UNIVERSIDAD NACIONAL DE COLOMBIA

Higher-order time derivative theories. Interpretation, instability and possible stabilization

Juan Mauricio Valencia Villegas

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá, Colombia
2017

Higher-order time derivative theories. Interpretation, instability and possible stabilization

Juan Mauricio Valencia Villegas

Trabajo de grado presentado como requisito parcial para optar al título de:
Magíster en Ciencias - Física

Director:
Ph. D. Justo Javier López Sarrión

Línea de Investigación:
Física matemática

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá, Colombia
2017

Say there be;
Yet nature is made better by no mean
But nature makes that mean: so, over that art,
Which you say adds to nature, is an art
That nature makes.

William Shakespeare, *A Winter's Tale*

Abstract

Higher-derivative field theories are well known to propagate ghost degrees of freedom (DOFs), an instability known as of Ostrogradsky. However, recent advances proposing conditions to stabilize this kind of models, with finitely many unstable DOFs in a non-minimal way, by including a stabilizer DOF and a kinetic coupling between both, have opened the question whether an extension of this methodology to relativistic field theories, also works.

In this thesis, the Pais-Uhlenbeck Lagrangian density with a higher derivative scalar field, which leads to an unstable theory, is considered as a basis for a toy model. Upon the requirement for the Lagrangian to be a Lorentz scalar, as well as for the transformation properties of both, the unstable and the stabilizer DOF, to be consistent with a kinetic constraint that controls the instability, it is first concluded that at the level of a free theory the stabilizer must be a vector field. The latter is also motivated to make plausible an extension to interacting higher derivative theories. This is, the kinetic instability should be controled already at the free theory, in such a way that the Feynman propagator does not show a ghost DOF.

A Hamiltonization with constraints is considered in order to deal with the imposed kinetic-constraint, which is at the core of the stabilization. This approach allows to examine the properties of the Ostrogradskian instability as it has been done up until now in the literature, therefore making evident the successful extension of the stabilization properties, at least in this toy model. Furthermore, the physical DOFs propagated by the theory are found, and the physical Hamiltonian written in terms of these, turns out to be positive definite and bounded from below in certain region of parameter space. In particular, a very interesting relation between the coupling parameter (α) of the higher-derivative term of the scalar field and the mass of the stabilizer field (m), arises as a requirement for the stabilization. The condition is a lower bound on the former, of the form $\alpha > 1/m$. Such relation was completely unexpected but more meaningful for the physical interpretation of the new higher-derivative structure, because it would show the energy scale at which these new terms may become important.

Keywords: higher derivatives, Ostrogradskian instability, quantum theories, quantization with constraints.

Resumen

Las teorías de campos con derivadas altas propagan grados de libertad inestables, *ghost DOFs*, una inestabilidad conocida como de Ostrogradsky. Sin embargo, avances recientes en que se proponen condiciones para estabilizar modelos de este tipo, con finitos grados de libertad inestables (DOFs), en una extensión no trivial, incluyendo un DOF estabilizador y un acople cinético entre ambos, han abierto nuevamente la pregunta de si una extensión de esta metodología a teorías de campos relativista, también funciona.

En esta tesis, la densidad Lagrangiana de Pais-Uhlenbeck con un campo escalar con derivadas altas, que da lugar a un modelo inestable, es considerada como base para un modelo de juguete. Demandando que el Lagrangiano sea un escalar de Lorentz, así como de requerir que las propiedades de transformación de ambos, DOFs inestable y estabilizador, sean consistentes con la ligadura cinética que controla la inestabilidad, se concluye en primera instancia que el campo estabilizador debe ser de hecho un campo vectorial. Esto último es motivado para hacer plausible una extensión a teorías interactuantes con derivadas altas. Es decir, la inestabilidad cinética debe ser controlada al nivel de la teoría libre, de tal manera que el propagador de Feynman no evidencie un *ghost* DOF.

Debido a que la ligadura cinética que se impone al modelo es clave para la estabilización, una formulación Hamiltoniana con ligaduras es adoptada para el análisis. Esta aproximación permite evaluar las propiedades de la inestabilidad de Ostrogradsky como se ha hecho previamente en la literatura, por tanto, haciendo evidente una extensión exitosa de las propiedades de estabilización, al menos en el modelo de juguete considerado. Adicionalmente, se identifican los grados de libertad físicos propagados por la teoría y el Hamiltoniano físico escrito en términos de estos últimos, resulta ser postivo y acotado inferiormente en cierta región del espacio de parámetros. En particular, de demandar la estabilización, resulta una relación muy interesante entre el parámetro de acople del término con derivadas altas (α) del campo escalar inestable y la masa del campo estabilizador (m). La condición es una cota inferior para el primero, del tipo $\alpha > 1/m$. Esta última relación, aunque completamente inesperada, resulta más enriquecedora para la interpretación de la nueva estructura de derivadas altas, porque esta daría idea de la escala de energías a la que estos nuevos términos podrían resultar importantes.

Palabras clave: derivadas altas, inestabilidad de Ostrogradsky, teorías cuánticas, cuantización con ligaduras.

Contents

1 Introduction

The most of the fundamental theories are written in terms of a Lagrangian that includes at most first order derivatives of the dynamical variables, but there seems to be no fundamental reason to start in that way besides the well-known appearence of ghosts in quantum theories [1]; propagating degrees of freedom with the "wrong" sign in the Feynman propagator, which signal a highly unstable, non-physical theory [1, 2, 3, 4, 5]. However, some very specific higher-derivative models have been found to be stable in this sense [6, 7, 8, 9, 10, 11, 12, 13], they avoid these peculiar ghosts, a instability also called as "of Ostrogradsky" [14, 15], which signals that the only possible fundamental reason to avoid higher derivatives, may just give further insights on how to build theories without ruling them out.

It is also truth that simpler models, *e.g.* those without higher-derivative terms, whose predictions seem to be in good accordance with the observed phenomena are preferred, and this is in indeed a good first approach. But it should be considered that at some point new phenomena may be observed, such that the mathematical structure that was up until that point conferred to the fundamental theories, may no longer be enough. Even worst may be the case in which trying to avoid a more general but unfamiliar formulation, cumbersome paths are taken in defense of a more convenient approach that so far had worked well, and that is convenient, precisely because enough time has been given for it to be understood.

The ideal approach would be to have a conclusive fact deciding instead of us, and our perhaps distorted sense of aesthetics, whether higher-derivative terms should or not be included. In this regard, progress with higher-derivative theories has to be made in order for them to be tested. Furthermore, there are already indications of the need of higher-derivative structure, for instance, in perturbation theory for gravity, as it is pointed out by Hawking [16]. In any case, it would be surprising that an assumption made so long ago, starting with second order equations of motion [2, 3], does not ever need to be changed, because when restricting to low-derivative theories an implicit assumption is being taken without proof: the dynamics to be described does not need of higher-order derivatives. This is a very strong assumption that is easy to forget with the many successful results encountered so far in physics, just by restricting to the lower-derivative case.

