 3.3 now reads:

$$T - T^\dagger = iTT^\dagger$$

(3.8)

### 3.1.2 Bogoliubov’s causality condition

The causality condition states, roughly speaking, that, chosen a time $T$, everything that happens at a time $t < T$ cannot depend on what happens at times $t > T$. Then, in order to implement causality as a restriction over the form of the $S$-matrix, we can imagine to divide the support $M$ into two parts, $M_1$ that is in the past of $T$, i.e. is the subset of $M$ such that $t < T$, and $M_2$, the future of $T$, the subset of $M$ such that $t > T$:

We can as well split the coupling function $g$ itself into two parts:

$$g(x) = g_1(x) + g_2(x)$$

(3.9)

where $g_1(x)$ ($g_2(x)$) has support only in $M_1$ ($M_2$). Causality means that the evolution of a state until the time $T$ depends only on $g_1(x)$:

$$|\varphi(t)\rangle = S(g_1) |\varphi(-\infty)\rangle , \ t < T$$

(3.10)

In particular, it holds when the time reaches $T$:

$$|\varphi(T)\rangle = S(g_1) |\varphi(-\infty)\rangle , \ t = T$$

(3.11)

But, for time coordinates $t > T$, the evolution is determined by the whole function $g$:

$$|\varphi(t)\rangle = S(g) |\varphi(-\infty)\rangle = S(g_1 + g_2) |\varphi(-\infty)\rangle , \ t > T$$

(3.12)

Then, splitting the time evolution into two parts using 3.11:

$$|\varphi(t)\rangle = S(g_2) |\varphi(T)\rangle = S(g_2) (S(g_1) |\varphi(-\infty)\rangle) , \ t > T$$

(3.13)
And comparing Eqs. 3.12 and 3.13, we get:

\[ S(g) = S(g_1 + g_2) = S(g_2)S(g_1) \]  

that is the finite causality condition. However, later in this chapter, we will need a differential formulation.

Let us then define another coupling function that differs from the function \( g(x) \) defined above, in Eq. 3.9, only for times later than \( T \):

\[ g'(x) = g_1(x) + g_2(x) + \delta g(x) \equiv g_1(x) + g_2'(x) \]  

where \( \delta g(x) \) - and subsequently \( g_2'(x) \) - has support only in \( M_2 \), as depicted in Fig. 3.2.

We notice that the finite causality condition 3.14, combined with the unitarity of \( S \), gives the following important result:

\[ S(g')S^\dagger(g) = S(g_2')S^\dagger(g_1)S^\dagger(g_2) \]  

that can be stated in general as:

(Bogoliubov’s causality condition) If two coupling functions \( g(x) \) and \( g'(x) \) coincide with each other for times smaller than a certain \( T \), than the product \( S(g')S^\dagger(g) \) must not depend on the simultaneous variation of \( g(x) \) and \( g'(x) \) by the same value in the region \( t < T \).

This is an alternative definition of causality, that we will refer to as Bogoliubov’s causality condition, since it was originally presented in Bogoliubov’s textbook [7].

We can put this condition in the form of an equation. Let us firstly rewrite the S matrix for \( g'(x) \) as:

\[ S(g') = S(g) + \delta S(g) \]  

with:

\[ \delta S(g) = \int_{y^0 > T} \frac{\delta S}{\delta g(y)} \delta g(y) \, dy \]  

And now the product in 3.16 can be expanded as:

\[ S(g')S^\dagger(g) = S(g)S^\dagger(g) + \delta S(g)S^\dagger(g) = 1 + \int_{y^0 > T} \frac{\delta S}{\delta g(y)} \delta g(y) \, dy \]  

By Bogoliubov’s condition, this product must not depend on the variation of the function \( g(x) \) with \( x_0 < T \). Then, deriving by \( g(x) \) and using the arbitrariness of the variation \( \delta g(y) \), we get:

\[ \frac{\delta}{\delta g(x)} \left( \frac{\delta S(g)}{\delta g(y)} S^\dagger(g) \right) = 0 \]  

(3.20)
Throughout the derivation, we have supposed that $x^0 < T < y^0$; the other case $y^0 < T < x^0$ can be obtained by this by exchanging $x$ and $y$:

$$
\frac{\delta}{\delta g(y)} \left( \frac{\delta S(g)}{\delta g(x)} S^\dagger(g) \right) = 0
$$

(3.21)

And we can put Eqs. 3.20 and 3.21 into a single equation by using two $\theta$-functions:

$$
\frac{\delta^2 S(g)}{\delta g(x) \delta g(y)} S^\dagger(g) + \theta(y^0 - x^0) \frac{\delta S(g)}{\delta g(y)} \frac{\delta S^\dagger(g)}{\delta g(x)} + \theta(x^0 - y^0) \frac{\delta S(g)}{\delta g(x)} \frac{\delta S^\dagger(g)}{\delta g(y)} = 0
$$

(3.22)

that is Bogoliubov’s condition, translated into an equation.

### 3.2 Building up a New Formalism

Both Bogoliubov’s causality condition 3.22 and the unitarity condition 3.8 have been stated in terms of the $S$-matrix. But we will always deal with Feynman diagrams, working in the framework of perturbation theory and we would like to apply these conditions to sets of diagrams, order by order in the perturbative series and then, summing over all the corrections, come back to the general expression in terms of the $S$-matrix.

The link between the perturbation theory and the more general requirements of unitarity and causality can be obtained using an elegant formalism, introduced by Cutkosky but also present - applied to gauge theories - in [21] and [22], that will lead us to the cutting equations.

We consider, at first, a theory with only bosonic propagators $G(x - y)$. We suppose that they satisfy the *Kallen-Lehmann representation*, that is, in momentum space:

$$
G(k) = \int_{a > 0} dM^2 \frac{\rho(M^2)}{k^2 - M^2 + i\epsilon}
$$

(3.23)

where $a$ is a real, positive constant and $\rho(M^2)$ is a positive definite *spectral density function*.

In the coordinate representation, this is equivalent to say that the following decomposition holds:

$$
G(x - y) = \theta(x^0 - y^0) G^+(x - y) + \theta(y^0 - x^0) G^-(x - y)
$$

(3.24)

with:

$$
G^\pm(x - y) = \frac{1}{(2\pi)^3} \int_{a > 0} dM^2 \rho(M^2) \times
$$

$$
\times \int d^4k e^{-ik(x-y)} \theta(\pm k_0) \delta(k^2 - M^2)
$$

(3.25)

In the case of a bare bosonic propagator with bare mass $m$, the spectral density function reduces to:

$$
\rho(M^2) = \delta(M^2 - m^2)
$$

(3.26)
and the propagator 3.24 can be cast in the usual form:

\[ G(x - y) = \frac{i}{(2\pi)^4} \int d^4k \, e^{-ik(x-y)} \frac{1}{k^2 - m^2 + i\epsilon} \]  

(3.27)

The \(\theta\)-function in 3.25 can also be defined as:

\[ \theta(k_0) = -\frac{1}{2\pi i} \int_{-\infty}^{\infty} d\tau \frac{1}{\tau + i\epsilon} e^{-ik_0\tau} \]  

(3.28)

in the limit \(\epsilon \to 0\).

Equipped with such propagators, a diagram can be written as a function

\[ \mathcal{F}(x_1, x_2, \ldots, x_n) \]  

(3.29)

where \(x_1, x_2, \ldots, x_n\) are the space-time coordinates of the vertices. For instance, let us consider a theory with vertices contributing only with a constant factor \(g\). We are now not interested in the number of legs that a vertex can have, since we deal with amputated correlation functions, but for sure a three vertex contribution should have the form:

\[ \mathcal{F}(x_1, x_2, x_3) = (ig)^3 G(x_1 - x_2)G(x_2 - x_3)G(x_3 - x_1) \]  

(3.30)

that is represented by the Feynman diagram:

\[ \text{Figure 3.3} \]

In diagrams like this, it is not clear how the energy flows: actually, it can flow inward or outward the three vertices with the only restriction that in each vertex the conservation of energy is satisfied. In order to implement the causality condition, we would like to distinguish the diagrams according to the direction of the energy flow.

For example, if we know that \(x_2\) and \(x_3\) are in the future with respect to the other vertex (that is, \(x_2^0 < x_1^0, x_3^0\)), we automatically have a restriction over the \(\theta\)-functions appearing when we expand the propagators in Eq. 3.29: the \(\theta\)-functions for which \(x_2^0 < x_1^0\) and those for which \(x_3^0 < x_1^0\) are simply zero. In such a case, the energy is forced to flow toward the vertices 2 and 3 from the vertex 1 and of course we expect the energy to flow outward the vertices 2 and 3 and inward the vertex 1 through external legs, in order to preserve the conservation of the energy:

Thus, from a more general point of view, coming back to Eq. 3.29, we can divide the vertex variables into two distinct subsets \(P = \{x_i\}\) and \(F = \{x_j\}\) such that each \(x_j\) is in the future of every \(x_i\), i.e \(x_j^0 > x_i^0 \land i, j\). To distinguish the variables of the two subsets, we underline the variables belonging to \(F\).
Hence, the function of the previous example would be \( \mathcal{F}(x_1, x_2, x_3) \). From a diagrammatic point of view, we represent the underlined variables by circling the respective vertices; again, the above diagram is now represented as:

\[
\begin{align*}
&x_1 \\
&x_2 \\
&x_3
\end{align*}
\]

where we intend the flow of energy going from the uncircled vertices toward the circled ones.

Between two nearby circled (or uncircled) vertices the energy can flow in both directions. We could of course introduce more than two subsets of variables, introducing other intermediate levels, but typical situations involve scattering among some particles, for which there are an initial and a final well defined states; the uncircled vertices (the past vertices) are associated with the incoming particles, while the circled ones (the future vertices) are associated with the outgoing particles.

This new formalism is very easy to handle, since every correlation function of both underlined and not underlined variables \( \mathcal{F}(x_1, \ldots, x_k, \ldots, x_j, \ldots, x_n) \) is defined starting from the original one \( \mathcal{F}(x_1, \ldots, x_k, \ldots, x_j, \ldots, x_n) \), with no variable underlined, following these rules:

1. \( G(x_i - x_j) \) is unchanged if neither \( x_i \) nor \( x_j \) are underlined;
2. \( G(x_i - x_j) \) is replaced by \( G^+(x_i - x_j) \) if \( x_i \) is underlined but \( x_j \) not;
3. \( G(x_i - x_j) \) is replaced by \( G^-(x_i - x_j) \) if \( x_j \) is underlined but \( x_i \) not;
4. \( G(x_i - x_j) \) is replaced by \( G^*(x_i - x_j) \) if \( x_i \) and \( x_j \) are both underlined;
5. for any underlined \( x_i \) one factor \( i \) has to be replaced by \( -i \).

Coming back to the example, we would get:

\[
\mathcal{F}(x_1, x_2, x_3) = (i\hbar)^3 G^+(x_1 - x_2) G^*(x_2 - x_3) G^+(x_3 - x_1) \quad \text{(3.31)}
\]

From now on, we also employ the following convention: all the incoming particles (i.e. the initial state) will be understood to be on the left of the diagram, while the outgoing particles will be on the right:

The energy flow can be, for instance:

\[\text{Figure 3.4}\]
It is now clear that all the contributions to a single process can be obtained by summing over all the possible combinations of underlined and not underlined vertices. However most of them will be zero, since they cannot satisfy the conservation of energy. For instance, the two diagrams:

\[ \text{Diagram 1} \quad \text{Diagram 2} \]

give no contribution to the process: in the first diagram, the energy can only flow from the nearby vertices to the vertex \( A \) and in the second diagram, the energy can only flow out from the vertex \( B \). Diagrams that do not give a vanishing contribution are, for example:

\[ \text{Diagram 3} \quad \text{Diagram 4} \]

Thus, in general, a diagram can give a non-zero contribution only if the circled vertices form connected regions that contain at least an outgoing line, while uncircled ones also form connected regions but involving at least an incoming line.

In order to simplify the notation and the structure of the diagrams further, we introduce the \textit{cuts}: instead of circling the vertices, we collect them inside a shadowed region, literally cutting the diagram. For example, the diagram of Fig. 3.4 becomes:

\[ \text{Diagram 5} \]

and the energy now flows from the unshadowed region through the shadowed one:

\[ \text{Diagram 6} \]

The rules outlined above to determine whether a diagram gives zero contribution or not can be stated as follows: a diagram gives non-zero contribution only if the cuts determine connected regions such that each unshadowed region involves at least one left (incoming) leg and each shadowed region involves at least one right (outgoing) leg.
3.3 Cutting Rules

Before going on, let us summarize here all the cutting rules encountered so far:

\[
G(k) = \frac{i}{k^2 - m^2 + i\varepsilon}
\]

\[
G^*(k) = -\frac{i}{k^2 - m^2 - i\varepsilon}
\]

\[
G^+(k) = 2\pi\theta(k_0) \delta(k^2 - m^2)
\]

\[
G^-(k) = 2\pi\theta(-k_0) \delta(k^2 - m^2)
\]

Table 3.1: Cutting rules for bosonic propagators.

In addition, if a vertex is in the shadowed region, we must replace one factor \(i\) with \(-i\).

More complex propagators, such as the photon propagator, are just a generalization of these. For instance, the cut gluon propagator:

\[
D^{++}_{\mu\nu}(k) = -\eta_{\mu\nu}\delta_{ab}D^+[k]
\]

(3.32)

where \(D^+[k]\) is the massless cut boson propagator.