Now, accordingly with these ideas, specially those first remarked in [16], research in general properties of higher derivative theories [1, 17, 18, 19, 20, 21, 22, 23], modifications of the Standard Model [24, 25], as well as theories of quantum gravity including higher derivatives [26, 27, 28, 29, 30, 31, 32, 33], and their implications in cosmology [34, 35, 36, 37, 38] has increased significantly in recent years. The later have also been motivated by the recent observation of gravitational waves that set limits in these modifications to general relativity, among which, many higher-derivatives theories, as Horndesky, have been even ruled out [35, 36].

Nevertheless, any progress for this topic is expected to be very slow, because it is unknown territory in many aspects. Interpretation may be the most difficult barrier, because assigning an interpretation in a theory to a new object, only can come after understanding how that new object enters in the existing theories, *i.e.* how it interacts with the already known objects and how the outcomes are modified. Even worst is for that new concept, if its naive introduction poses immediately problems that seem to be related to the object itself, when, however, it could be that the previous construction methods are inappropiate to introduce the new concept.

The latter may be the case for higher derivatives. This is: firstable, the interpretation has, in general, not been assigned yet. Second, these models are not well understood because in some cases they lead to theories with nice properties, while in many other cases they do not. Furthermore, what leads to theories with nice properties, from now on called "stable", and to unstable theories, is also not completely understood. Therefore, it is possible to think that the "objects" themselves - the new higher-derivative structure being conferred to the theories - is not the problem, but the way in which the structure is being introduced.

Important advances regarding the second problem have been obtained recently in [39, 40, 41], where some conditions for the higher derivative theories to be stable have been proposed. These are only valid for classical models with a finite number of degrees of freedom (*dof* s), and a formal extension to field theories has not been established yet. Furthermore, they can be understood as a modification of the methodologies in which the new structure, higher derivatives, had naively and unsuccessfully been introduced up until now. What remains to be seen, is if that methodology for finite *dof*s, is just a limiting case of a broader and more general methodology that still needs to be formulated, that would allow the higher-derivative structure to be introduced into field theories, also making them susceptible to be quantized.

Despite the previously mentioned conditions may not be the only possibility to reach stable higher-derivative theories, they are in fact the only well established methodology that has been proposed so far, and there are some authors that held the opinion that it is indeed the only possible way to reach the above-mentioned objective [2, 3, 39, 40, 41]. For that reason, the intention with this thesis is to contribute to the understanding of how the introduction of these new structures by means of these new methodologies, can be made consistent specifically in continuum field theories. A general extension of the methodologies is not in the objectives of this introductory approach, and therefore, only some, very particular models are proposed.

Knowing before hand that higher derivatives introduced carelessly in a theory lead to instabilities, the next less naive approach has been taken. This is: the proposed toy models have been built by starting from a naive extension of the methodologies given in [39], even if they were only intended for finite *dof*s. A constraint between the two fields arises in the system, because the methodology requires a very special kind of coupling between a stable low-derivative field, and possibly, we claim, a field whose dynamics require higher-derivatives to be described. In short, this will turn out to be a system with constraints, and hoping to acquire some insight about the interpretation, the extended Hamiltonian formalism with constraints has been taken. In this formalism, a physical Hamiltonian of the initial system is found by defining the Dirac brackets and writing a new set of variables that are canonical in these new brackets. This equivalent system has much nicer properties that greatly simplify the analysis of the dynamics of the system. Furthermore, they allow to check the stability. Some models with finite *dof*s have also been studied mainly motivated for further interpretation of the results obtained in field theory.

Finally, it has been verified that after some subtleties regarding the construction of the field theory model and in particular the transformation properties of the stabilizer field, it is possible to build a classical continuum field theory that can, in principle, be brought to the starting point of a higher-derivative quantum field theory. The quantization is done in the canonical formalism, because the already found Physical Hamiltonian, that served for the stability analysis, and the Dirac brackets, are the main objects in this scheme.

Furthermore, precisely related to the interpretation of this new structure, it turns out that the only way to stabilize the higher-derivative scalar field for this particular toy model, forces a relation between the coupling parameter of the higher-derivative scalar field (α), and the mass parameter of the stabilizer field (m). The condition is, at least for this particular field toy model, not the most interesting upper bound, but only a lower bound of the kind $\alpha > \frac{1}{m}$. However, the existance of such relation was unexpected and more meaningful for the physical interpretation of higher-derivative terms.

Even though it is not claimed that this kind of relation should appear in every possible stable higher-derivative field theory. The sole fact that this toy model with these interesting properties exists, speaks about possibly physically interesting higher-derivative theories, that include such kind of relations. If some higher-derivatives are needed to describe some dynamics in a physical phenomema, this situation would be much more interesting than a completely "disconnected" term, from all the other objects in the theory. Again, coming back to the first arguments of this discussion.

2 Objective

2.1 General objective

Identify the source of the stability in the sense of Ostrogradsky, for some particular higher-order Lagrangians with a finite number of degrees of freedom (HTDt - fDOF), characterize this source giving it a physical interpretation and recognize the difficulties involved when extending this successfully stable conditioned HTDt - fDOF to Lagrangians with infinite degrees of freedom (HTDt - iDOF) as a way to understand the possible extensions to quantum field theory.

2.2 Specific Objectives

- Understand the mechanism through which the classically imposed conditions on the HTDt - fDOF Lagrangians, successfully guarantee the stability of the associated theories.
- Recognize possible difficulties when extending successfully stable conditioned HTDt - fDOF to infinite degrees of freedom (iDOF) and identify the source of these potential, stability-instability changes [2, 3], when carrying out the extension.
- Perform the quantization of the particular HTDt - iDOF that have inherited the stability properties of the corresponding finite degree of freedom counterpart, as a first approximation to scalar field theories.
- Examine the vacuum stability of the quantized HTDt - idf

3 Theoretical framework and recent advances in Higher-order time derivative theories (HTDt)

In this section, important advances on current research on HTDt are thoroughly examined and the basic ideas about fundamental concepts to be employed later are reviewed. Upon the former, all the results and contributions to the subject, that shall be exposed in section 4, will be based.

To begin with, a brief introduction to Higher-order time derivative theories (HTDt) is given. The whole introduction to the topic is presented in the simplest possible scheme, which is classical mechanics with finite degrees of freedom. This will later help to emphasize, when the instability is discussed, that the fundamental problems arising in HTDt do not only appear when quantizing the theories. Then, the emergence of the instability and its consequences to the theories, followed by a detailed development of the fundamental ideas about estabilization and common missconceptions about this procedure, are carefully studied.

This review, along with a section on quantization of theories with constraints, shall serve as a basis for the later discussion and will also be the foundation for new ideas to be proposed, regarding the interpretation of the whole stabilization procedure (See sections 4 and 5).