3.4 The Largest Time Equation and Its Consequences

Let us suppose that, among all the time components \(\{x_0^i\}\) appearing in Eq. 3.29, we can identify \(x_0^k\) as the largest one, i.e. \(x_0^k > x_0^i \forall i \neq k\). Then the largest time equation holds:

\[
F(x_1, \ldots, x_k, \ldots, x_j, \ldots, x_n) = -F(x_1, \ldots, x_k, \ldots, x_j, \ldots, x_n)
\]

(3.33)

that is, every function in which \(x_k\) is not underlined equals minus the same function, with \(x_k\) underlined. This is obvious: in fact, since \(x_0^k\) is the largest
time, underlining it or not does not make any change and the minus sign is
due to the point (5) of Sec. 3.2.
For instance, let us come back to three-vertex example examined in Sec. 3.2:
\[
\mathcal{F}(x_1, x_2, x_3) = (ig)^3 G(x_1 - x_2)G(x_2 - x_3)G(x_3 - x_1)
\]  
(3.34)
and let us suppose that \(x_1^0\) is the largest time component. Then:
\[
G(x_1 - x_2) \equiv G^+(x_1 - x_2), \quad G(x_3 - x_1) \equiv G^-(x_3 - x_1)
\]
\[
\Rightarrow \mathcal{F}(x_1, x_2, x_3) = (ig)^3 G^+(x_1 - x_2)G(x_2 - x_3)G^-(x_3 - x_1)
\]
and the function with only \(x_1\) underlined is:
\[
\mathcal{F}(x_1, x_2, x_3) = -(ig)^3 G^+(x_1 - x_2)G(x_2 - x_3)G^-(x_3 - x_1) = -\mathcal{F}(x_1, x_2, x_3)
\]
The equation 3.33 is not so important \textit{per se} (it is not possible in general to
tell which, among all the time components, is the largest one), however its
consequences are fundamental.
For example, we get:
\[
\sum_{\mathcal{U}} \mathcal{F}(x_1, \ldots, \underline{x_i}, \ldots, x_j, \ldots, x_n) = 0
\]  
(3.35)
where \(\mathcal{U}\) is the set of all the \(2^N\) possible underlinings.
The proof is an immediate consequence of Eq. 3.33: supposing again that \(x_k^0\)
is the largest time component, for each term
\[
\mathcal{F}(x_1, \ldots, x_k, \ldots, \underline{x_j}, \ldots, x_n)
\]
there is a term
\[
\mathcal{F}(x_1, \ldots, \underline{x_i}, \ldots, x_j, \ldots, x_n)
\]
that is the opposite of the previous one, making the terms cancel out in
pairs. Since Eq. 3.35 contains all types of underlinings, we are actually not
even interested in which point has the largest time component.
However, more frequently, we will deal with diagrams in momentum rep-
resentation. Now, underlining a momentum in the same manner of a vertex
has no particular meaning, thus we cannot use the same notation, but the
Fourier transform \(\hat{\mathcal{F}}\) of each term of 3.35 is well defined. Hence, we can
rewrite 3.35 as:
\[
\sum_{c \in \mathcal{C}} \hat{\mathcal{F}}_c(k_1, \ldots, k_m) = 0
\]  
(3.36)
where with \(\hat{\mathcal{F}}_c\) we have denoted the Fourier transform of the term of 3.35
corresponding to the cutting \(c\) and with \(\mathcal{C}\) the set of all possible cuttings.
Now, equipped with this formalism and the largest time equation, we can
state the unitarity condition 3.8 and Bogoliubov’s causality condition 3.22 in
terms of the cutting rules.
3.4.1 Unitarity

First, let us rewrite Eq. 3.36 isolating the Fourier transform of the function where no variables are underlined, which we denote by $\hat{F}$, and the one where all the variables are underlined, denoted by $\hat{F}$:

$$\hat{F}(k_1, \ldots, k_m) + \hat{F}(k_1, \ldots, k_m) = -\sum_{\xi} \hat{F}_c(k_1, \ldots, k_m)$$  (3.37)

where now $\xi$ denotes the sum over all the internal cuts. A diagrammatic representation of this equation is:

![Diagram](image)

Figure 3.7: Unitarity condition

where the *cut blob* denotes the sum over all the internal cuts.

Actually, Eq. 3.37 has the same structure as the unitarity condition 3.8. But, while Eq. 3.8 holds for all the diagrams contributing to a process, Eq. 3.37 holds only for a single diagram, even if it is taken with all the possible cuts. Thus, if Eq. 3.37 effectively implies unitarity, further properties must hold:

- the diagrams in the shadowed region must be those that occur in $S^\dagger$;
- the sum over the intermediate states must be projected onto the sum over the physical states.

Let us suppose that these two additional hypotheses hold. Then, starting from Eq. 3.7, we see that (neglecting the identity) the T-matrix can be obtained from the S-matrix by a multiplication of $-i$. Thus, multiplying Eq. 3.37 by $-i$ we get:

$$-i\hat{F}(k_1, \ldots, k_m) - i\hat{F}(k_1, \ldots, k_m) = i \sum_{\xi} \hat{F}_c(k_1, \ldots, k_m)$$  (3.38)

Let us examine the terms one by one; the first one:

$$-i\hat{F}(k_1, \ldots, k_m)$$  (3.39)

once integrated over internal momenta, gives directly a contribution to the T-matrix. Summing over all the diagrams contributing to the process, then:

$$\sum_{D} \prod_{\text{internal } k_j} d^4 k_j \left(-i\hat{F}(k_1, \ldots, k_m)\right) = T$$  (3.40)

For the second term:

$$-i\hat{F}(k_1, \ldots, k_m)$$  (3.41)

we notice that, from Eq. 3.6, only if the Lagrangian that defines the S-matrix is its own conjugate, we get:

$$\hat{F}(k_1, \ldots, k_m) = \hat{F}^\ast(k_1, \ldots, k_m)$$  (3.42)
Thus:
\[-i\vec{\tau}(k_1, \ldots, k_m) = -(-i\vec{\tau}(k_1, \ldots, k_m))^*\]  \hspace{1cm} (3.43)

and, comparing this to Eq. 3.40, summing over all the diagrams we make the identification:
\[
\sum_D \int \prod_{\text{internal } k_j} d^4k_j \left( -i\vec{\tau}(k_1, \ldots, k_m) \right) = -T^\dagger
\]  \hspace{1cm} (3.44)

Finally, the term on the right hand side of Eq. 3.38, that combines both terms belonging to the S-matrix and terms belonging to S^\dagger, should be somehow related to the product TT^\dagger.
\[
\sum_D \int \prod_{\text{internal } k_j} d^4k_j \left( i \sum_{\epsilon_i} \vec{\tau}_c(k_1, \ldots, k_m) \right) = iT^\dagger T
\]  \hspace{1cm} (3.45)

Here the second hypothesis (S and S^\dagger can be connected only by physical states) is crucial: the identification 3.45 is possible only if the cuttings of the right hand side of Fig. 3.7 reduce to the physical states only. We will discuss it in the next paragraph for the case of QCD.

### 3.4.2 Causality

In order to view causality in terms of this formalism, we start with Eq. 3.35. In addition, we suppose that, among all the x_i, we can choose two variables, x_a and x_b, such that there is a time order between them, say x_0^a < x_0^b. Then, calling \U_a the set of underlinings where x_a is not underlined, we can rewrite Eq. 3.35 as:
\[
\sum_{\U_a} \mathcal{F}(x_1, \ldots, x_a, \ldots, x_n) = 0
\]  \hspace{1cm} (3.46)

In fact, since x_0^a is never the largest time and always smaller that x_0^b, we do not need the terms with x_a underlined to have cancellations in pairs. But we could suppose that x_0^b < x_0^a as well, obtaining a relation analogous to Eq. 3.46, but without considering x_b underlined:
\[
\sum_{\U_b} \mathcal{F}(x_1, \ldots, x_n) = 0
\]  \hspace{1cm} (3.47)

We can combine Eq. 3.46 and Eq. 3.47 into a unique expression by using two \theta-functions:
\[
\theta(x_0^b - x_0^a) \sum_{\U_a} \mathcal{F}(x_1, \ldots, x_n) + \\
+ \theta(x_0^a - x_0^b) \sum_{\U_b} \mathcal{F}(x_1, \ldots, x_n) = 0
\]  \hspace{1cm} (3.48)

But we can improve the notation further: first, let us isolate the term with no variable underlined:
\[
\mathcal{F}(x_1, \ldots, x_n) = -\theta(x_0^b - x_0^a) \sum_{\U = \{U_a, U_b\}} \mathcal{F}(x_1, \ldots, x_n) - \\
- \theta(x_0^a - x_0^b) \sum_{\U = \{U_a, U_b\}} \mathcal{F}(x_1, \ldots, x_n)
\]  \hspace{1cm} (3.49)
where \( U - \{ U_a, 0 \} \) (\( U - \{ U_b, 0 \} \)) means all the underlinings, but the ones involving \( x_a \) (\( x_b \)) and the one with no underlinings. We can now proceed noting that, among all the terms, there are many for which neither \( x_a \) nor \( x_b \) is underlined; we then isolate them to obtain:

\[
\mathcal{F}(x_1, \ldots, x_n) = - \sum_{U - \{ U_a, U_b \}} \mathcal{F}(x_1, \ldots, x_n)
\]

\[
- \theta (x_0^a - x_0^b) \sum_{U_b - U_a} \mathcal{F}(x_1, \ldots, x_n) -
\]

\[
- \theta (x_0^b - x_0^a) \sum_{U_a - U_b} \mathcal{F}(x_1, \ldots, x_n)
\]

(3.50)

where \( U_b - U_a \) denotes the underlinings, where \( x_b \) is underlined and \( x_a \) is not - vice versa for \( U_a - U_b \). We can express it from a diagrammatic point of view as:

\[
-\theta(x_0^b - x_0^a)
\]

\[
-\theta(x_0^a - x_0^b)
\]

Figure 3.8: Bogoliubov causality condition in coordinate representation

If the hypotheses assumed for unitarity hold, then Eq. 3.50 is formally equivalent to Bogoliubov’s condition 3.22: in fact, the first term of Eq. 3.22 is the one for which no time ordering between \( x_a \) and \( x_b \) is necessary, i.e. \( x_a \) and \( x_b \) are both outside the cut, and corresponds to the term on the left hand side of Eq. 3.50 plus the first on the right hand side. The second and the third terms of Eq. 3.22 correspond (assuming the correspondence \( x_a \mapsto x, x_b \mapsto y \)) respectively to the second and the third term of the right hand side of Eq. 3.50.

Because of the presence of the \( \theta \)-functions in 3.50, Bogoliubov’s condition is not so simple to handle in momentum space. However, for completeness, we report here the way to deal with it. In particular, we need another ingredient: starting from the definition of \( \theta \)-function 3.28, we can define an auxiliary propagator:

\[
G_\theta(k) = \frac{(2\pi)^3 i}{k_0 + \Gamma k} \delta^3(k)
\]

that represents just the \( \theta \) in momentum space and it is clearly not covariant. Thus, the representation of Fig. 3.8 is now translated into:
As anticipated, we want to show that summing over cut intermediate states equals the sum over physical states for a local gauge theory, that is to say QCD. We will give a diagrammatic proof, starting from the rules stated in Sec. 3.3 and we begin with two particle intermediate states: the general case will follow straightforwardly by induction.

A fundamental ingredient will be the Ward identity demonstrated in [21], that is:

\[ o, \alpha_1 \sum_{i=1}^{n} o, \alpha_i = o, \alpha_n \]

Figure 3.9

Here \( o \) means that the leg has to be taken on shell and \( \alpha_i \) denotes the collection of indices (color, Lorentz, etc.) associated with the \( i \)-th leg. The wavy line denotes the gluon propagator, while the dashed line denotes the ghost propagator. The double line at the end of the gluon lines denotes the multiplication by \( -ip^\mu \), where \( p^\mu \) is the momentum flowing into the vertex the leg is attached to. We will also need the additional propagators:

\[
\begin{align*}
\eta_{\mu \nu}^{tr} & = -\eta_{\mu \nu} - \frac{k_\nu \bar{k}_\mu + k_\mu \bar{k}_\nu}{2(k \cdot \zeta)^2} \\
\kappa_\mu & = k_\mu - 2(k \cdot \zeta) \epsilon_\mu
\end{align*}
\]

and \( \zeta_\mu \) is the vector \((1, 0, 0, 0)\). Thus, \( \eta_{\mu \nu}^{tr} \) allows the propagation of only physical degrees of freedom.

On shell, we also have

\[ k^2 = 0, \quad \bar{k}_\mu k^\mu = -2(k \cdot \eta)^2 \quad (3.51) \]
And the following equality holds:

\[ \eta^{\mu
u} = -\eta_{\mu\nu} - (ik_\nu) \frac{-ik_\mu}{2(k \cdot \eta)^2} = \frac{i k_\nu}{2(k \cdot \eta)^2} = \eta_{\mu\nu} - k_\nu \eta_{\mu\nu} + \frac{k_\nu k_\mu}{2(k \cdot \eta)^2} \]

In fact:

\[ -\eta^{\mu\nu} = -\eta_{\mu\nu} - (ik_\nu) \frac{-ik_\mu}{2(k \cdot \eta)^2} = \frac{i k_\nu}{2(k \cdot \eta)^2} = -\eta_{\mu\nu} - k_\nu \eta_{\mu\nu} + \frac{k_\nu k_\mu}{2(k \cdot \eta)^2} \]

Applying the Ward identity expressed in Fig. 3.10 twice we get:

The last equality comes from the fact that:

\[ D^+(k) \frac{i \xi_\mu}{2(k \cdot \eta)^2} \times i k^\mu = D^+(k) \]

which is exactly a cut ghost propagator.

Considering a physical two particle intermediate state (i.e. where only physical degrees propagate), applying Fig. 3.10, we obtain
But the last two terms are zero: applying the Ward identity we would have a transverse propagator contracted with a momentum. Then:

Figure 3.13

Using again the result of Fig. 3.10 we get:
and the first and the second diagram on the right hand side are identical to the one of Fig. 3.11. Therefore:

![Diagram](image)

Figure 3.15

The meaning of this equation is the following: the sum over the cuts of physical two particle intermediate states is equal to the sum over all the cuts of two particle intermediate states, i.e. the gauge field states and the two ghosts states, that are correctly taken with a minus sign. The generalization to N-particle intermediate states can be easily obtained by induction, starting from the case here examined $N = 2$.

### 3.6 The Case of Non Local Theories

Let us summarize the hypotheses that we have used to derive the largest time equation, that is the basis to prove both unitarity and causality for a local field theory:

- we have assumed that the vertices contribute only with a constant factor $ig$;
- the propagators satisfy the Kallen-Lehmann representation 3.23.

But these two hypotheses are too restrictive. For sure, we can relax the first one, by introducing vertices that depend also on derivatives: in terms of Fourier transforms it simply means to multiply by additional momenta. However, the second hypothesis cannot be generalized any further: only if the propagators satisfy the Kallen-Lehmann representation 3.23, we can define cut propagators like the ones of Sec. 3.3. Thus, only theories that satisfy these hypotheses can be treated using the formalism examined throughout this chapter. And, in general, it is very hard to find propagators that, taken alone, without further restrictions, satisfy the decomposition 3.24.

In order to understand the issue better, let us first examine the local Yang-Mills theory, that is for sure unitary and we expect the largest time equation to hold. The gauge field propagator is:

$$D_{\mu \nu}^{ab}(x-y) = -\delta_{ab} \left( \eta_{\mu \nu} - (1 - \xi) \frac{\partial_\mu \partial_\nu}{\Box} \right) D(x-y)$$

(3.52)
where $D(x - y)$ is the scalar propagator for a massless particle. In momentum space, the propagator is:

$$D_{\mu\nu}^{ab}(k) = -\delta_{ab} \left( \eta_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right) D(k)$$  \hspace{1cm} (3.53)

We know that we can decompose:

$$D(x - y) = \theta(x^0 - y^0) D^+(x - y) + \theta(y^0 - x^0) D^-(x - y)$$  \hspace{1cm} (3.54)

but, when substituted into Eq. 3.52, because of the presence of the derivatives (that act on $\theta$-functions as well), some contact terms - i.e terms proportional to $\delta (x^0 - y^0)$ and its derivatives - can be produced (in particular, if the time derivatives are two or more), preventing 3.24 to be satisfied for $x^0 = y^0$. That is:

$$D_{\mu\nu}^{ab}(x - y) = \theta(x^0 - y^0) D_{\mu\nu}^{ab}(x - y) +$$
$$+ \theta(y^0 - x^0) D_{\mu\nu}^{-ab}(x - y), x^0 \neq y^0$$  \hspace{1cm} (3.55)

Moreover, even the Kallen-Lehmann representation 3.23 is not satisfied, due to the presence of the double $k = 0$ pole in the part proportional to $k_\mu k_\nu$ of the propagator 3.53.

However, choosing to work in the Feynman gauge, where $\xi = 1$, the derivatives are eliminated, the decomposition is valid for every time component $x^0$, the propagator is coherent with the Kallen-Lehmann representation and the largest time equation holds as well. But for gauge theories we do know that other restrictions are imposed through Ward identities: in particular, every physical result must be gauge independent. Thus, once verified - to say - the unitarity, that is a physical property, in Feynman gauge, we can then switch to any other gauge and unitarity will hold there too.

As far as non local theories are concerned, the situation is more involved. For instance, let us consider the propagator of a non local version of Yang-Mills theory [24]:

$$D_{\mu\nu}^{NLab}(x - y) = -\delta_{ab} \left( \eta_{\mu\nu} - (1 - \xi) \frac{\partial_\mu \partial_\nu}{\Box} \right) h(\Box)^{-1} D(x - y)$$  \hspace{1cm} (3.56)

where $h(z)$ is a transcendental function, or the non local gravity propagator 2.18.

But also going to momentum space:

$$D_{\mu\nu}^{NLab}(k) = -\delta_{ab} \left( \eta_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right) h(k^2)^{-1} D(k) \equiv$$
$$\equiv h(k^2)^{-1} D_{\mu\nu}^{NLab}(k)$$  \hspace{1cm} (3.57)

it is evident that, beside the double pole $k = 0$, the nonpolynomial function at the denominator strongly prevent the Kallen-Lehmann representation from being satisfied. Hence, in these conditions, choosing the Feynman gauge in 3.56 or the Landau gauge in 2.18 does not improve the situation.

In [24], it is proposed to use $D_{\mu\nu}^{ab}(k)$ itself as the propagator of the theory - so as to have the same propagator as in the local theory - by rewriting the propagator as:

$$D_{\mu\nu}^{ab}(k) = h(k^2)^{1/2} D_{\mu\nu}^{NLab}(k) h(k^2)^{1/2}$$
and then conveniently divide the nonlocal vertex by $\bar{\hbar}(k^2)^{1/2}$ for each leg, so as to transfer the nonpolynomial function to vertices only. However this cannot solve the problem. In fact, a more serious problem plagues nonlocal theories: due to the nonlocal nature of vertices, they are made of an infinite set of terms, making it impossible to single out a vertex with the largest time. In sum, the largest time equation $3.35$ does not hold and the cutting rules, that are the basis of the demonstration of unitarity and causality for local gauge theory, are useless in the framework of nonlocal theories. So far, there do not exist other demonstrations of unitarity and causality of a field theory as complete as the ones we have examined. Thus, even if nonlocal theories have been introduced with the hope to solve the lack of unitarity of higher derivative models (and they do that at tree level), we are still in need of a new demonstration, that does not rely on Cutkosky’s formalism, which can prove unitarity and causality, starting from the perturbation theory.
To complete the Feynman rules of Kuz’min’s nonlocal theory of gravity, we need to find out the structure of the vertices. As we have seen, in quantum gravity, terms with an arbitrary number of fields can be generated, allowing the existence of vertices with an arbitrary number of legs as well. The structure of the $N$-leg vertex can be read by the expectation value of $N$ gravitational fields:

$$\langle \phi_{\mu_1 \nu_1}(x_1) \phi_{\mu_2 \nu_2}(x_2) \ldots \phi_{\mu_N \nu_N}(x_N) \rangle = \frac{\int [d\phi] \exp(iS_{NL}(\phi)) \langle \phi_{\mu_1 \nu_1}(x_1) \phi_{\mu_2 \nu_2}(x_2) \ldots \phi_{\mu_N \nu_N}(x_N) \rangle}{\int [d\phi] \exp(iS_{NL}(\phi))} \tag{4.1}$$

Splitting, as usual, the Lagrangian into the free field part and the interacting part, stopping at the tree level of the perturbative expansion and then applying Wick’s theorem, we should expect the general following form:

$$\langle \phi_{\mu_1 \nu_1}(x_1) \phi_{\mu_2 \nu_2}(x_2) \ldots \phi_{\mu_N \nu_N}(x_N) \rangle = \int dx \prod_{i=1}^N (P_{NL\mu_i \nu_i \rho_i \sigma_i}(x_i - x)) V_{NL}^{(N)\rho_1 \sigma_1 \ldots \rho_N \sigma_N}(x_1, \ldots, x_N) \tag{4.2}$$

In order to get the Feynman rules for vertices, we amputate the external legs (i.e. we eliminate the propagators appearing in Eq. 4.2) and do not consider the integration over the internal variable $x$.

Considering the form of the nonlocal terms of the Lagrangian 2.16, a vertex that involves only gravitons can be written in a more useful way as:

$$V_{NL}^{(N)}(x; x_1, \ldots, x_N) = \left. \frac{\delta^n'' \left( \sqrt{-g} R \right)}{\delta \phi^{n''}} \right|_{\phi=0} \left. \frac{\delta^n \left( \phi^n \right)}{\delta \phi^{n}} \right|_{\phi=0} \delta_{x_1} \ldots \delta_{x_N} \tag{4.3}$$

with the constraint:

$$N = n' + n + n'' \tag{4.4}$$

and where the dependence on Lorentz indices is understood; moreover, we have set:

$$\frac{\delta^n}{\delta \phi^n} \equiv \left. \frac{\delta^n}{\delta \phi(x_{i1}) \ldots \delta \phi(x_{in})} \right|_{\phi=0} \tag{4.5}$$

and with $R$ we denote any contraction of the Riemann tensor.

Eq. 4.3 exhibits the different contributions from the Riemann tensors and the ones - here written inside $\psi^{(n)}$ - that come from the nonpolynomial
functions, that we generically call \( h(-\Box_c) \), understanding the energy scale factor \( \mu \). The specific expression of \( \nu^{(n)} \) can be directly derived from the Lagrangian \( 2.16 \):

\[
\nu^{(n)} = \frac{\delta^n}{\delta\phi^n} \sum_{r=0}^{\infty} a_r (-\Box_c)^r \bigg|_{\phi=0}
\]

(4.6)

where we have expanded \( h(-\Box_c) \).

However, as shown by Eq. 4.3, the structure of vertices is very difficult to handle: for instance, looking at the contributions in 4.6, derivatives act on everything on their right and each \( \Box_c \) also depends on the fields.

Then, let us first set:

\[
-\Box_c = -\Box + (-\Box_c + \Box) \equiv -\Box + J
\]

(4.7)

where now all the dependence on the fields is encoded in \( J \). Hence, we can write \( (-\Box_c)^r \) in Eq. 4.6 as a combination of powers of \( J \) and \( -\Box \):

\[
\nu^{(n)} = \frac{\delta^n}{\delta\phi^n} \sum_{r=0}^{\infty} a_r \sum_{\lambda} \mathcal{T}_{l,r}^\lambda \left((-\Box)^{r-l}, (J)^l\right) \bigg|_{\phi=0}
\]

(4.8)

Here \( \mathcal{T}_{l,r}^\lambda \) is the term with 1 \( J \)-terms with a given ordering \( \lambda \) in between \( r-l \) \( \Box \)s. In fact, in general, \( \Box \)s and d’Alambertians do not commute and their order has to be taken into account.

We can rewrite the sum in Eq. 4.8 as:

\[
\sum_{r=0}^{\infty} \sum_{l=0}^{r} \rightarrow \sum_{l=0}^{\infty} \sum_{r=0}^{\infty}
\]

(4.9)

Then:

\[
\nu^{(n)} = \frac{\delta^n}{\delta\phi^n} \sum_{l=0}^{\infty} \sum_{r=0}^{\infty} a_r \sum_{\lambda} \mathcal{T}_{l,r}^\lambda \left((-\Box)^{r-l}, (J)^l\right) \bigg|_{\phi=0}
\]

(4.10)

The functional derivatives act only on \( J \)-factors and, since \( J \) can contain an arbitrary number of gravitational fields, the \( n \) derivatives appearing in Eq. 4.6 can act all on a single \( J \) or be distributed among \( n \) (or some) of them. Thus, we can finally write Eq. 4.6 as:

\[
\nu^{(n)} = \sum_{l=0}^{n} \sum_{\sigma} \sum_{r=1}^{\infty} a_r S^\sigma_{r,l,n} \left((-\Box)^{r-l}, \left(\frac{\delta^b_j}{\delta\phi^b} \bigg|_{\phi=0}\right)^l\right)
\]

(4.11)

where now \( S^\sigma_{r,l,n} \) denotes the sum over all the possible ways of distributing \( r-l \) powers of \( -\Box \) among the \( l \) \( J \) terms, with a specific sequence of derivatives \( \sigma \).

To make things clearer, let us consider a general term appearing in \( S^\sigma_{r,l,n} \):

\[
(-\Box)^{p_1} B_{l_1} (-\Box)^{p_2} B_{l_2} \ldots (-\Box)^{p_{l-1}} B_{l_{l-1}}
\]

(4.12)

where we have used the shorthand notation:

\[
B_{l} \equiv \frac{\delta^b_j}{\delta\phi^b} \bigg|_{\phi=0}
\]

(4.13)

with the constraint:

\[
\sum_{i=1}^{l} b_i = n
\]

(4.14)
Thus, with $\sigma$, in Eq. 4.11, we denote the sequence $\{b_i\}$ of derivatives. We can also count the numbers of these terms using the following argument. Let us consider the $lI$ factors in 4.11 - for now we neglect the d’Alambertians:

$$J_1 J_2 \ldots J_l$$

(4.15)

We have to distribute $n$ derivatives among them. Hence, we first attach one derivative to each $J_i$ (since, at the end, we must set $\phi = 0$, if only one term in Eq. 4.15 is not derived, it gives a null contribution):

$$\frac{\delta J_1}{\delta \phi} \frac{\delta J_2}{\delta \phi} \ldots \frac{\delta J_l}{\delta \phi}$$

(4.16)

We are then left with $n - l$ derivatives to be placed: these can be attached to only one $J_i$ or distributed among them. Thus, the maximum number of derivatives that can be attached to a single $J_i$ is $n - l + 1$. Then:

$$b_i \in \{1, n - l + 1\}$$

(4.17)

Since each of these $n - l$ derivatives can be attached in $l$ ways to the terms in 4.15 and considering the permutations of the $l$ $J$s, the total number of the ordered $\sigma$ sequences is:

$$\Gamma = l^{(n-l)} l!$$

(4.18)

However, this does not specify the structure of a single term of Eq. 4.8 completely. In fact, chosen $\sigma$, we can distribute the $r - l$ d’Alambertians in all the possible ways in the $l + 1$ positions among the $B_i$s: the sum of all such terms is understood in $S^\rho_{r,l,n}$.

Their total number $\Lambda$ is equivalent to the number of permutations of $r - l$ identical objects (i.e. the d’Alambertians) with $l$ separators (i.e. the field dependent factors $J$):

$$\frac{(-\Box)(-\Box) \ldots (-\Box)(-\Box)}{r \text{ objects}}$$

Then:

$$\Lambda = \frac{r!}{l! (r - l)!} = \binom{r}{l}$$

(4.19)

It is convenient to express the vertex structure of Eq. 4.11 in momentum space. Hence, let us consider a $N$-leg vertex:

![N-point vertex](image)

Figure 4.1

where $q_i$ is the momentum entering the $i$-th leg (with $i = 1, \ldots, N$).

From Eq. 4.33, let us consider:

$$\nu^{(n)}(x; x_1, \ldots, x_n) \frac{\delta^n R}{\delta \phi^n} \bigg|_{\phi = 0}$$

(4.20)
Each derivative appearing in the vertex structure $v^{(n)}$ acts on everything on its right: that is, each d’Alambertian in the $k$-th position in Eq. 4.12 produces a squared momentum, that is the sum of all $j$ momenta, with $j = k, \ldots, l$, of the corresponding $B_k$ legs, and the momenta coming through the $\frac{\delta^n \phi}{\delta \phi'}$ factor. Thus, calling $q_{l+1}$ the sum of the momenta that stem from the $\frac{\delta^n \phi}{\delta \phi'}$ factor, in momentum space we have to make the replacement:

$$(-\Box)^{p_k} \rightarrow (Q_k^2)^{p_k} \quad (4.21)$$

with:

$$Q_k = \sum_{j=k}^{l+1} q_k$$ \quad (4.22)

Thus, for fixed $r$, in $v^{(n)}$ we have:

$$\prod_{i=1}^{l+1} (Q_i^2)^{p_i} \quad (4.23)$$

Recalling the structure 4.12, we also have all the possible $\{p_i\}_{r-1}$ combinations (that is, all the $p_i$s such that $\sum_{i=1}^{l+1} p_i = r - 1$) to contribute to a vertex. We can then directly define:

$$S_l(Q_1, Q_2, \ldots, Q_{l+1}) \equiv \sum_{r=1}^{\infty} a_r \sum_{\{p_i\}_{r-1}}^{1+1} \prod_{i=1}^{l+1} (Q_i^2)^{p_i} \quad (4.24)$$

Indeed, all the $B_k$s give rise to a factor $\Phi_{\sigma_l,n}^{(r)} \{\{Q\}\}$, that carries all the needed indices.