3.1 Introduction to HTDt

The most of the theories that have been written to describe fundamental processes in nature, can be deduced from a Lagrangian with derivatives that at most, include first order in the time derivatives. With this in mind, by definition, every Lagrangian with more than one time-derivative shall be included in the set of higher-order time derivative Lagrangians. These, by means of the least action principle, lead to higher-order time derivative theories (HDTt) [2, 3].

Definition 3.1 (HTDt) *A higher-order time derivative theory results from a higher-order time derivative Lagrangian, by considering the least action principle. The Lagrangian cannot be trivially reduced by partial integration to an equivalent one with low-order time derivatives, otherwise it leads to a normal theory as opposed to HTDt. [2, 3, 39, 40, 41]*

Even though the instability will be discussed in the following sections, it is worth mentioning the main ideas at this regard before getting into the details.

Theorem 3.1 (Ostrogradsky's theorem) *A non-degenerate, higher-order time derivative Lagrangian, leads to a theory with an Ostrogradskian instability [2, 3], which is a kinetic instability with an arbitrarily fast time scale [6].*

The *Ostrogradskian instability*, can be tracked to a linear term in the Hamiltonian that makes it unbounded from below [2, 3, 39], and as will be further explained, leads to negative norm states or lost of unitarity in quantum theory.

These ideas are made precise in what follows and some examples of the Ostrogradskian instability are also considered.

3.1.1 Higher-order time derivative lagrangians with a finite number of degrees of freedom

Consider a Lagrangian that depends on one degree of freedom and its N time derivatives. Denoting $x^{(i)}$ as the i-th time derivative, the Lagrangian is written as:

$$L = L(x, x^{(1)}, \ldots, x^{(N)}) \tag{1}$$

Following the definition given in section 3.1, the Lagrangian (1) cannot be trivially integrated to obtain an equivalent low-order Lagrangian[1]. On the other hand, let us consider a non-degenerate Lagrangian, just as was required in the above given definition.

Non-degeneracy means,

$$\frac{\partial^N L}{\partial (x^{(N)})^N} \neq 0 \tag{2}$$

or equivalently, $\left(\frac{d}{dt}\right)^N \frac{\partial L}{\partial x^{(N)}} \neq 0$, which implies that there exists a function $\mathcal{A} = \mathcal{A}(x, x^{(1)}, \ldots, x^{(N-1)}) \neq 0$, *i.e.* it is possible to solve for $x^{(N)} = \mathcal{A}(x, \ldots, x^{(N-1)}, P_N)$ [2, 3], where the conjugate momenta P_N will be defined later. In other terms, this condition means that the equations of motion are of order 2N, that the Hessian matrix has non-vanishing determinant. This is relevant to identify the source of the instability and also, to identify the ways to estabilize these theories without making them trivially reducible to low-order Lagrangians, leading to second order equations of motion.[2]

The equations of motion are derived as usual, by means of the least action principle. $x(t)$ and $x^{(1)}(t), \ldots, x^{(N-1)(t)}$ are fixed at the boundary ∂. With all this in mind, for a general variation δx, the expression reduces to:

$$\begin{aligned}
\delta S &= \int \left(\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial x^{(1)}} + \cdots + \left(-1\frac{d}{dt}\right)^N \frac{\partial L}{\partial x^{(N)}} \right) \delta x dt = 0 \\
0 &= \left(\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial x^{(1)}} + \cdots + \left(-1\frac{d}{dt}\right)^N \frac{\partial L}{\partial x^{(N)}} \right) \\
0 &= \sum_{i=0}^{N} \left(-\frac{d}{dt}\right)^i \frac{\partial L}{\partial x^{(i)}} \qquad (3)
\end{aligned}$$

The equation (3) gives the equations of motion for a general N-order Lagrangian under the assumption that the degree of freedom, as well as its N-1 time derivatives are kept fixed at the boundaries.

[1] The equivalence between Lagrangians differing by total derivatives, is provided at the level of equivalent equations of motion, when the least action principle is taken into account.

[2] Further explanation regarding the non-degeneracy, is given below in an example involving the Pais-Uhlenbeck Oscillator.

Also notice from equation (3), that the equations of motion are of order 2N, and thus, the phase space of the theory has equal number of dimensions. The canonical variables are chosen to be [2, 3, 14],

$$X_i = x^{(i-1)} \quad P_i = \sum_{j=i}^{N} \left(-\frac{d}{dt}\right)^{j-i} \frac{\partial L}{\partial x^{(j)}} \tag{4}$$

This selection generates time evolution by means of the Hamiltonian built by the Legendre transform and the Hamilton equations of motion. However, particularly in field theory, there are non trivial options that are going to be considered for a discussion in section (4.2). Therefore, as an aside, the following possible definition of canonical fields is introduced for future reference [42]:

$$\begin{aligned} P^{\mu_1 \ldots \mu_i} &= \frac{\partial \mathcal{L}}{\partial(\partial_{\mu_1} \ldots \partial_{\mu_i} \varphi)} - \partial_{\mu_{i+1}} P^{\mu_1 \ldots \mu_i \mu_{i+1}} \\ i &= 1, \ldots m-1 \text{ , being m the highest derivative in the Lagrangian density} \\ P^{\mu_1 \ldots \mu_m} &= \frac{\partial \mathcal{L}}{\partial(\partial_{\mu_1} \ldots \partial_{\mu_m} \varphi)} \end{aligned} \tag{5}$$

The Hamiltonian density is built using these definitions, as usual, by means of the Legendre transform. Since this is just an extension to fields that would unecessarily complicate the following introduction to the Ostrogradskian instability, we come back from this detour to finite dofs.

It is also worth mentioning that in theories with constraints it is necessary to adopt for consistency, as a first approach, an "extended Hamiltonian formalism" [43, 44, 45, 46, 47]. The basic ideas are reviewed in section (3.4).

With the definition given above in [14], the Hamiltonian is built as:

$$\begin{aligned} H &= \sum_{i=1}^{N} P_i x^{(i)} - L(x, x^{(1)}, \ddots, x^{(N)}) \\ H &= P_1 X_2 + \cdots + P_N x^{(N)} - L(X_1, \ldots, X_N, x^{(N)}) \end{aligned} \tag{6}$$

and the Hamilton equations are written as usual:

$$\dot{X}_i = \frac{\partial H}{\partial P_i} \quad \dot{P}_i = -\frac{\partial H}{\partial X_i} \tag{7}$$

Before going on with the basic ideas about HTDt, let us consider one example to help clarifying the whole idea: the Pais-Uhlenbeck oscillator. This model will be recurrent throughout this work and corresponds to N=2, *i.e.* Lagrangian with time-derivatives up to order 2.