Finally, considering Eq. 4.24, the Fourier transform of Eq. 4.11 is:

$$v^{(n)} (\{Q_l\}) = \sum_{l=0}^{n} \sum_{\sigma} \sum_{r=1}^{\infty} a_r S_{\sigma l,n}^{(r)} (\{Q_l\}) =$$

$$= \sum_{l=0}^{n} \sum_{\sigma} S_l (\{Q_l\}) \Phi_{\sigma l,n}^{(r)} (\{Q_l\}) \quad (4.25)$$

which is a compact formula, but still not enough. It is not clear how the vertex behaves in the high momentum limit because of the presence of the function:

$$S_l(z_1, z_2, \ldots, z_{l+1}) = \sum_{r=1}^{\infty} a_r \sum_{\{p_i\}_{r-1}}^{1+1} \prod_{i=1}^{l+1} z_i^{p_i} \quad (4.26)$$

Since the coefficients $a_r$ appear, we could relate somehow $S_l$ to $h(z)$ or its derivatives. This will be our goal in the next paragraph, where we will see that is not only possible, indeed $S_l$ assumes a very simple form.

### 4.2 Properties of $S_l(z)$

A first way to rewrite Eq. 4.26 is:
\[ S_l(z_1, z_2, \ldots, z_{l+1}) = \sum_{i=1}^{l+1} \frac{h(z_i)}{\prod_{j \neq i} (z_i - z_j)} \]  
(4.27)

**Proof.** We demonstrate this formula by induction over \( l \). Therefore, let us start examining the base case \( l = 0 \):

\[ \{l = 0\} \] In this case, formula (4.27) trivially becomes:

\[ S_0(z) = h(z) \]
as well as (4.26):

\[ S_0(z) = \sum_{r=0}^{\infty} a_r z^r = h(z) \]

\[ \{l = 1 \Rightarrow l \} \] Let us now suppose that (4.27) is true for \( l - 1 \); then, it is also true for \( l \). Firstly, looking at (4.26), we notice that we can rewrite:

\[ \sum_{r=0}^{\infty} a_r \sum_{\{p_i\}} \prod_{i=1}^{l+1} z_i^{p_i} \]
in a more compact expression as:

\[ \sum_{\{p_i\}} a_{l+\sum p_i} \prod_{i=1}^{l+1} z_i^{p_i} \]  
(4.28)

where \( \{p_i\} \) now means all the possible \( p_i \) permutations with no restriction and we have expressed \( r \) using the constraint:

\[ r - 1 = \sum_{i=1}^{l+1} p_i \]  
(4.29)

Now, let us focus on the monomials in \( z_1 \) and \( z_2 \):

\[ \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} z_1^{p_1} z_2^{p_2} \]

We rewrite it as:

\[ \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \left(\frac{z_1}{z_2}\right)^{p_1} z_2^{p_2+p_1} \]

And setting \( p = p_2 + p_1 \), we get:

\[ \sum_{p_1=0}^{\infty} \sum_{p=p_1}^{\infty} \left(\frac{z_1}{z_2}\right)^{p_1} z_2^{p} \]  
(4.30)

But, as we have done for Eq. 4.8, we can rearrange the sum as:

\[ \sum_{p_1=0}^{\infty} \sum_{p=p_1}^{\infty} \sum_{p_1=0}^{\infty} \sum_{p=0}^{p_1} \]
Then, Eq. 4.30 becomes:
\[
\sum_{p=0}^{\infty} \sum_{p_1=0}^{p} \left( \frac{z_1}{z_2} \right)^{p_1} z_2^p
\]  
(4.31)
where the sum over \( p_1 \) is a truncated geometric series; summing then over \( p_1 \), we get:
\[
\sum_{p=0}^{\infty} 1 - \left( \frac{z_1}{z_2} \right)^{p+1} z_2^p = \sum_{p=0}^{\infty} \frac{z_2^{p+1} - z_1^{p+1}}{z_2 - z_1}
\]  
(4.32)
Inserting this result into Eq. 4.28 and considering that \( p_1 + p_2 = p \), we obtain:
\[
S_l = \sum_{(p_1)_{i \geq 3}} \sum_{p=0}^{\infty} a \sum_{l+p-1=0}^{l=p+1} \frac{z_2^{p+1} - z_1^{p+1}}{z_2 - z_1} \prod_{i=3}^{l+1} z_1^{p_i}
\]  
(4.33)
Let us now translate \( p \rightarrow p - 1 \) and add the null term corresponding to \( p = 0 \):
\[
S_l = \sum_{(p_1)_{i \geq 3}} \sum_{p=0}^{\infty} a \sum_{l+p-1=0}^{l=p+1} \frac{z_2^{p+1} - z_1^{p+1}}{z_2 - z_1} \prod_{i=3}^{l+1} z_1^{p_i}
\]  
(4.34)
Splitting the sum into two parts, we get the first, important result:
\[
S_l(z_1, z_2, \ldots, z_{l+1}) = \frac{1}{z_2 - z_1} \left( \sum_{(p_1)_{i \geq 3}} \sum_{p=0}^{\infty} a \sum_{l+p-1=0}^{l=p+1} \frac{z_2^{p+1} - z_1^{p+1}}{z_2 - z_1} \prod_{i=3}^{l+1} z_1^{p_i} \right)
\]  
(4.35)
And now that we have expressed \( S_l \) in terms of two \( S_{l-1} \) depending terms, we can use the inductive hypothesis to rewrite these terms using Eq. 4.27:
\[
S_l(z_1, z_2, \ldots, z_{l+1}) = \frac{1}{z_1 - z_2} \left( \sum_{l' \geq 3} \sum_{i=3}^{l+1} h(z_i) \prod_{j \neq l_2} (z_i - z_j) \right) + \frac{1}{z_1 - z_2} \sum_{l' \geq 3} \sum_{i=3}^{l+1} h(z_i) \prod_{j \neq l_1} (z_i - z_j)
\]  
(4.36)
We can isolate the first term in each sum to get:
\[
S_l(z_1, z_2, \ldots, z_{l+1}) = \frac{1}{z_1 - z_2} \prod_{j \neq l_2} (z_1 - z_j) + \frac{1}{z_2 - z_1} \prod_{j \neq l_1} (z_2 - z_j)
\]  
(4.37)
The first two terms can be rewritten as:
\[
\prod_{\substack{i \neq 1 \atop j \neq 1}} \frac{1}{(z_i - z_j)} (h(z_1) + h(z_2))
\]
(4.38)
and the last two sum up to give:
\[
\sum_{i=3}^{l+1} \frac{h(z_i)}{\prod_{\substack{j \neq i \atop j \neq 3}} (z_i - z_j)} \left( \frac{1}{z_i - z_1} - \frac{1}{z_i - z_2} \right) = \sum_{i=3}^{l+1} \frac{h(z_i)}{\prod_{\substack{j \neq i \atop j \neq 3}} (z_i - z_j)}
\]
(4.39)
Finally Eq. 4.37 becomes:
\[
\mathcal{S}_l(z_1, z_2, \ldots, z_{l+1}) = \sum_{i=1}^{l+1} \frac{h(z_i)}{\prod_{\substack{j \neq i \atop j \neq 1}} (z_i - z_j)}
\]
(4.40)
that is exactly Eq. 4.27.

From Eq. 4.26 it can seem that, if we take the limit of coinciding arguments, \( \mathcal{S}_l \) can diverge. But, if we look back at its original expression of Eq. 4.26, it is clear that it cannot be the case. Then, let us come back to Eq. 4.35, where we set \( z_1 \equiv z_2 + \Delta z_2 \):
\[
\mathcal{S}_l(z_2 + \Delta z_2, z_2, \ldots, z_{l+1}) = \mathcal{S}_{l-1}(z_2 + \Delta z_2, z_3, \ldots, z_{l+1}) - \mathcal{S}_{l-1}(z_2, z_3, \ldots, z_{l+1})
\]
(4.41)
Taking now the limit \( \Delta z_2 \to 0 \):
\[
\lim_{\Delta z_2 \to 0} \mathcal{S}_l(z_2 + \Delta z_2, z_2, \ldots, z_{l+1}) = \frac{\partial}{\partial z_2} \mathcal{S}_{l-1}(z_2, z_3, \ldots, z_{l+1})
\]
(4.42)
Iterating these steps other \( l-1 \) times, we obtain:
\[
\prod_{i=1}^{l} \lim_{\Delta z_i \to 0} \mathcal{S}_l(z_{i+1} + \Delta z_i, \ldots, z_{l+1} + \Delta z_l, z_{l+1}) = \frac{\partial^l}{\partial z_{l+1}^l} \mathcal{S}_0(z_{l+1}) = \frac{\partial^l}{\partial z_{l+1}^l} h(z_{l+1})
\]
(4.43)
but \( h(z) \) is an entire function and its derivative is simply the derivative of its series, thus we arrive at the following result:
\[
\lim_{z_1 \to z_{l+1}} \lim_{z_2 \to z_{l+1}} \ldots \lim_{z_l \to z_{l+1}} \mathcal{S}_l(z_1, z_2, \ldots, z_{l+1}) = \frac{\partial^l}{\partial z_{l+1}^l} h(z_{l+1}) = \sum_{r=l}^{\infty} a_r r! (r-1)! z_{l+1}^{r-1}
\]
(4.44)
Now we are ready to study the UV behavior of the vertices. Thus, let us consider a $N$-point vertex: $n$ legs are associated to the factor $v^{(n)}$, while $n'$ and $n''$ of them are originated - respectively - from the factors $R$ on the right and $\sqrt{-g}R$ on the left:

\[
\begin{align*}
\frac{\delta^{n'}R}{\delta\phi^{n'}} & \quad v^{(n)} \\
\frac{\delta^{n''}(\sqrt{-g}R)}{\delta\phi^{n''}} & \quad N\text{-point vertex}
\end{align*}
\]

Figure 4.2

However, while studying the divergence of a diagram, we will be interested only in the $m \leq N$ legs that enter inside the diagram, not in the external, amputated legs.

Looking at vertex structure in Eq. 4.3, it is clear that, in the high energy limit a vertex behaves like:

\[
v^{(N)}([Q]) \sim \frac{\delta^{n''}(\sqrt{-g}R)}{\delta\phi^{n''}} ([Q]) v^{(n)} ([Q]) \frac{\delta^{n'}R}{\delta\phi^{n'}} ([Q]) \bigg|_{\phi=0} \phi=0 \sim \Lambda \delta^R_{\text{int}}
\]  

(4.45)

The factors depending on Riemann tensors and their contractions can have a complicated form, but for sure each term will contribute with two derivatives. Thus, let us call $\delta^R_{\text{int}}$ the number of derivatives coming from these factors, restricting only to internal legs; then:

\[
\delta^R_{\text{int}} \leq 4
\]  

(4.46)

and the behavior of the two $R$-factors for high momenta is:

\[
\frac{\delta^{n''}(\sqrt{-g}R)}{\delta\phi^{n''}} ([Q]) \bigg|_{\phi=0} \times \frac{\delta^{n'}R}{\delta\phi^{n'}} ([Q]) \bigg|_{\phi=0} \sim \Lambda \delta^R_{\text{int}}
\]  

(4.47)

where we have introduced a high energy scale $\Lambda$.

Considering Eq. 4.25, we see that the high energy behavior of the each term appearing in $v^{(n)}$ is:

\[
v^{(n)} ([Q]) \sim S([Q^2]) \Phi_{l,n}^{\sigma} ([Q])
\]  

(4.48)

We recall that $\Phi_{l,n}^{\sigma}$ stems from the $l B_l$ factors in 4.12. Since in each of them, in the coordinate representation, two derivatives appear, then in momentum space their behavior is, at most:

\[
\Phi_{l,n}^{\sigma} ([Q]) \sim \Lambda^{2l}
\]  

(4.49)
This happens when all the $Q_i$ diverge, i.e:

$$|Q_i| \sim \Lambda \to \infty \forall i = 1, \ldots, n$$  \hspace{1cm} (4.50)

Considering the high energy behavior of $S_1(\{Q_i^2\})$, it is convenient to define, along with $\gamma$ of Eq. 2.33:

$$\gamma \equiv \lim_{|z| \to \infty} \lim_{\Im z \to 0} \left( \frac{\ln |h(z)|}{\ln |z|} \right)$$  \hspace{1cm} (4.51)

also:

$$\gamma_{(1)} \equiv \lim_{|z| \to \infty} \lim_{\Im z \to 0} \left( \frac{\ln |h^{(1)}(z)|}{\ln |z|} \right)$$  \hspace{1cm} (4.52)

and considering again the worst case of Eq. 4.50, when all $Q_i$ have the same UV behavior, we can use Eq. 4.44 to get:

$$S_1(\{Q_i^2\}) \sim h^{(1)}(\Lambda^2) \sim \Lambda^{2\gamma_{(1)}}$$  \hspace{1cm} (4.53)

Finally, collecting Eqs. 4.47, 4.49 and 4.53, we get the high energy behavior of the vertex:

$$\mathcal{V}^{(N)}(\{Q_l\}) \sim \Lambda^{\delta_{\text{int}} + 2\gamma_{(1)} + 2l}$$  \hspace{1cm} (4.54)

Hence we can define the superficial degree of divergence a $N$-point vertex $\omega_N$ as:

$$\omega_N = \delta_{\text{int}} + 2\gamma_{(1)} + 2l$$  \hspace{1cm} (4.55)

But this result can seem quite unsatisfying, because $\omega_N$ can strictly depend on the structure of a diagram. Then, following [7, Ch. VIII], we can define the maximum degree of divergence $\Omega_N$ of the $N$-point vertex, which is a characteristic of the theory, by considering all the legs of the vertex internal. Then, recalling 4.46, we have:

$$\Omega_N = 4 + 2\gamma_{(1)} + 2l$$  \hspace{1cm} (4.56)

### 4.4 UV Behavior of Diagrams

We want now to prove the super-renormalizability of Kuz’min’s nonlocal quantum gravity by power counting.

Then, let us now consider a generic diagram $D$; in the same notation as Sec. 1.6, it gives a contribution:

$$I_D \sim \int d^4k_1 \ldots d^4k_L \prod_{i=1}^{\mathcal{V}} \left( \mathcal{V}_{NL,i}^{(N)}([k_L, p_L]) \right) \prod_{j=1}^{\mathcal{P}} \left( P_{NL,j}(\{k_j, p_j\}) \right)$$  \hspace{1cm} (4.57)

We have to consider the ultraviolet behavior of $I_D$, that is, we suppose that each internal line carries a momentum $|k_i| \to \Lambda$ and then send $\Lambda \to \infty$. From Eq. 2.18, it is clear that, in the high energy regime, the graviton propagator behaves like:

$$P_{NL}(k) \sim \frac{1}{\Lambda^{4+2\gamma}}$$  \hspace{1cm} (4.58)
Thus, the superficial degree of divergence $\omega_D$ of the diagram is:

$$\omega_D = 4L + \sum_{i=1}^{V} \omega_{N_i} - (4 + 2\gamma)I \quad (4.59)$$

where $\omega_{N_i}$ denotes the divergence degree of the $i$-th vertex with $N_i$ legs. However, $\sum_{i=1}^{V} \omega_{N_i}$ strictly depends on the number of external legs and then on the particular structure of the diagram. Since we would like to get only an upper bound on $\omega_D$, we can consider the maximum degree of divergence of Eq. 4.56. Then:

$$\sum_{i=1}^{V} \omega_{N_i} = \sum_{i=1}^{V} \left( 2\gamma(1_i) + 2l_i \right) \leq \sum_{i=1}^{V} \Omega_{N_i} = \sum_{i=1}^{V} \left( 2\gamma(1_i) + 2l_i \right) + 4V =$$

$$= \sum_{i=1}^{V} \left( 2\gamma(1_i) + 2l_i \right) + 4(1 - L + 1) \quad (4.60)$$

And Eq. 4.59 gives:

$$\omega_D \leq 4L + \sum_{i=1}^{V} \left( 2\gamma(1_i) + 2l_i \right) + 4(1 - L + 1) - (4 + 2\gamma)I =$$

$$= \sum_{i=1}^{V} \left( 2\gamma(1_i) + 2l_i \right) - 2I \gamma + 4 \quad (4.61)$$

Expressing $I = L + V - 1$, we obtain:

$$\omega_D \leq 2 \sum_{i=1}^{V} \left( \gamma(1_i) + l_i - \gamma \right) - 2(L - 1)\gamma + 4 \quad (4.62)$$

A diagram can diverge if:

$$\omega_D \geq 0 \quad (4.63)$$

that means:

$$2 \sum_{i=1}^{V} \left( \gamma(1_i) + l_i - \gamma \right) - 2(L - 1)\gamma + 4 \geq 0 \quad (4.64)$$

This can be interpreted as a restriction on the number of loops that a diagram can have to generate divergences:

$$L \leq \frac{2 + \gamma + \sum_{i=1}^{V} \left( \gamma(1_i) + l_i - \gamma \right)}{\gamma} \quad (4.65)$$

So far, however, we have not fully used the property of polynomial growth of the non local function $h(z)$. In fact, if in the conical region specified by Eq. 2.32 $|h(z)| \rightarrow \infty$ as $|z|^\gamma$, then, in this region, its $l$-th derivative grows at most as $|h^{(l)}(z)| \rightarrow \infty$ as $|z|^{\gamma - 1}$; this implies that, for each $1_i$:

$$\gamma(1_i) + l_i - \gamma \leq 0 \quad (4.66)$$

Substituted in Eq. 4.65, it gives:
\[
L \leq \frac{2}{\gamma} + 1 \quad (4.67)
\]

Hence, finally, the number of loops a connected divergent diagram can have are summarized in following table:

<table>
<thead>
<tr>
<th>GROWTH DEGREE $\gamma$</th>
<th>MAXIMUM NUMBER OF LOOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = 1$</td>
<td>$L = 3$</td>
</tr>
<tr>
<td>$\gamma = 2$</td>
<td>$L = 2$</td>
</tr>
<tr>
<td>$\gamma \geq 3$</td>
<td>$L = 1$</td>
</tr>
</tbody>
</table>

Table 4.1: The maximum number of loops which a diagram can have to give rise to divergence in function of the growth degree $\gamma$ of the nonlocal function.

In particular, if we choose $h(z)$ such that in the conical region defined by Eq. 2.32 $|h(z)| \mid z \rightarrow \infty \mid |z|^3$ only one loop diagrams diverge and choosing higher powers of growth does not improve the situation.

4.5 STRUCTURE OF COUNTERTERMS

To complete the proof of super-renormalizability of nonlocal quantum gravity, some issues remain to be clarified, such as:

- does the renormalization procedure generate nonlocal divergences?
- does nonlocal quantum gravity correspond to a higher derivative theory in the high energy limit?
- how many parameters do we need in order to renormalize the theory?

4.5.1 Locality of counterterms

We now prove that, for a convenient choice of the polynomial limit of the nonlocal function $h(z)$, no nonlocal counterterms are generated. This is not a trivial property: in fact, if we consider the general structure of nonlocal vertices of Eqs. 4.3, 4.25, it is evident that, when we symmetrize the vertex, some of the terms depend on the nonlocal function $h(z)$, evaluated only on the external momenta.

Such a situation happens when all the legs associated with the right $\mathcal{R}$ factor in Eq. 4.3 are external:

\[
\frac{\delta^n \phi}{\delta \phi^n} \bigg|_{\phi=0} \left( q_m^\prime, q_n^\prime, p_j^\prime, p_k \right) v^{(in)} \left( q_n^\prime, p_j^\prime, p_k \right) \times
\]

\[
\times \frac{\delta^{n'} \mathcal{R}}{\delta \phi^{n'}} \bigg|_{\phi=0} (p_k)
\]

where we have called $\{p_k\}$ the external momenta associated with the $n'$ legs that come from the right $\mathcal{R}$ factor and $\{p_j^\prime\}$ the external momenta of $r < n$.
legs that stem from $v^{(n)}$; moreover, $\{q^{'}_{n}\}$ denotes the set of the internal momenta of the remaining $v^{(n)}$ legs and $\{q''_{n}\}$ the one of the internal momenta brought by the $\sqrt{-g} \mathcal{R}$ legs. The situation is depicted in Fig. 4.3, where the blob collects all the internal legs.

Figure 4.3

Considering Eq. 4.27, among all the terms appearing in $v^{(n)}$, there are some that behave as

$$v^{(n)} (q^{'}_{n}, p^{'}_{j}, p_{k}) \sim \frac{h(p^2)}{\prod_{i} (Q_{i}^2 - p^2)} \Phi_{l,n}^{q'} (q^{'}_{n}, p^{'}_{j}, p_{k})$$

(4.69)

where $p$ is a linear combination of both $\{p_{i}\}$ and $\{p^{'}_{j}\}$. When computing divergences, the function $h(p^2)$ is not integrated and a nonlocal divergence can be generated.

However, recalling that the factor $\Phi_{l,n}^{q'}$ is generated from the Fourier transform of the $\mathcal{B}_{i}$ factors in 4.13 and calling $m$ the number of the $\mathcal{B}_{i}$s that are internal to the diagram, in the high energy limit:

$$\Phi_{l,n}^{q'} \sim \Lambda^{2m}$$

(4.70)

with the constraint $m \leq l$.

From the structure of the vertices 4.27, we note that the denominator of Eq. 4.69 contains the product of all the terms $Q_{i}^2 - p^2$, with $Q_{i}$ defined in 4.22: among all the $Q_{i}$s there are the ones associated to the internal $\mathcal{B}_{i}$ factors, making the denominator of Eq. 4.69 behave as:

$$\frac{1}{\prod_{i} (Q_{i}^2 - p^2)} \sim \frac{1}{\Lambda^{2m}}$$

(4.71)

Therefore, in the ultraviolet regime, $\Phi_{l,n}^{q'}$ cancel the denominator of Eq. 4.69 and the vertices 4.68 grow at most as $\Lambda^2$, due to the presence of the $\sqrt{-g} \mathcal{R}$ factor in 4.68.

Let us then suppose that our diagram is made by two sets of vertices: $V_{1}$ vertices with the nonlocal function that acts on internal legs only while $V_{2}$
vertices are of the type of Eq. 4.68. We can correct the power counting of Eq. 4.60 inserting the new vertices as:

\[ \omega_D \leq 4L - 2I(\gamma + 2) + \sum_{i=1}^{V_1} \left( 2\gamma l_i + 2l_i + 4 \right) + 2V_2 = \]

\[ = 4 - 2V_2 - 2I\gamma + \sum_{i=1}^{V_1} 2\left( \gamma l_i + l_i \right) \leq \]

\[ \leq 4 - 2(L - 1)\gamma - 2V_2(1 + \gamma) \] (4.72)

And if we are dealing with a diagram with at least one vertex of the kind of Eq. 4.68 and with at least one loop, we obtain that for

\[ \gamma \geq 2 \] (4.73)

all the diagrams converge. Thus, \( \gamma \geq 2 \) is an essential requirement in order not to get nonlocal divergences.

4.5.2 Correspondence between nonlocal and higher derivative theories

The nonlocal function \( h(z) \) enters in the vertices only via \( \delta_1 \) of Eq. 4.27. Let us rewrite it as:

\[ \delta_1(\{Q_i^2\}) = \sum_{i=1}^{l+1} \frac{CQ_i^{2\gamma} + \xi(Q_i^2)}{\prod_{j \neq i} (Q_i^2 - Q_j^2)} \] (4.74)

The function \( \xi(z) \) is defined as:

\[ \xi(z) \equiv h(z) - Cz^\gamma \] (4.75)

where \( C \) is the constant defined in Eq. 2.33.

Because of the hypothesis (vi) stated in Sec.2.5, we have:

\[ \lim_{z_i \to \infty} \left( \frac{\xi(z_i)}{\prod_{j \neq i} (z_i - z_j)} P(\{z_k\}) \right) = 0 \] (4.76)

where \( P(\{z_k\}) \) is a polynomial function in the variables \( z_k \) that embeds all the other contributions to the vertex of Eq. 4.3. Eq. 4.76 tells us that, in the high energy limit, all the differences between nonlocal and higher derivative theories vanish: in other words, in such a limit, we can compute the divergences of the nonlocal theory using its polynomial limit.

4.5.3 The renormalized Lagrangian of nonlocal quantum gravity

In sum, if we choose \( \gamma \geq 3 \), Kuz'min’s nonlocal quantum gravity is super-renormalizable, divergences arise only from one loop diagrams, the counterterms are local and in the ultraviolet regime it resembles a higher derivative theory.

Therefore, only covariant counterterms, whose dimension is at most four, appear, that is to say:

\[ R, \ R^2, \ R_{\mu\nu}R^{\mu\nu}, \ \lambda \] (4.77)
We have not included the term $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ because of the Gauss-Bonnet theorem \ref{gauge-bonnet}. $\lambda$ is the cosmological constant: in our nonlocal, bare theory \ref{cosmological-constant} it was set equal to zero, thus it appears only to cancel the divergences. Hence only four parameters are renormalized: the cosmological constant $\lambda$, the gravitational constant $\chi$ and the coefficients of $R^2$ and $R_{\mu\nu}R^{\mu\nu}$, i.e. the first term of the expansions of $h_1(z)$ and $h_2(z)$ (say - respectively $a_0$ and $b_0$). The renormalized nonlocal Lagrangian is:

\[
\mathcal{L}_{\text{ren}}^{NL} = -\frac{2}{\chi R} R - 2\alpha R h_1 \left( -\frac{\Box c}{\mu^2} \right) R - 2\beta R_{\mu\nu} h_2 \left( -\frac{\Box c}{\mu^2} \right) R_{\mu\nu} - 2\alpha (a_0 - a_0) R^2 - 2\beta (b_0 - b_0) R_{\mu\nu}^{\mu\nu} R_{\mu\nu} + \lambda R
\]  

(4.78)


## Finite Nonlocal Gravity

### 5.1 In Search for a Finite Theory of Quantum Gravity

Having explored the issue of renormalizability, now we shall add a last requirement to the nonlocal gravitational theory, its finiteness. Here, with finiteness we mean that with a slight adjustment of Kuz’min’s theory of Eq. 2.16, it is possible to eliminate all its divergences. Such modifications have been studied only recently in [13] or [14], that will be our starting point for this chapter.

Let us modify the nonlocal action 2.16 by adding a potential $V$:

$$S_{NL, \text{ finite}} = \int d^4x \sqrt{-g} (\mathcal{L}_{NL} + V) \quad (5.1)$$

$V$ depends only on the Riemann tensors and their derivatives and has to fulfill some basic requirements:

- it is a Lorentz generally covariant scalar;
- it is at least cubic in Riemann tensor;
- it can contain at most $2\gamma + 4$ derivatives of the metric tensor.

In such a way, $V$ does not modify the behavior of the propagator and the correspondence between the nonlocal theory and a higher derivative theory of degree $2\gamma + 4$ continues to hold, since the new terms do not distort the ultraviolet behavior of the vertices.

We can write $V$ as:

$$V = \sum_{n=3}^{\gamma+2} \mathcal{O}_{2n}(g_{\mu\nu}) \quad (5.2)$$

where $\mathcal{O}_{2n}(g_{\mu\nu})$ stands for the sum of all the covariant terms with $2n$ derivatives of the metric tensors (which are at least three). Denoting schematically with $\Box_i r^j$ a generic contraction of $j$ Riemann tensors and $2i$ covariant derivatives, the $\mathcal{O}_{2n}$ operators can be classified as:

- $\mathcal{O}_6 = \{r^3\}$
- $\mathcal{O}_8 = \{r^4, \Box_c r^3\}$
- ...
- $\mathcal{O}_{2\gamma+4} = \{r^{\gamma+2}, \Box_c r^{\gamma+1}, \ldots, \Box^\gamma r^{-1} r^3\}$

We write the generic operator $\mathcal{O}_{2n+4}$ as

$$\mathcal{O}_{2n+4} = \sum_{j=3}^{n+2} \sum_{k} \eta_{n,k}^{k,j} \left(\Box_c^{n+2-j} r^j\right)_k \quad (5.3)$$
with $\eta^{i,j}_n$ dimensionful constants and $k$ denoting a particular contraction of Lorentz indices.

It will also be useful to consider the high energy limit of the nonlocal Lagrangian $\mathcal{L}_{NL}$, where it assumes the form of a higher derivative theory:

$$\mathcal{L}_{NL} \xrightarrow{k^2 \to \infty} \omega_{-1} R + \sum_k \omega^k_0 (\mathcal{R}^2)_k + \sum_k \omega^k_1 (\mathcal{R} \Box c \mathcal{R})_k + \ldots + \sum_k \omega^k_\gamma (\mathcal{R} \Box^{\gamma} \mathcal{R})_k$$  \hspace{1cm} (5.4)

It is also worth noting for the analysis of the following sections that $\forall i, j, k$ the dimensions of the constants $\omega^{i-j}_\gamma - n$ and $\eta^{i,j}_\gamma - n$ are such that:

$$\left[ \frac{\omega^{i-j}_\gamma - n}{\omega^k_\gamma} \right] = \left[ \frac{\eta^{i,j}_\gamma - n}{\omega^k_\gamma} \right] = 2n$$  \hspace{1cm} (5.5)

### 5.2 Influences of the New Terms on Renormalization

In order to get finiteness, the new operators written in $V$ do have to contribute to the renormalization. For the rest of the chapter we will assume $\gamma \geq 3$: the theory is super-renormalizable and only one-loop divergences survive. They can be computed starting from the integral:

$$\int d^4 k \prod_i V_i^{(n_i)}(p_i;k) \prod_j W_j^{(n_j)}(p_j;k) \prod_l P_l(p_l;k)$$  \hspace{1cm} (5.6)

We understand the Lorentz indices to simplify the formulas.