- **Pais-Uhlenbeck oscillator. N=2:** The general Lagrangian is [3],

$$L = L(x, x^{(1)}, x^{(2)})$$

[3]Recall that the notation $x^{(i)}(t)$ stands for the i-th time derivative of $x(t)$.

and applying 3, the Euler-Lagrange equation is simply,

$$-\frac{d^2}{dt^2}\frac{\partial L}{\partial x^{(2)}}+\frac{d}{dt}\frac{\partial L}{\partial x^{(1)}}-\frac{\partial L}{\partial x}=0$$

Let us consider what happens if the Lagrangian is degenerate. Applying expression 2 or its equivalent, degeneracy means that,

$$\frac{d^2}{dt^2}\frac{\partial L}{\partial x^{(2)}}=0 \tag{8}$$

Which implies that the equation of motion reduces trivially to the one obtained for a Lagrangian of the form $L=L(x,x^{(1)})$. Note that for degenerate Lagrangians, it is not possible to solve for $x^{(N)}=\mathcal{A}(x,\ldots,x^{(N-1)},P_N)$ because $x^{(N)}$ does not even appear in the equations of motion. However, for N=2, a well known case is recovered, *i.e.* $x^{(2)}=\mathcal{A}(x,\ldots,x^{(1)},P_1)$, which is simply the expression for the acceleration [2, 3].

With this in mind, requiring nondegeneracy, the equation of motion is of fourth order, which implies a totally different form in comparison with the usual Newton's equation, $x^{(4)}=\mathcal{F}(x,x^{(1)},x^{(2)},x^{(3)})$ [2, 3].

Having considered the general implications of a non-degenerate second order (in time) Lagrangian, let us examine the Pais-Uhlenbeck oscillator.

Definition 3.2 (Pais-Uhlenbeck oscillator) *The Pais-Uhlenbeck [15] oscillator is defined by the Lagrangian.*

$$L=-\frac{\epsilon}{2w^2}x^{(2)2}+\frac{1}{2}x^{(1)2}-\frac{w^2}{2}x^2 \tag{9}$$

This Lagragian, gives the equation of motion,

$$\frac{\epsilon}{w^2}x^{(4)}+x^{(2)}+w^2x=0 \tag{10}$$

whose solution can be expanded as $x(t)=C_+\cos(k_+t)+S_+\sin(k_+t)+C_-\cos(k_-t)+S_-\sin(k_-t)$, where the constants depend on the initial values given by $(x_0,x_0^{(1)},x_0^{(2)},x_0^{(3)})$, and the frequencies are

$$k_\pm=w\sqrt{\frac{1\mp\sqrt{1-4\epsilon}}{2\epsilon}} \tag{11}$$

Writing the Hamiltonian by means of the usual expression (6) but with the extended definition for the canonical variables as in (4), the Hamiltonian can be rewritten in two different ways [2]:

$$H=P_1X_2-\frac{w^2}{2\epsilon}P_2^2-\frac{1}{2}X_2^2+\frac{w^2}{2}X_1^2 \tag{12}$$

where the term P_1X_2 is linear in P_1, signaling an unboundedness from below for the energy. On the other hand, it can aslo be rewritten as,

$$H = \frac{1}{2}\sqrt{1-4\epsilon}k_+^2(C_+^2 + S_+^2) - \frac{1}{2}\sqrt{1-4\epsilon}k_-^2(C_-^2 + S_-^2) \tag{13}$$

It is evident in (13), that the second term carries negative energy, due to the "wrong" minus sign in front of it. However, one may argue that the Hamiltonian may not be the energy in these extended theories. At this regard, the canonical variables selection made by Ostrogradsky, here shown in (7), solves this problem, because it is possible to verify that this Hamiltonian generates time evolution, which allows us to identify it with the energy.

Another important property of the Pais-Uhlenbeck oscillator, is that it can be rewritten as two different harmonic oscillators, one of them carrying negative energy [7, 9, 10, 11, 16], just as it is suggested by the form that takes the Hamiltonian in equation (13). This property will be of ample use later in section 4, and therefore, it is worth giving some details.

Consider the following Lagrangian,

$$L = \frac{1}{2}(\dot{x}_1^2 - k_1^2x_1^2) - \frac{1}{2}(\dot{x}_2^2 - k_2^2x_2^2) \tag{14}$$

where the notation $x^{(1)} = \dot{x}$ will be adopted from now on, and $k_{1,2}$, are given by:

$$k_{1,2} = \frac{1}{2g}(1 \mp \sqrt{1-4gw^2}) \tag{15}$$

It is easy to verify that under the following transformation (16), the equation (14) is in fact the Pais-Uhlenbeck Lagrangian (17).

$$x_1 = (\partial_t^2 - k_2^2)x \qquad x_2 = (\partial_t^2 - k_1^2)x \tag{16}$$

Applying the equations (16) to (14), there is only one extra term of the form $\dddot{x}\dot{x}$ that in fact only differs from $\ddot{x}^2$ by a total derivative, and therefore, under the equivalence of both Lagrangians, the following is obtained:

$$L = -\frac{g}{2}\ddot{x}^2 + \frac{1}{2}\dot{x}^2 - \frac{1}{2}w^2x^2 \tag{17}$$

Notice that (14) is, as previously mentioned, the Lagrangian for two decoupled harmonic oscillators. However, one of them, (x_2), has the "wrong" sign. Signaling again that this mode carries negative energy.

The simple fact of a term carrying negative energy in the Hamiltonian (see (13), or the corresponding Lagrangian (17)), which can also be traced out to a linear term in the conjugate momenta (see (12)), lies at the very center of the instability problem.

3.2 The Ostrogradskian instability

The Ostrogradskian theorem is sometimes considered as a quite general restriction to develop fundamental theories with higher derivatives [2, 3, 8]. In what follows, the arguments that are usually exposed to support the latter, are reviewed. Most of the introductory ideas presented in this section, are largeley based on recent contributions, as in [2, 3, 6].

Before going into the details, it is worth pointing out that a way to avoid all the disadvantages to be exposed in brief, can be bypassed by allowing the Lagragians to be degenerate while involving a special kind of coupling to other degrees of freedom [2, 3, 39, 40, 41]. This kind of degeneracy does not lead to a trivial reduction into a low-order time derivative theory and therefore may serve as a basis for new theories with better renormalization properties[4], among others [6, 16, 39]. In section 4 some contributions to the interpretation of these stabilization procedures, as well as an application to the construction of a stable higher-derivative quantum field theory, are given.

The following review on the instability properties of HTDt is divided into two parts. The first section gives the fundamental ideas about the instability, which apply not only to quantum theory but also to classical mechanics, and therefore are approached from the latter formalism. This approach is based on the idea that a classical theory is quantized (which has a lot to do with the historical development), but the full meaning of the instabilities should instead arise fundamentally from quantum theory, and then by the Correspondance Principle should converge to classical theory. However, both approaches are equivalent stating that the instability survives quantization (and second quantization)[3], or that the instability in quantum theory survives when the classical limit is taken.

The second part of this brief review is completely devoted to the instability implications that only pertain to quantum theory, or that are better understood in this formalism.

3.2.1 General implications of the Ostrogradskian instability

To begin with, let us recall the general form that takes the Hamiltonian for a non-degenerate Lagrangian with up to N time derivatives and only one degree of freedom, $L = L(x, x^{(1)}, \ldots, x^{(N)})$:

$$H = P_1 X_2 + \cdots + P_N \mathcal{A}(X_1, \ldots, X_N, P_N) - L(X_1, \ldots, X_N, x^{(N)}) \tag{18}$$

where the canonical variables were defined as[5]:

$$X_i = x^{(i-1)} \quad P_i = \sum_{j=i}^{N} \left(-\frac{d}{dt}\right)^{j-i} \frac{\partial L}{\partial x^{(j)}} \tag{19}$$

[4]The same criteria to realize that a higher derivative field, not coupled to other dofs, propagates ghost dofs, can be used to form a basic idea of the renormalization properties of HTDTs. This is, since the equations of motion for a higher derivative field are of higher order, and the Green's function is the inverse of the differential operator, one expects that the Feynman propagator in momentum space, goes as the inverse of higher powers of momenta than 2, then the superficial degree of divergence may be better behaved and the UV behaviour of these theories may be better than lower derivative theories. Nevertheless, this is a very rough idea that has indeed many subtleties. For details on renomalization properties of higher derivative field theories, see for instance [26].

[5]For details on this formulation, see section 3.1

In equation (18), the explicit function $\mathcal{A} = x^{(N)}$ is written, as a consequence of the non-degeneracy. However, note that, in all the other terms, the canonical variables X_i are not invertible in terms of the conjugate momenta P_i, and therefore, they entail the unboundedness of the energy spectrum[6]. Since any P_i ($\neq P_N$) is not functionally dependent to another conjugate momenta, neither quadratic, it can be very high while X_{i+1} is in turn low, while keeping the state in the same energy hypersurface. These linear momenta in the Hamiltonian (18) are called Ghosts, and its implications are easily seen in specific models.

For instance, consider again the example shown in section (3.1), about the Pais-Uhlenbeck oscillator (P-U). It was shown that the Lagrangian can be written as that of two harmonic oscillators where one of them has the "wrong sign", making it carry negative energy:

$$L = \frac{1}{2}(\dot{x}_1^2 - k_1^2 x_1^2) - \frac{1}{2}(\dot{x}_2^2 - k_2^2 x_2^2) \tag{20}$$

When going to the Hamiltonian formalism, it is evident that for a constant energy, x_1 modes can excite to arbitrarily high energies, while the x_2 "excite" to very negative ones, in order to make up for the energy taken by x_1. Furthermore, as noted in [3], as more excited x_1 gets, more ways are open for the negative oscillator x_2 to compensate the difference and keep the system in the same energy hypersurface in phase space.

The latter can be understood in the following way. If one restricts to the x_1 and p_1 sector of the phase space (being p_1 the conjugate momenta associated with x_1), as p_1 grows, more states are created in this sector of the corresponding energy hypersurface. To clarify this, consider the usual harmonic oscillator, where the energy hypersurface is the ellipse $H(p,q) = \frac{p^2}{2m} + \frac{kq^2}{2}$. If $|\vec{p}|$ grows, the perimeter of the ellipse grows, which in general would be stated as: the number of microstates in the energy hypersurface has increased. In this little detour from the main example, the energy has in fact changed for the normal harmonic oscillator, because the amplitude of oscillation has increased. However, in the P-U oscillator, there is still another possibility, which is what makes these theories different. Recall that we have restricted to the x_1 and p_1 sector of the phase space, and then, let us now explicitly consider the sector x_2 and p_2. In this sector, the energy created is negative, or equivalently, the surface of the hypersurface could be understood to count to the total, as negative, or in other words, the number of microstates in this section of the hypersurface, could be asigned with a negative sign, though they count indeed and the system can occupy states in this sector. With all this in mind, it is easy to see that, while staying on the same energy hypersurface, the number of microstates increase for larger excited modes of x_1 and x_2, since additional area ($\propto$ number of microstates) created in the hypersurface (given a certain reference), can vanish identically between the two sectors. The latter is usually characterized in the literature [2, 3], as:

- *Large $|\vec{p}|$ do not decouple.*

First, note that no interaction has been brought to discussion yet. Therefore, it is important to clarify that all the positive - negative energy exchange, between the modes, that has been described

[6] The unboundedness of the energy spectrum is granted, when the Hamiltonian is recognised as the energy. This is non-trivial for these extended HTDt, and thus, a simple proof will be given at the end of this section.

above as ”made by hand”, is in fact quite real whenever the two sectors interact (by means of an intercation term). This can be stated as:

- *The negative energy modes are not observable whenever there is no interaction, but if the positive and negative sectors couple, the instability becomes important.*

Notice that the above-mentioned creation of microstates in phase space is, as noted in [2], favoured by entropy production. One can also notice that this fact is independent of the energy reference itself. Thus, if a vacuum is defined in these non-stabilized theories, it immediately decays in many particles of positive and negative energy, while keeping the total energy 0.

It is also important to bring to attention that the real Pais-Uhlenbeck variable is x in the Lagrangian $L = -\frac{g}{2}\ddot{x}^2 + \frac{1}{2}\dot{x}^2 - \frac{1}{2}w^2x^2$ instead of x_1 and x_2. This should emphasize how grave the instability is, given that the same variable creates both, positive energy as well as negative energy modes. This becomes clearer in quantum theory, where the variable can be written in terms of two pairs of creation and annihilation operators. One pair for negative energy, and the other for positive energy. However, this is further specified in the next section 3.2.2 and fully developed in 4 where a free Pais-Uhlenbeck quantum field is quantized.

- *The same variable creates both, positive energy as well as negative energy modes.*

Another way to characterize the instability, is, as pointed out in [6],

- *The Ostrogradsky instability is a kinetic instability with an arbitrarily fast time scale.*

The latter comes from the fact that the number of microstates in the energy hypersurface does not strictly depend on the parameters of the interaction but on the only fact that there exists an interaction. Then, the instability seems to have an ”arbitrarily fast time scale” [6].

3.2.2 Restricted implications of the instability to quantum theory

As has been pointed out previously, some implications of the instability that apply to classical theory, have a direct counterpart in quantum theory, where as will become apparent, are somewhat more natural to interpret. Let us start with those previously discussed.

- *The vacuum decays into an even growing number of particles with positive and negative energies. All this is driven by entropy production. This is exclusive to interacting theories which couple the negative and positive sectors, allowing for the exchange of energy [3].*

- *The P-U operator is x instead of x_1 and x_2 separately, and then, it can be written in terms of a pair of creation and annihilation operators. One of them creating particles of negative energy, while the other pair, creates particles of positive energy.*

  The latter implies that for a local interaction, the two sectors (+ and - energy) become coupled, leading to the instability [3].