Here $V_i^{(n_i)}(p_i;k)$ denotes a $n_i$-leg vertex coming from the operators quadratic in Riemann tensor $\mathcal{R} \Box^n \mathcal{R}$ already contained in $\mathcal{L}_{NL}$ and $W_j^{(n_j)}(p_j;k)$ a $n_j$-leg vertex that originates from the operators in $V$.

We recall that, in the high energy regime, the quadratic operator behaves as:

$$Q(k) \sim \sum_i a^i_\gamma \omega^i_\gamma k^{2\gamma + 4} + \sum_i a^i_{\gamma - 1} \omega^i_{\gamma - 1} k^{2\gamma + 2} + \sum_i a^i_{\gamma - 2} \omega^i_{\gamma - 2} k^{2\gamma} + \mathcal{O}(k^{2\gamma - 2})$$  \hspace{1cm} (5.7)

where $a^i_{\gamma - n}$ are some dimensionless constants; thus, the propagator falls off as:

$$P(k) \sim \frac{1}{\sum_i a^i_\gamma \omega^i_\gamma k^{2\gamma + 4}} \left( 1 + \sum_j \frac{a^j_{\gamma - 1}}{k^2} \sum_i \frac{\omega^i_{\gamma - 1}}{a^i_\gamma \omega^i_\gamma} + \sum_j \frac{a^j_{\gamma - 2}}{k^4} \sum_i \frac{\omega^i_{\gamma - 2}}{a^i_\gamma \omega^i_\gamma} + \mathcal{O} \left( \frac{1}{k^6} \right) \right)$$  \hspace{1cm} (5.8)

We will work in the framework of the dimensional regularization (the notation is listed in Appendix B): only logarithmic divergences survive, i.e. the ones corresponding to integrands that fall off like $1/k^3$. Then, we shall examine the structure of the vertices to find out which ones contribute.
5.2.1 Renormalization of the curvature and the cosmological constant

To the lowest order, the curvature and the cosmological constant both start linearly in the graviton field. Therefore we have to examine the tadpole diagram

![Tadpole Diagram](image)

Figure 5.1

then, by covariance, we can get the renormalization of the cosmological constant $\lambda_R$ and the Newton constant $\chi_R$. We have only two choices for the vertex:

$$\int d^4k V^{(3)}(p,k)p(k), \quad \int d^4k W^{(3)}(p,k)p(k)$$

(5.9)

Let us first consider $V^{(3)}(p,k)$, that comes from the terms $\mathcal{R}\Box^{-n}\mathcal{R}$. Their ultraviolet behavior is $k^2\gamma+4−2n$; however, we have also to pay attention to lower order terms, that can contribute to the divergence. Explicitly, let us come back to the structure of vertex originated by terms like:

$$\mathcal{R}\Box^{-n}\mathcal{R}$$

(5.10)

Of course, if the fields $\phi_2$ and $\phi_3$ are *internal*, the behavior is $k^2\gamma+4−2n$, but the symmetrization process also produces terms that grow as $k^2\gamma+2−2n$, if the internal fields are $\phi_1$ and $\phi_3$. Thus we get:

$$V^{(3)}(p,k) \sim \sum_i \alpha_i^1\omega_i^0 k^{2\gamma+4} +$$

$$+ \left( \sum_i \alpha_i^0(p) \omega_i^0 + \sum_i \alpha_i^1 \omega_i^1 \right) k^{2\gamma+2} +$$

$$+ \left( \sum_i \alpha_i^1(p) \omega_i^1 + \sum_i \alpha_i^2 \omega_i^2 \right) k^{2\gamma} +$$

$$+ O(k^{2\gamma-2})$$

(5.11)

where $\alpha_i^j$ are dimensionless constants and $\alpha_i^{-n}(p)$ are quadratic functions of the external momentum $p$. Due to the behavior of the propagator 5.8, we do not need other terms: the renormalization process is blind to the terms $\mathcal{R}\Box^{-3}\mathcal{R}, \mathcal{R}\Box^{-4}\mathcal{R}, \ldots, \mathcal{R}\Box\mathcal{R}, \mathcal{R}^2$.

Logarithmic divergences can only come from two sources: terms that contain combinations of

$$\frac{\omega_i^2}{\omega_i^0} \quad \text{or} \quad \frac{\omega_i^1}{\omega_i^0}$$

(5.12)

which are independent of the external momentum; they have dimension 4 and then they renormalize the cosmological constant; or terms

$$\frac{\omega_i^1}{\omega_i^0} f(p)$$

(5.13)
where \( f(p) \) is a quadratic function of the external momentum (actually, it is proportional to the Fourier transform of the curvature, taken at the lowest order): indeed they renormalize the curvature.

The vertex \( W^{(3)} \) comes from the operators cubic in Riemann tensor contained in the potential; to the lowest order, they are of the form

\[
(\Box \phi_1)(\Box \phi_2)(\Box^{\gamma-n} \phi_3)
\]

(5.14)

Since one of the fields has to be external, the highest power of internal momentum that can appear in \( W^{(3)} \) is \( 2\gamma + 2 \):

\[
W^{(3)}(p,k) \sim \sum_i \beta^{i,3}_\gamma(p)\eta^{i,3}_{\gamma-1}k^{2\gamma+2} + \sum_i \beta^{i,3}_{\gamma-1}(p)\eta^{i,3}_{\gamma-1}k^{2\gamma} + O(k^{2\gamma-2})
\]

(5.15)

where \( \beta^{i,3}_{\gamma-1}(p) \) are quadratic functions of the external momentum.

Divergences are now originated from terms that depend on:

\[
\sum_k c_k \omega_\gamma \beta^{i,3}_{\gamma-1}(p), \quad \sum_k c_k \omega_\gamma \beta^{i}_{\gamma-1}(p)
\]

(5.16)

They contribute to the renormalization of the curvature only.

5.2.2 Renormalization of the quadratic operators

The renormalization of the operators that are quadratic in Riemann tensor is more complex, but we will see that only the terms with the maximum number of derivatives (i.e. \( 2\gamma + 4 \)) contribute.

Let us consider the diagram:

![Figure 5.2](image)

that corresponds to

\[
\int d^4k V^{(4)}(p,k)P(k), \quad \int d^4k W^{(4)}(p,k)P(k)
\]

(5.17)

This diagram (and the following one) contributes also to the divergent part of the cosmological constant and the Newtonian constant: we will not consider them again here.

The vertex \( V^{(4)} \), that comes from terms \( \sim \phi^2(\Box \phi)(\Box^{\gamma-n} \phi) \), has the same behavior as the three-leg vertex in Eq. 5.11: the divergence of the operators \( \mathcal{R}^2 \) would require a dimensionless constant, but none of such terms can generate it. The vertices \( W^{(4)} \) come from the operator quartic in the graviton field in the potential

\[
(\Box \phi)(\Box \phi)(\Box \phi)(\Box^{\gamma-n} \phi)
\]

(5.18)
that is
\[ W^{(4)}(p, k) = \sum_i \beta^{i,A}_\gamma(p) \eta^{i,A}_\gamma k^{2\gamma} + \mathcal{O}(k^{2\gamma-2}) \quad (5.19) \]

where \( \beta^{i,A}_\gamma(p) \) are quartic in the external momentum. Such divergences depend on the dimensionless ratios:
\[ \frac{\eta^{i,A}_\gamma}{\sum_k c^k \omega^k} \quad (5.20) \]
times a quartic function of \( p \), equal to a linear combination of the Fourier transforms of the lowest orders of \( R^2 \) and \( R^2_{\mu\nu} \). Moreover, we have to include the following diagram:

![Figure 5.3](image)

whose divergent part can be computed starting from:
\[
\int d^4k V^{(3)}(p, k)V^{(3)}(p, k)P(p, k)P(k), \\
\int d^4k V^{(3)}(p, k)W^{(3)}(p, k)P(p, k)P(k), \\
\int d^4k W^{(3)}(p, k)W^{(3)}(p, k)P(p, k)P(k) \quad (5.21)
\]

In this case more combinations are allowed; considering the behavior of the vertices in Eqs. 5.11, 5.15, again the divergences can only come from the terms with \( 2\gamma + 4 \) derivatives, through nonlinear (rational) functions of the coefficient \( \omega^i_\gamma \) or through terms proportional to
\[ \eta^{i,3}_\gamma \eta^{i,3}_\gamma, \quad \eta^{i,3}_\gamma \omega^{i,3}_\gamma \quad (5.22) \]

All the results obtained in this section are summarized in Table 5.1 (the curly brackets denote any order of the factors).
Table 5.1: Renormalization of finite nonlocal gravity.

Then it is clear that, if we want to avoid the renormalization of the cosmological constant, the limit Lagrangian must not contain the terms $\mathcal{R} \Box^{-1} \mathcal{R}$ and $\mathcal{R} \Box^{-2} \mathcal{R}$; the finiteness of the curvature can be achieved only with a potential that is cubic in Riemann tensor, with at most $2\gamma + 2$ derivatives, while the finiteness of the quadratic in curvature terms can be obtained with quartic operators with $2\gamma + 4$ derivatives: this is the minimal choice to get a finite theory. Sometimes we will refer to such operators as killers.

5.3 An explicit example

Now we show an example where the considerations of the first two sections are applied. A similar calculation was considered in [13], using the Barvinsky-Vilkovisky method; instead, here we use the standard Feynman diagrams approach.

Let us consider the high energy limit of the nonlocal action with $\gamma = 3$:

\[
S_{NL, \text{limit}} = -\frac{2}{\chi^2} \int d^4x \sqrt{-g} \left( R + \alpha \chi^2 \frac{\Box^3}{\mu^6} R + \beta \chi^2 \mathcal{R}_{\mu\nu} \frac{\Box^3}{\mu^6} \mathcal{R}^{\mu\nu} + \frac{c_1}{\mu^8} \mathcal{R}^2 \Box c R^2 + \frac{c_2}{\mu^8} \mathcal{R}_{\mu\nu} \mathcal{R}^{\mu\nu} \Box c R_{\rho\sigma} R^{\rho\sigma} \right) + S_{gf}
\]

(5.23)

with quartic killers and the gauge fixing:

\[
S_{gf} = -\frac{2}{\chi^2} \int d^4x \left( \frac{1}{2\xi} \partial_\mu \phi \partial^\mu \omega \left( -\frac{\Box}{\mu^2} \right) \partial^\nu \phi_{\nu\rho} \right)
\]

(5.24)
where $\omega(z)$ falls off at least as $z^4$. Let us now list the Feynman rules that we need; the propagator of this theory is:

$$
P_{\mu\nu,\rho\sigma} = \frac{i\chi^2}{k^2} \left( \frac{p_{(2)}_{\mu\nu,\rho\sigma}}{1 + \beta \chi^2 \frac{k^8}{\mu^8}} - \frac{p_{(2)}_{\mu\nu,\rho\sigma}}{2 \left( 1 - 2\chi^2 \beta \frac{k^8}{\mu^8} - 6\alpha \chi^2 \frac{k^6}{\mu^6} \right)} \right)$$

In the renormalization process, the only diagram we need to consider is the one of Fig. 5.2; therefore we need the four-leg vertex in order to compute its divergent part. Some contributions come from the curvature and the two terms $R^2 \Box c R^2$: they are hard to compute, but not necessary to our purposes. Actually we are looking for a way to eliminate them, so we focus on the contributions that stem from the killer $R^2 \Box c R^2$: the two four-leg vertices are easy to compute and they come only from the lowest order of the Lagrangian terms:

$$
- \frac{2c_1}{\chi^2 \mu^8} \sqrt{-g} (R(1))^2 \Box c (R(1))^2
$$

$$
- \frac{2c_2}{\chi^2 \mu^8} \sqrt{-g} (R(1))^3 \Box c (R(1))^2 (R(1))^3
$$

with the notation of Appendix A.

We call $W_1$ the vertex proportional to $c_1$ and $W_2$ the one proportional to $c_2$. Going in momentum space (all the momenta being defined as flowing into the vertex):

![Figure 5.4](image)

the vertex $W_1$ is:

$$
(W_1)_{\mu_1 \nu_1 \mu_2 \nu_2}^{\mu_3 \nu_3 \mu_4 \nu_4}(p_1, p_2, p_3, p_4) = -\frac{2ic_1}{\mu^8 \chi^2} \times
\times (p_1 \mu_1 p_1 \nu_1 - p_1^2 \eta_{\mu_1 \nu_1})(p_2 \mu_2 p_2 \nu_2 - p_2^2 \eta_{\mu_2 \nu_2})
\times (-p_3^2 - p_4^2 - 2p_3 p_4) \times
\times (p_3^2 p_5^2 - p_3^2 \eta_{\mu_3 \nu_3})(p_4^2 p_5^2 - p_4^2 \eta_{\mu_4 \nu_4})
$$

and $W_2$ is:

$$
(W_2)_{\mu_1 \nu_1 \mu_2 \nu_2}^{\mu_3 \nu_3 \mu_4 \nu_4}(p_1, p_2, p_3, p_4) = \frac{ic_2}{8\mu^8 \chi^2} \times
\times (p_1 \mu_1 P_1 \nu_1 - p_1^2 \eta_{\mu_1 \nu_1})(p_1 \mu_1 \nu_1 - p_1^2 \eta_{\mu_1 \nu_1})(p_1 \nu_1 - p_1^2 \eta_{\nu_1 \nu_1}) \times
\times (p_2^2 p_2 \nu_2 - p_2^2 \eta_{\nu_2 \nu_2})(p_2^2 p_2 \nu_2 - p_2^2 \eta_{\nu_2 \nu_2})(p_2^2 \eta_{\nu_2 \nu_2}) \times
\times (-p_3^2 - p_4^2 - 2p_3 p_4) \times
\times (p_3^2 p_3^2 \eta_{\mu_3 \nu_3})(p_3^2 p_3^2 \eta_{\mu_3 \nu_3})(p_3^2 p_3^2 \eta_{\nu_3 \nu_3}) \times
\times (p_4^2 p_4^2 \eta_{\mu_4 \nu_4})(p_4^2 p_4^2 \eta_{\mu_4 \nu_4})(p_4^2 p_4^2 \eta_{\nu_4 \nu_4})
$$

(5.27)
that are to be symmetrized under exchanges of different couples \((\mu_i, \nu_i)\) as well as exchanges of \(\mu_i\) and \(\nu_i\) within a single couple (then conveniently divided by the symmetry factor). We call \(W_i^{\text{sym}}\) such vertices.