Given that a vacuum would decay in an un-stabilized HTDt, it is not clear how to define it. As the vacuum decays in many different states of many postive particles with their counterpart, negative energy particles, while keeping the reference energy level at 0, it could be said that the vacuum is not unique, which is not allowed in quantum theory. This will be one of the problems to tackle when stabilizing HTDt.

On the other hand, let us see what happens if a vacuum is naively defined for the Pais-Uhlenbeck quantum oscillator [2, 3, 7], that was defined in section 3.1.

- Consider a Fock space built by the direct sum of the subspaces of $\hat{x}_1$ and $\hat{x}_2$, *i.e.* by the pairs $\hat{b}_1^\dagger$, $\hat{b}_1$ and $\hat{b}_2^\dagger$, $\hat{b}_2$ respectively. Then, let us define the vacuum, $|\Omega\rangle$, as that annihilated independently by $\hat{b}_1$ and $\hat{b}_2$; this is:

$$\hat{b}_1|\Omega\rangle = 0 \qquad \hat{b}_2|\Omega\rangle = 0 \tag{21}$$

In section 3.1 the solution $x(t)$ was written as $C_+\cos(k_+t) + S_+\sin(k_+t) + C_-\cos(k_-t) + S_-\sin(k_-t)$, now, promoting it to an operator, writing it in terms of plane waves, and identifying the creation and annihilation operators, the following is obtained [3, 7, 16]:

$$\hat{x}(t) = \frac{1}{2}\left(\hat{b}_1 e^{-ik_1t} + \hat{b}_1^\dagger e^{ik_1t} + \hat{b}_2 e^{ik_2t} + \hat{b}_2^\dagger e^{-ik_2t}\right) \tag{22}$$

where the frequencies $k_{1,2}$ where defined in 11 as $k_{+,-}$ respectively. Note that for the negative part of the P-U (subscript 2), the creation operator $\hat{b}_2^\dagger$ because the plane wave with k_2 carries negative energy modes. $\hat{b}_1 \propto \hat{C}_+ + i\hat{S}_+$ and $\hat{b}_2 \propto \hat{C}_- - i\hat{S}_-$. Given the four pieces of initial data $(x_0, \dot{x}_0, \ddot{x}_0, \dddot{x}_0)$, the coefficients are:

$$\begin{aligned} \hat{C}_+ &= \frac{k_2^2\hat{x}_0 + \ddot{\hat{x}}_0}{k_2^2 - k_1^2} \qquad & \hat{S}_+ &= \frac{k_2^2\dot{\hat{x}}_0 + \dddot{\hat{x}}_0}{k_1(k_2^2 - k_1^2)} \\ \hat{C}_- &= -\frac{k_1^2\hat{x}_0 + \ddot{\hat{x}}_0}{k_2^2 - k_1^2} \qquad & \hat{S}_- &= -\frac{k_1^2\dot{\hat{x}}_0 + \dddot{\hat{x}}_0}{k_2(k_2^2 - k_1^2)} \end{aligned} \tag{23}$$

This, written in terms of the canonical variables (4), $X_1 = \hat{x}$, $X_2 = \dot{\hat{x}}$, $P_1 = \dot{\hat{x}} + \frac{\epsilon}{w^2}\dddot{\hat{x}}$ and $P_2 = -\frac{\epsilon}{w^2}\ddot{\hat{x}}$, takes the form:

$$\begin{aligned} \hat{b}_1 &\propto (1 + \sqrt{1-4\epsilon})\frac{k_1}{2}X_1 - (1 - \sqrt{1-4\epsilon})\frac{i}{2}X_2 + iP_1 - k_1P_2 \\ \hat{b}_2 &\propto (1 - \sqrt{1-4\epsilon})\frac{k_2}{2}X_1 + (1 + \sqrt{1-4\epsilon})\frac{i}{2}X_2 - iP_1 - k_2P_2 \end{aligned} \tag{24}$$

and identifying $P_i = -i\hbar\frac{\partial}{\partial X_i}$, there appears in fact an unique solution [3, 7, 16],

$$\Omega(X_1, X_2) = Ne^{\left(-\frac{\sqrt{1-4\epsilon}}{2\hbar(k_1+k_2)}(k_1k_2X_1^2 + X_2^2) - \frac{i\sqrt{\epsilon}}{\hbar}X_1X_2\right)} \tag{25}$$

This wave function can be normalized and therefore the complete set of normalized states can be built, but the Hamiltonian has eigenvalues carrying negative energy (With the usual definition of number operator $\hat{N}_i = \hat{b}_i^\dagger \hat{b}_i$)

$$\hat{H}|N_1, N_2\rangle = \hbar((N_1 + \frac{1}{2})k_1 - (N_2 + \frac{1}{2})k_2)|N_1, N_2\rangle \tag{26}$$

for positive norm states,

$$|N_1, N_2\rangle = \frac{1}{\sqrt{N_1! N_2!}}(\hat{b}_1^\dagger)^{N_1}(\hat{b}_2^\dagger)^{N_2}|0,0\rangle \tag{27}$$

This is, therefore, the consequence of defining an apparent vacuum in quantum theory for a HTDt, an ill defined energy spectrum. Even if the negative energy is not observable, upon interaction the two sectors would couple, giving rise to the vacuum instability. To emphasize the latter, consider the case of "a vacuum state". One expects it to be the lowest energy state, *e.g.* to have 0 energy, but there are infinite possibilities to obtain that result, then a vacuum would be degenerate, or also, a lower energy state can always be found.

On the other hand, as [3, 7] point out, there is a way to bypass the ill defined energy spectrum problem. Nevertheless, it entails another difficulty to the theory, which is perhaps another interpretation of the Ostrogradskian instability in quantum theory: lost of unitarity.

- Consider a vacuum redefinition. It will be, from now on in this example, the state that is independently annihilated by the annihilation operator of the positive energy sector, and by the creation operator of negative particles; this is

$$\hat{b}_1|\Omega\rangle = 0 \qquad \hat{b}_2^\dagger|\Omega\rangle = 0 \tag{28}$$

following the same procedure sketched above, an unique non-normalizable vacuum wave function is found [3, 7, 16]. If the Fock space is built upon this new vacuum by means of ((29)), the states are also non-normalizable. However, the energy spectrum turns out to be "well" defined, because it is positive defined, $\hat{H}|N_1, N_2\rangle = \hbar((N_1 + \frac{1}{2})k_1 + (N_2 + \frac{1}{2})k_2)|N_1, N_2\rangle$. This seems promising but the redefined sector has to be examined more carefully. Since the vacuum has been redefined by (28), the negative particle-sector of the Fock space, is built by repeated action of $\hat{b}_2$ instead of $\hat{b}_2^\dagger$; *i.e.*,

$$|N_1, N_2\rangle = \frac{1}{\sqrt{N_1! N_2!}}(\hat{b}_1^\dagger)^{N_1}(\hat{b}_2)^{N_2}|\Omega\rangle \tag{29}$$

Thus, for instance, the state $|N_1, 1\rangle = \frac{1}{\sqrt{N_1!}}(\hat{b}_1^\dagger)^{N_1}\hat{b}_2|0,0\rangle$, has the following norm

$$\begin{aligned}
\langle N_1, 1|N_1, 1\rangle &= \langle N_1, 0|\hat{b}_2^\dagger \hat{b}_2|N_1, 0\rangle \\
\langle N_1, 1|N_1, 1\rangle &= (\langle N_1, 0|\hat{b}_2 \hat{b}_2^\dagger|N_1, 0\rangle - \langle N_1, 0|[\hat{b}_2, \hat{b}_2^\dagger]|N_1, 0\rangle) \\
\langle N_1, 1|N_1, 1\rangle &= -\langle N_1, 0|N_1, 0\rangle
\end{aligned} \tag{30}$$

where $[\hat{b}_2, \hat{b}_2^\dagger] = 1$ has been used, as well as the fact that $\hat{b}_2^\dagger$ annihilates the redefined vacuum (28).

These are called ghost states [2, 3] and must not be considered in a quantum theory due to the impossibility of a probabilistic interpretation. Therefore, if one were to project them out of the Fock space, it would lead to non-unitary scattering processes, as is pointed out in [2, 3, 6, 7, 9, 10, 25, 29, 40], because as shown above, the scattering process couples the positive and negative energy sectors, or in other words, nothing avoids to compute the inner product between well defined positive-norm and a ill negative-norm states in the theory.

The latter remarks can be summarized in the following statement:

- *The Ostrogradskian instability leads, in a quantum theory, to an ill defined energy spectrum, or equivalently, to ghost states of negative norm and lost of unitarity [2, 3, 6, 7, 9, 10, 25, 29, 40].*

3.3 Possible ways to stabilize these theories

The main research on higher-time derivative theories has addressed specific HTDT identifying the parameter space in which these restricted cases turn out to be stable, but without considering the reason for such stability. Therefore, if all these cases share common properties is not yet known [6, 7, 8, 9, 10, 11, 12, 13]. Nevertheless, it seems that the only way for a HTDt to be stable in the sense of Ostrogradsky is when the Lagrangian is degenerate, which is usually claimed to make it reducible to a lower derivative Lagrangian.

If the Lagrangian is non-degenerate it is possible to solve the phase space variables for $x^{(N)}$ (with $L = L(x, x^{(1)}, \ldots, x^{(N)})$). More specifically, it is possible to find a function $\mathcal{A} = \mathcal{A}(X_1, \ldots, X_N)$ such that $\frac{\partial L}{\partial X^{(N)}}\Big|_{x^{(i-1)}=X_i, x^{(N)}=\mathcal{A}} = P_N$ [2, 3]. With this conditions, L does not give just 2nd order equations of motion as the well known case of first time derivatives in the Lagrangian. However, there is a subtlety in this analysis. On this regard, Motohashi *et. al* [39] and Klein [40] have separately found very general rules, that rely on degeneracy, for Lagrangians with multiple and finitely many DOFs that interact in a very particular way, which guarantee the associated theories to be stable and not trivially reducible to lower derivative theories. In the next section the general idea is shown.

3.3.1 Conditions on fDOF Lagrangians that guarantee stability in the sense of Ostrogradsky

As was previously discussed, there have been established very general conditions to be imposed on Lagrangians with a finite number of DOF and higher-time derivatives (HTDt -fDOF), that guarantee the stability in the sense of Ostrogradsky. Here, the approach followed by Motohashi *et. al* [39] is to be presented, even though these conditions have also been reached in a different way by Klein and Roest [40].

In general, the Lagrangians to be considered are of the following form:

$$L = L(\ddot{\phi}^a, \dot{\phi}^a, \phi^a; \dot{q}^i, q^1) \tag{31}$$

where $a = 1, \ldots, n$ and $i = 1, \ldots, m$, and $\dot{\phi} = \frac{d\phi}{dt}$. It must be noted that a conclusion in [39], is that every system containing any number of just "problematic" variables, here denoted by ϕ^a, without any "healthy variables" (q^a), is unstable if it is non-degenerate. Thus, all the theories to be considered here will have at least one healthy variable.

- Let us first consider the simplest case of just one healthy variable and one problematic [39].

$$L^{\times} = L^{\times}(\ddot{\phi}, \dot{\phi}, \phi; \dot{q}, q) \tag{32}$$

Furthermore, let us define the following constraint $Q = \dot{\phi}$, which will make the things easier to extend to more DOF, and furthermore, allows the theory to be interpreted as a usual Lagrangian with constraints. Correspondingly, the Lagrangian will be written (with $\dot{\lambda} = 0$):

$$L(\dot{Q}, Q, \phi; \dot{q}, q; \lambda, \dot{\lambda}) = L(\dot{Q}, Q, \phi; \dot{q}, q) + \lambda(\dot{\phi} - Q) \tag{33}$$

Now, recalling that the Ostrogradskian Instability can be identified in the linearity of the momenta in the Hamiltonian after the canonical transformation has been done, it is better to begin with the Hamiltonian analysis:

  – The canonical momenta are:

$$P = L_{\dot{Q}} \qquad p = L_{\dot{q}} \qquad \pi = L_{\dot{\phi}} = \lambda \qquad \rho = L_{\dot{\lambda}} = 0 \tag{34}$$

From the last two equations, we get two primary constraints which only arise because of the way we have rewritten the Lagrangian and therefore include no further physical meaning.

$$\Phi = \pi - \lambda \approx 0 \qquad \Psi = \rho \approx 0 \tag{35}$$

where the $\approx$ symbol means weak equality; i.e. the constraint is restricted to be zero only on a certain hypersurface of the phase space [46]. The latter also implies that $\{\Phi, \Psi\} = -1$.

Now, the core of the stabilization procedure can be understood as demanding a constraint between momenta of one of two degrees of freedom propagated by the higher derivative variable[7] and the healthy additional variable in the Lagrangian. Given (33), with Q being a variable related to the higher derivative one (ϕ) and q, the healthy DOF, it is natural to consider an infinitesimal variation of momenta associated to them and impose the constraint described above [39], but before, let us motivate this new constraint:

$$\begin{pmatrix} \delta P \\ \delta p \end{pmatrix} = \begin{pmatrix} L_{\dot{Q}\dot{Q}} & L_{\dot{Q}\dot{q}} \\ L_{\dot{Q}\dot{Q}} & L_{\dot{q}\dot{q}} \end{pmatrix} \begin{pmatrix} \delta \dot{Q} \\ \delta \dot{q} \end{pmatrix} \tag{36}$$

[7] In quantum field theory, this can be easiy seen by computing the Feynman propagator of, for instance, a higher derivative (HD) scalar field, which can be seen to propagate at least two DOFs, one of them ghost like, due to a "wrong" sign in the propagator in momentum space (see [26]). In classical mechanics, let it be field theory or with finitely many DOFs, this can be seen by rewriting the Lagrangian for one HD variable, in terms of two different low derivative variables, one of them with the "wrong" signs. This was shown with the P-U Lagrangian, as an example, in section 3.1.