Let us come to the computation of the divergent part of the diagram the diagram 5.2. Calling \(p\) the external momentum and \(k\) the loop momentum:

\[
\begin{align*}
\Delta_{\mu\nu, \rho\sigma}(p) &= 12 \int d^4k P_{\alpha\beta, \gamma\delta}(k) \times \\
&\times \left[ (W_i^{\text{sym}})_{\mu\nu, \rho\sigma}(-p, p, -k, k) + (W_2^{\text{sym}})_{\mu\nu, \rho\sigma}(-p, p, -k, k) \right] \\
&\equiv -\frac{2}{\chi^2} \left( i\chi^2 \Delta_1 \mathcal{D}_{1\mu\nu\rho\sigma}(p) + i\chi^2 \Delta_2 \mathcal{D}_{2\mu\nu\rho\sigma}(p) \right)
\end{align*}
\]

We can say something more about the functional form of \(\Delta_{\mu\nu, \rho\sigma}(p)\): since we know that only the terms \(R^2\) and \(R_{\alpha\beta} R^{\alpha\beta}\) are renormalized, \(\Delta_{\mu\nu, \rho\sigma}\) is a linear combination of their Fourier transforms, both taken at the lowest order:

\[
\begin{align*}
R^{(1)}(-k)R^{(1)}(k) &\equiv \frac{1}{2} \phi_{\mu\nu}(-k) D_{1\mu\nu\rho\sigma}(k) \phi_{\rho\sigma}(k) \\
R^{(1)}(-k)R^{(1)}(k) &\equiv \frac{1}{2} \phi_{\mu\nu}(-k) D_{1\mu\nu\rho\sigma}(k) \phi_{\rho\sigma}(k)
\end{align*}
\]

Then:

\[
\Delta_{\mu\nu, \rho\sigma}(p) = -\frac{2}{\chi^2} \left( i\chi^2 \Delta_1 \mathcal{D}_{1\mu\nu\rho\sigma}(p) + i\chi^2 \Delta_2 \mathcal{D}_{2\mu\nu\rho\sigma}(p) \right)
\]

and \(\Delta_i\) are linear in \(c_i\).

Inserting the expressions 5.25, 5.27 and 5.28 into Eq. 5.29, with the renormalization conventions enlisted in Appendix B, using a Mathematica program, we get the following divergences:

\[
-\frac{3ic_1}{2\pi^2\chi^2\mu^2\varepsilon(3\alpha + \beta)} \left( p^2 \eta_{\mu\nu} \eta_{\rho\sigma} - p^2 p_{\mu} p_{\nu} \eta_{\rho\sigma} - p^2 p_{\rho} p_{\sigma} \eta_{\mu\nu} + p_{\mu} p_{\nu} p_{\rho} p_{\sigma} \right)
\]

\[
\text{Figure 5.5}
\]
and

\[ \frac{ic_2}{32\pi^2\chi^2\mu^2\epsilon(3\alpha + \beta)} \left[ -3(p^2)^2\eta_{\rho\phi}\eta_{\mu\nu} - \frac{7}{6}(p^2)^2\eta_{\sigma\nu}\eta_{\mu\rho} - \frac{10\alpha}{3\beta}(p^2)^2\eta_{\sigma\nu}\eta_{\mu\rho} - \frac{10\alpha}{3\beta}(p^2)^2\eta_{\sigma\mu}\eta_{\nu\rho} + \frac{7}{6}p^2\eta_{\nu\rho}p_\rho p_\mu + \frac{10\alpha}{3\beta}p^2\eta_{\nu\rho}p_\phi p_\mu + \frac{7}{6}p^2\eta_{\mu\rho}p_\rho p_\nu + \frac{10\alpha}{3\beta}p^2\eta_{\mu\rho}p_\sigma p_\nu + \frac{7}{6}p^2\eta_{\sigma\mu}p_\rho p_\nu + \frac{10\alpha}{3\beta}p^2\eta_{\sigma\mu}p_\phi p_\nu + 3p^2\eta_{\mu\nu}p_\rho p_\sigma + 3p^2\eta_{\phi\rho}p_\phi p_\nu + \frac{16}{3}p_\mu p_\nu p_\rho p_\sigma - \frac{20\alpha}{3\beta}p_\mu p_\nu p_\rho p_\sigma \right] \] (5.33)

or we can rearrange them as in Eq. 5.31, with

\[ \Delta_1 = \frac{3}{8\pi^2(3\alpha + \beta)}\chi^2\mu^2\epsilon c_1 + \frac{\beta - 10\alpha}{192\pi^2\beta(3\alpha + \beta)}\chi^2\mu^2\epsilon c_2 \] (5.34)

\[ \Delta_2 = \frac{10\alpha + 7\beta}{96\pi^2\beta(3\alpha + \beta)}\chi^2\mu^2\epsilon c_2 \] (5.35)

in agreement with the results obtained in [13].

At this order, to eliminate them, we must add

\[ -\frac{2}{\chi^2} \left( -\frac{\chi^2}{2} \Delta_1 \phi_{\mu\nu} D_{\mu\nu} \phi_{\rho\sigma} - \frac{\chi^2}{2} \Delta_2 \phi_{\mu\nu} D_{\mu\nu} \phi_{\rho\sigma} \right) \] (5.36)

to the Lagrangian 5.23 By covariance principle, the counterterms 5.36 correspond to

\[ -\frac{2}{\chi^2} \sqrt{-g} \left( -\chi^2 \Delta_1 R^2 + \chi^2 \Delta_2 R_{\mu\nu} R^{\mu\nu} \right) \] (5.37)

This implies that the renormalization coefficients \( a_{0,R} \) and \( b_{0,R} \) of Eq. 4.78 are now substituted by:

\[ \tilde{a}_{0,R} = a_{0,R} - \frac{\Delta_1}{\alpha}, \quad \tilde{b}_{0,R} = b_{0,R} - \frac{\Delta_2}{\beta} \] (5.38)

and, with an appropriate choice of \( c_1 \) and \( c_2 \), it is possible to set \( \tilde{a}_{0,R} = \tilde{b}_{0,R} = 0 \).

### 5.4 The General Case

The example shown in the previous section is too restrictive and we now see a general method for dealing with the one loop divergences. In this section we keep \( \gamma = 3 \) and consider again the minimal choice of the potential as a combination of operators quartic in Riemann tensor:

\[ S_{NL, \text{limit}} = -\frac{2}{\chi^2} \int d^4x \sqrt{-g} \left( R + \alpha \chi^2 R \Box_\mu \Box_\nu R + \beta \chi^2 R_{\mu\nu} \Box_\mu \Box_\nu R_{\mu\nu} + \sum_i \frac{c_i}{\mu^8} \left\{ \Box_\mu \Box_\nu R_{\mu\nu} \right\} + S_{gf} \right) \] (5.39)
where $i$ denotes a particular contraction of Lorentz indices. Actually, all these quartic operators can be written as a linear combination of the following ones:

$$\mathcal{K}_1 \equiv T_1 R_{\mu_1 \nu_1 \rho_1 \sigma_1} R_{\mu_2 \nu_2 \rho_2 \sigma_2} (D_\gamma R_{\mu_3 \nu_3 \rho_3 \sigma_3}) (D_\delta R_{\mu_4 \nu_4 \rho_4 \sigma_4})$$

$$\mathcal{K}_2 \equiv T_2 R_{\mu_1 \nu_1 \rho_1 \sigma_1} R_{\mu_2 \nu_2 \rho_2 \sigma_2} R_{\mu_3 \nu_3 \rho_3 \sigma_3} R_{\mu_4 \nu_4 \rho_4 \sigma_4}$$

The tensors $T_i$ are independent of the graviton fields and they are a combination of flat metrics with eighteen indices so as to make the operators $\mathcal{K}_i$ scalar.

As shown in Sec. 5.2, it is clear that divergences can only appear if the derivatives of $\mathcal{K}_1$ and $\mathcal{K}_2$ act both on internal legs as depicted in Fig. 5.6.

![Figure 5.6](image)

Since we do know that such divergences correct only the terms $R^2$, we can keep the other two Riemann tensors outside the diagram. That is, all the divergences can be computed from

$$D_1(k) = k_\gamma R_{\mu_3 \nu_3 \rho_3 \sigma_3}(k)(-k)_\delta R_{\mu_4 \nu_4 \rho_4 \sigma_4}(-k) \times P(k)$$

$$D_2(k) = R_{\mu_3 \nu_3 \rho_3 \sigma_3}(k)(-k)_\gamma (-k)_\delta R_{\mu_4 \nu_4 \rho_4 \sigma_4}(-k) \times P(k)$$

where $k$ is the loop momentum and the field indices are understood.

The divergence $D_1$ - where we understand the Lorentz indices - of the first diagram of Fig. 5.6 can be computed integrating $D_1(k)$; we write it as:

$$D_1 = \frac{i}{768\pi^2}\beta(3\alpha + \beta)\chi \mu^2 \left[2(3\alpha + \beta)\eta_{\mu_3 \rho_3} \eta_{\mu_4 \rho_4} - (4\alpha + \beta)\eta_{\mu_3 \rho_3} \eta_{\mu_4 \rho_4} + 2(3\alpha + \beta)\eta_{\mu_3 \mu_4} \eta_{\rho_3 \rho_4}\right] \times$$

$$\times \left[\eta_{\nu_3 \delta} \eta_{\nu_4 \delta} \eta_{\sigma_3 \gamma} + \eta_{\nu_3 \delta} \eta_{\nu_4 \gamma} \eta_{\sigma_3 \delta} + \eta_{\nu_3 \gamma} \eta_{\nu_4 \delta} \eta_{\sigma_3 \delta} + \eta_{\nu_3 \gamma} \eta_{\nu_4 \gamma} \eta_{\sigma_3 \delta} + \eta_{\nu_3 \delta} \eta_{\nu_4 \delta} \eta_{\sigma_3 \gamma} + \eta_{\nu_3 \delta} \eta_{\nu_4 \gamma} \eta_{\sigma_3 \gamma} + \eta_{\nu_3 \gamma} \eta_{\nu_4 \delta} \eta_{\sigma_3 \delta} + \eta_{\nu_3 \gamma} \eta_{\nu_4 \gamma} \eta_{\sigma_3 \delta} + \eta_{\nu_3 \delta} \eta_{\nu_4 \delta} \eta_{\sigma_3 \gamma} + \eta_{\nu_3 \gamma} \eta_{\nu_4 \delta} \eta_{\sigma_3 \gamma} + \eta_{\nu_3 \delta} \eta_{\nu_4 \gamma} \eta_{\sigma_3 \gamma} + \eta_{\nu_3 \gamma} \eta_{\nu_4 \gamma} \eta_{\sigma_3 \delta}ight]$$

Actually this is not the complete divergence, that can be obtained from Eq. 5.44 by symmetrizing with respect to the couples $(\mu_i, \nu_i) \leftrightarrow (\rho_i, \sigma_i)$ ($i = 3, 4$) and antisymmetrizing with respect to $\mu_1 \leftrightarrow \nu_1, \rho_1 \leftrightarrow \sigma_1$ ($i = 3, 4$). However, since at the end the divergence $D_1$ has to be contracted with the tensor $T_1$ of Eq. 5.40, it is much more convenient to embed such symmetrizations directly into $T_1$. 

The divergence that comes after integrating $D_2(k)$, that we call $D_2$, is the same as $D_1$ but for the sign:

$$D_2 = -D_1$$  \tag{5.45}

Moreover, the two divergences have to be weighted by different combinatoric factors (i.e. $\sigma_1 = 2$ for the first diagram of Fig. 5.6 and $\sigma_2 = 6$ for the second one).

Therefore, all the counterterms that stem from the diagrams of Fig 5.6, at the lowest order, can be written as:

$$R^{(1)}_{\mu_1\nu_1\rho_1\sigma_1}(-k)R^{(1)}_{\mu_2\nu_2\rho_2\sigma_2}(k) \left( \sigma_1 T_{\text{sym}}^1 D_1 + \sigma_2 T_{\text{sym}}^2 D_2 \right)$$  \tag{5.46}

where $T_{\text{sym}}^i$ are the conveniently symmetrized versions of the tensors $T_i$ of Eqs. 5.40 and 5.41. We also stress that $T_2$ needs an additional symmetrization in the indices $1,2,3$, because each one of the three $R$ factors (without derivatives attached) can contribute to the divergence. Moreover, because of the Gauss-Bonnet theorem 1.129, we know that all the divergences reduces to linear combinations of $R^2$ and $R_{\mu\nu}R^{\mu\nu}$ only. Therefore, as done in the previous section, the Lagrangian counterterms are obtained from Eq. 5.46 by covariance and they can all be cast in the form:

$$-\frac{2}{\chi^2} \sqrt{-g} \left( -\chi^2 \Delta_1 R^2 - \chi^2 \Delta_2 R_{\mu\nu}R^{\mu\nu} \right)$$  \tag{5.47}

where $\Delta_1$ and $\Delta_2$ are linear combinations of $D_1$ and $D_2$.

The example studied in the previous section can also be obtained using this general case. Provided the correspondence:

$$R^2 D^2 R^2 = 2R^2 D R D R + 2\chi^2 D^2 R$$  \tag{5.48}

if we choose the two tensors:

$$T_1 = \eta_{\gamma\delta} \left( \prod_{i=1}^{4} \eta_{\mu_i\rho_i} \right) \left( \prod_{i=1}^{4} \eta_{\nu_i\sigma_i} \right)$$  \tag{5.49}

$$T_2 = \eta_{\gamma\delta} \left( \prod_{i=1}^{4} \eta_{\mu_i\rho_i} \right) \eta_{\nu_1\nu_2} \eta_{\nu_3\nu_4} \eta_{\sigma_1\sigma_2} \eta_{\sigma_3\sigma_4}$$  \tag{5.50}

we obtain again the divergences of Eqs. 5.32, 5.33.

It is also clear that some choices of the killers cannot provide a finite theory, if $\Delta_1$ or $\Delta_2$ is zero. For instance, the killers $R^2 \Box R^2$ and $R^3 \Box R$ are not enough. They generate divergences that parametrized as in Eq. 5.47 are:

$$\Delta_1(R^2 \Box R^2) = \frac{3}{8\pi^2(3\alpha + \beta)\chi^2 \mu^2 \varepsilon}, \quad \Delta_2(R^2 \Box R^2) = 0$$

$$\Delta_1(R^3 \Box R) = \frac{9}{32\pi^2(3\alpha + \beta)\chi^2 \mu^2 \varepsilon}, \quad \Delta_2(R^3 \Box R) = 0$$

and it is not possible to make the constant $b_0, R$ of Eq. 5.38 vanish.
CONCLUSION AND FUTURE OUTLOOK

In this work we have seen that, if we renounce the principle of locality, we can build a quantum theory of gravity that has chances to be fully consistent with quantum mechanics.

Our starting point was Kuz’min’s nonlocal Lagrangian defined in Eq. 2.16, which involves nonpolynomial functions that satisfy very strict hypotheses. First of all, to get unitarity at the tree level, the bare propagator should have no pole besides the one at $k^2 = 0$. This implies that the nonpolynomial functions have to be transcendental. Moreover, general relativity has to be obtained as the low energy limit of the nonlocal theory, while in the high energy limit the nonlocal theory must resemble a higher derivative one. There are no other nonlocal theories, alternative to Kuz’min’s, that satisfy such hypotheses.

Then, we have demonstrated the super-renormalizability of Kuz’min’s theory by power counting and shown that, if the nonpolynomial functions grow as a polynomial of degree (at least) three, divergences arise only at one loop and the counterterms are local. Actually, only four parameters (the cosmological constant, the Newton constant and the coefficients of the terms $R^2$ and $R_{\mu\nu}R^{\mu\nu}$) are subject to renormalization.

We have also shown that, adding corrections that are at least cubic in the graviton field, we can make the nonlocal theory finite. We have explicitly computed the divergences in the case of a linear combination of $R^2\Box R^2$ and $R_{\mu\nu}R^{\mu\nu}\Box R_{\rho\sigma}R^{\rho\sigma}$ and then given a general recipe to deal with corrections that are quartic in the Riemann tensor.

Although the lack of unitarity in higher derivative models was the main reason that led us to the nonlocal theory, it is not simple to demonstrate the unitarity of nonlocal theories beyond the tree level. This is due essentially to the fact that the largest time equation does not hold in the nonlocal framework and consequently the cutting rules - which successfully provide a demonstration of unitarity and causality for local gauge theories - here fail.

However, only in recent years there has been a growing interest in infinite derivative models. Nonlocal theories are largely unexplored and many issues are yet to be solved, for instance:

- **are nonlocal theories truly unitary?** As we have shown in this work, nonlocality can imply unitarity only at the tree level, but it is necessary a new demonstration, that does not rely on the Cutkosky’s formalism, in order to prove unitarity for nonlocal theories, starting from the perturbation theory;

- **how does nonlocality influence cosmology?** In fact, cosmology is plagued by the presence of singularities and nonlocality can help to remove them. Moreover, nonlocality could also have had a role in the evolution of the early universe. For example, in [11], it is studied how nonlocal gravity can help to better address the problem of black holes singularities.
can nonlocality be tested? Future experiments should tell whether nonlocal phenomena emerge in the high energy regime (or, equivalently, at short distances), otherwise they will cast severe constraints over the nonlocal theory. An example of such an experiment is already in preparation at LENS laboratory in Florence [4].

If these difficulties are overcome, the nonlocal theory of gravity can be merged with the Standard Model of particle physics towards a grand unified theory: this possibility is worth of attention, because it can provide a valid alternative to other more exotic theories of quantum gravity.
NOTATION AND USEFUL EXPANSIONS

A.1 Notation and Definitions

The flat metric is
\[ \eta_{\mu\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \] (A.1)

The explicit expression of the connection \( \Gamma_{\mu,\nu\rho} \) is:
\[ \Gamma_{\mu,\nu\rho} = \frac{1}{2} \left( \frac{\partial g_{\mu\sigma}}{\partial x^\rho} + \frac{\partial g_{\mu\rho}}{\partial x^\nu} - \frac{\partial g_{\nu\rho}}{\partial x^\mu} \right) \] (A.2)

or
\[ \Gamma^\mu_{\nu\rho} = \frac{1}{2} g^{\mu\alpha} \left( \frac{\partial g_{\alpha\nu}}{\partial x^\rho} + \frac{\partial g_{\alpha\rho}}{\partial x^\nu} - \frac{\partial g_{\nu\rho}}{\partial x^\alpha} \right) \] (A.3)

and it is symmetric for exchange \( \nu \leftrightarrow \rho \).

We denote the covariant d’Alambertian as:
\[ \Box_c \equiv D_\mu D^\mu \] (A.4)

and in flat space:
\[ \Box \equiv \partial_\mu \partial^\mu \] (A.5)

The Riemann tensor is:
\[ R_{\mu\nu\rho\sigma} = \frac{1}{2} \left( \frac{\partial^2 g_{\mu\sigma}}{\partial x^\nu \partial x^\rho} + \frac{\partial^2 g_{\nu\rho}}{\partial x^\sigma \partial x^\mu} - \frac{\partial^2 g_{\mu\rho}}{\partial x^\sigma \partial x^\nu} - \frac{\partial^2 g_{\nu\sigma}}{\partial x^\mu \partial x^\rho} \right) + g_{\alpha\beta} \left( \Gamma^\alpha_{\nu\rho} \Gamma^\beta_{\mu\sigma} - \Gamma^\alpha_{\nu\sigma} \Gamma^\beta_{\mu\rho} \right) \] (A.6)

with the properties:
\[ R_{\mu\nu\rho\sigma} = -R_{\nu\mu\rho\sigma} = -R_{\mu\nu\sigma\rho} \]
\[ R_{\mu\nu\rho\sigma} = R_{\rho\sigma\mu\nu} \] (A.7)

and the two Bianchi identities hold:
\[ D_\alpha R_{\mu\nu\rho\sigma} + D_\rho R_{\mu\nu\sigma\alpha} + D_\sigma R_{\mu\nu\alpha\rho} = 0 \] (A.8)

\[ D_\alpha R_{\mu\nu\rho\sigma} + D_\rho R_{\mu\nu\sigma\alpha} + D_\sigma R_{\mu\nu\alpha\rho} = 0 \] (A.9)

The Ricci tensor is
\[ R_{\mu\nu} = g^{\alpha\beta} R_{\alpha\beta\mu\nu} \] (A.10)

and it is symmetric for exchange \( \mu \leftrightarrow \nu \). The curvature is:
\[ R = g^{\alpha\beta} R_{\alpha\beta} \] (A.11)

We will call \( g \) the determinant of the metric:
\[ g = \det g_{\mu\nu} \] (A.12)
A.2 USEFUL EXPANSIONS

If we set
\[ g_{\mu\nu} \equiv \eta_{\mu\nu} + \phi_{\mu\nu} \] (A.13)
in order to get \( g^{\mu\alpha} g_{\alpha\nu} = \delta^\mu_\nu \), we have
\[ g^{\mu\nu} = \eta^{\mu\nu} - \phi^{\mu\nu} + O(\phi^2) \] (A.14)

Here we list some useful expansions that we need throughout this work:

**Connection**
\[ \Gamma^{(0)}_{\nu\rho} = 0 \] (A.15)
\[ \Gamma^{(1)}_{\nu\rho} = \frac{1}{2} \eta^{\mu\alpha} (\partial_\rho \phi_{\alpha\nu} + \partial_\nu \phi_{\alpha\rho} - \partial_\alpha \phi_{\nu\rho}) \] (A.16)

**Riemann tensor**
\[ R^{(0)}_{\mu\nu\rho\sigma} = 0 \] (A.17)
\[ R^{(1)}_{\mu\nu\rho\sigma} = \frac{1}{2} (\partial_\nu \partial_\rho \phi_{\mu\sigma} + \partial_\mu \partial_\sigma \phi_{\nu\rho} - \partial_\nu \partial_\sigma \phi_{\mu\rho} - \partial_\mu \partial_\rho \phi_{\nu\sigma}) \] (A.18)
\[ R^{(2)}_{\mu\nu\rho\sigma} = \eta_{\alpha\beta} \left( \Gamma^{(1)}_{\nu\rho} \Gamma^{(1)}_{\mu\sigma} - \Gamma^{(1)}_{\nu\sigma} \Gamma^{(1)}_{\mu\rho} \right) \] (A.19)

**Ricci tensor**
\[ R^{(0)}_{\nu\sigma} = 0 \] (A.20)
\[ R^{(1)}_{\nu\sigma} = \eta^{\mu\rho} R^{(1)}_{\mu\nu\rho\sigma} = \frac{1}{2} \left( \partial_\nu \partial^\mu \phi_{\mu\sigma} + \partial^\mu \partial_\sigma \phi_{\mu\nu} - \partial_\nu \partial_\sigma \phi^\mu_{\mu\rho} - \Box \phi_{\nu\sigma} \right) \] (A.21)
\[ R^{(2)}_{\nu\sigma} = \eta^{\mu\rho} R^{(2)}_{\mu\nu\rho\sigma} - \phi^{\mu\rho} R^{(1)}_{\mu\nu\rho\sigma} \] (A.22)

**Curvature**
\[ R^{(0)}_{\nu\rho} = 0 \] (A.23)
\[ R^{(1)}_{\nu\rho} = \eta^{\nu\sigma} R^{(1)}_{\nu\sigma} = \partial_\mu \partial_\nu \phi_{\mu\nu} - \Box \phi_{\nu\sigma} \] (A.24)
\[ R^{(2)}_{\nu\rho} = \eta^{\nu\sigma} R^{(2)}_{\nu\sigma} - \phi^{\nu\sigma} R^{(1)}_{\nu\sigma} = \frac{1}{4} \left( -\partial_\mu \phi_{\sigma}^{\alpha} \partial^\mu \phi_{\rho}^{\beta} + 2 \partial_\rho \phi_{\mu\nu} \partial^\mu \phi_{\nu\sigma} - \partial_\mu \phi_{\nu\sigma} \partial^\mu \phi_{\nu\sigma} \right) \] (A.25)
In this work, renormalization is carried out in the framework of dimensional regularization. We parametrize divergences with $\epsilon$ defined by $D = 4 - \epsilon$. If we consider a rotationally invariant function $f(p^2)$, we have:

$$\int_{-\infty}^{\infty} \frac{d^D p}{(2\pi)^D} f(p^2) = \frac{1}{2^{D-1} \pi^{D/2} \Gamma(D/2)} \int_0^{\infty} dp \, p^{D-1} f(p^2) \quad (B.1)$$

In $D$ dimensions, the integral of the solid angle is:

$$\int d\Omega_D = \frac{2\pi^{D/2}}{\Gamma(D/2)} \quad (B.2)$$

$\Gamma$ is defined as the function such that:

$$\Gamma(x + 1) = x\Gamma(x) \quad (B.3)$$

In particular, for $n$ integer:

$$\Gamma(n + 1) = n! \quad (B.4)$$

For the lowest values of $D$, $\Gamma$ and the solid angle are:

<table>
<thead>
<tr>
<th>$D$</th>
<th>$\Gamma(D/2)$</th>
<th>$\int d\Omega_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\sqrt{\pi}$</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$2\pi$</td>
</tr>
<tr>
<td>3</td>
<td>$\sqrt{\pi}$</td>
<td>$4\pi$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>$2\pi^2$</td>
</tr>
</tbody>
</table>

We have the expansions:

$$\Gamma\left(\frac{n}{2}\right) = \sqrt{\pi} \frac{(n - 2)!}{2^{n-1}/2} \quad (B.5)$$

$$\Gamma(z) = \frac{1}{z} - \gamma_E + O(z) \quad (B.6)$$

$$\Gamma(z) = \sqrt{\pi} \left[ 1 + \left(z - \frac{1}{2}\right) \psi^{(0)}\left(\frac{1}{2}\right) + O\left((z - \frac{1}{2})^2\right) \right] \quad (B.7)$$

where $\gamma_E = 0.5772...$ is the Euler-Macheroni constant and $\psi^{(0)}\left(\frac{1}{2}\right) = -1.9635100$. 


B.2 CONVERGENCE

The integral:
\[ \int_{a}^{\infty} x^{\alpha} \, dx, \quad \text{with } a > 0 \]
converges if \( \alpha > -1 \).

The integral:
\[ \int_{0}^{\alpha} x^{\alpha} \, dx, \quad \text{with } a > 0 \]
converges if \( \alpha < -1 \).

Hence, if we consider the integral:
\[ \int_{-\infty}^{\infty} dD \frac{d^{D}p}{(2\pi)^{D}} f(p^{2}) \]
with \( D \) a complex number and \( f(p^{2}) \) such that:
\[ f(p^{2}) \sim (p^{2})^{\alpha_{UR}} \text{ as } p^{2} \to \infty \]
\[ f(p^{2}) \sim (p^{2})^{\alpha_{IR}} \text{ as } p^{2} \to 0 \]
it is convergent if:
\[ -2\alpha_{IR} < \Re D < -2\alpha_{UR} \quad (B.8) \]

B.3 FEYNMAN PARAMETERS AND OTHER USEFUL INTEGRALS

Here we list some useful formulas, that are needed when computing divergences:

\[ \frac{1}{AB} = \int_{0}^{1} dx \frac{1}{|Ax + B(1-x)|^{2}} \quad (B.9) \]

\[ \frac{1}{A^{\alpha}B^{\beta}} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{0}^{1} dx \frac{x^{\alpha-1}(1-x)^{\beta-1}}{|Ax + B(1-x)|^{\alpha+\beta}} \quad (B.10) \]

\[ \int_{0}^{1} dx \, x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (B.11) \]

\[ \int \frac{d^{D}p}{(2\pi)^{D}} \frac{(p^{2})^{\beta}}{(p^{2} - m^{2})^{\alpha}} = \frac{i(-1)^{\alpha+\beta}\Gamma(\beta + \frac{D}{2})\Gamma(\alpha - \beta - \frac{D}{2})}{(4\pi)^{\frac{D}{2}}\Gamma(\alpha)\Gamma(\frac{D}{2})} \left( m^{2} \right)^{\frac{D}{2} - \alpha + \beta} \quad (B.12) \]

\[ \int \frac{d^{D}p}{(2\pi)^{D}} \frac{1}{(p^{2} - m^{2})^{\alpha}} = \frac{i(-1)^{\alpha}\Gamma(\alpha - \frac{D}{2})}{(4\pi)^{\frac{D}{2}}\Gamma(\alpha)} \left( m^{2} \right)^{\frac{D}{2} - \alpha} \quad (B.13) \]