Where the resultant matrix is the kinetic matrix K. It is possible to see that if this matrix is not singular, one can solve for $\dot{Q}$ and $\dot{q}$ in terms of P and p.

By the Legendre transform, the Hamiltonian[8] is:

$$H(P,p,\pi,Q,q,\phi,\mu,\nu) = \pi Q + P\dot{Q}(P,p,\dots) + p\dot{q}(P,p,\dots) - L(\dot{Q}(P,p,\pi,\dots),\dot{q}(P,p,\pi,\dots),\phi,Q,q) + \mu\Phi + \nu\Psi \tag{37}$$

Following the Dirac programme, which will be introduced for more general cases in the following section, we demand the constraints to be conserved; i.e.

$$\frac{d\Phi}{dt} \stackrel{!}{=} 0 \qquad \frac{d\Psi}{dt} = \{\Psi, H\} = \{\Psi, \pi Q + P\dot{Q} + p\dot{q} - L + \mu\Phi + \nu\Psi\} \stackrel{!}{=} 0 \tag{38}$$

$$\{\Psi, H\} = \{\Psi, \mu\Phi\} = -\mu \rightarrow \mu = -\{\Psi, H\} \stackrel{!}{=} 0 \tag{39}$$

similarly $\nu = \{\Phi, H\} = 0$, and thus, all the Lagrange multipliers are left fixed and there are no secondary constraints, which permits to see that there is a linear term in the momenta in (37), πQ, because $Q = \dot{\phi} \neq \dot{\phi}(\pi)$. This is the ghost-like term that gives rise to the instability (see section 3.2.1) [2, 3, 39]. Thus, it has been shown that the theory without extra constraints has the Ostrogradskian instability.

Now, it is evident that everything is solved if Q can be written in terms of π. Furthermore, let us note that the terms $P\dot{Q}(P,p,\dots)$, $p\dot{q}(P,p,\dots)$ do not contribute to the instability, because they, in principle, can be bounded in phase space[9]. This can only be true, because we identified that the kinetic matrix K in (36) is not singular and therefore it is possible to solve for $\dot{Q}$ and $\dot{q}$ in terms of P and p and other canonical variables, but not π, which leaves us with no possibility to find some $Q = Q(\pi, P, p, \dots)$. Thus, the condition for stabilization is to impose an additional constraint in phase space between the canonical momenta[10] that avoids to simply invert Q in terms of $(P, p, \dots)$ but that also includes π. Let us write such constraint and see which conditions have to be imposed:

$$\Xi \equiv P - F(p, Q, \phi, q) \approx 0 \tag{40}$$

With this constraint the Hamiltonian can be rewritten with an additional Lagrange multiplier, and demanding again stability under time evolution,

$$\dot{\Phi} \approx 0 \qquad \dot{\Psi} = \mu \approx 0 \qquad \dot{\Xi} = \{\Xi, H\} - \mu F_\phi \approx 0$$

[8] It is worth emphasizing that if the kinetic matrix is not invertible, the construction of the Hamiltonian in terms of canonical variables is not possible and an "Extended Hamiltonian Formalism" [45] is required.

[9] The fact that the function turns out to be unbounded on P or p, $P\dot{Q}(P,p,\dots)$, $p\dot{q}(P,p,\dots)$ in the Hamiltonian and consequently be still unstable, is no longer considered as the Ostrogradskian instability, because this kind of instability relies on the special form of the Lagrangian, and may not be related to higher-order time derivatives.

[10] This constraint cannot be defined globally [39].

Finally finding:

$$\pi \approx \{\Xi, H_0\} - F_\phi Q \quad (41)$$

Where evidently, the canonical momenta π is written in terms of the other variables, in particular Q, and it can be inverted upon obvious conditions on F, being $F_\phi =: \frac{\partial F}{\partial \phi}$ and H_0 is the Hamiltonian without any constraint nor Lagrange multipliers. Notice the importance of the $\approx$ weak equality symbol that does not allow to set μ to zero, before computing all the time evolution (For more details, see Dirac's Programme below, or [44, 45, 46, 47]).

This turns out to eliminate the linearity in the Hamiltonian, therefore eliminating the Ostrogradskian instability [39, 40]. The latter can be interpreted if one recalls that the original instability came from the fact that there were infinite possible microstates compatible with the energy fixed macrostate, which defined a hypersurface. Thus, if an additional constraint is added, it is easy to expect that the number of possible microstates that the system can occupy is reduced.

It is also possible to verify that the previously imposed constraint implies [39]:

$$det(K) = L_{\dot{Q}\dot{Q}} L_{\dot{q}\dot{q}} - L^2_{\dot{q}\dot{Q}} = 0 \quad (42)$$

Therefore, the condition imposed by Motohashi *et. al* to avoid the Ostrogradskian instability is: there must exist a constraint between the canonical momenta $\Xi \equiv P - F(p, Q, \phi, q) \approx 0$, or equivalently $det(K) = 0$.

- Now, considering the most general case[39]

$$L = L(\ddot{\phi}^a, \dot{\phi}^a, \phi^a; \dot{q}^i, q^1)$$

with $a = 1, \ldots, n$ and $i = 1, \ldots, m$, i.e., n problematic variables and m ordinary ones.

Now, the conditions are deduced in a very similar way but being aware of some subtleties when the constraint analysis is developed. The aforementioned imposition of a constraint between canonical momenta, here extends to as many constraints as the number of unstable DOF the Lagrangian includes (n). Thus, the first n conditions are stated as:

$$\Xi_a \equiv P_a - F_a(p_i, q^i, Q^b, \phi^b) \approx 0 \quad (43)$$

but here, when demanding the time invariance of the constraints, the following relation is obtained[11]:

$$\dot{\Xi}_a = \{\Xi_a, H\} + \xi^b \{\Xi_a, \Xi_b\} \approx 0 \quad (44)$$

As pointed out in [39], if the matrix whose components are the Poisson brackets between the primary imposed constraints $\{\Xi_a, \Xi_b\}$ is invertible, then all the Lagrange multipliers in the

[11] This is very similar to the equation obtained for the simpler case, but here, there are in fact n time invariance requirements in matrix form

Hamiltonian are left fixed and there are not enough secondary constraints to eliminate the linear momenta. Thus a necessary condition would be that

$$det(\{\Xi_a, \Xi_b\}) = 0 \tag{45}$$

but, a sufficient, though restrictive condition, would be

$$\{\Xi_a, \Xi_b\} = 0 \tag{46}$$

Therefore, a general approach to evade Ostrogradskian instability when there is a finite number of unstable (n) and stable (m) DOF, is to impose n primary constraints between the canonical momenta (Ξ_a) and also demanding that $det(\{\Xi_a, \Xi_b\}) = 0$ [39].

3.4 Canonical quantization with constraints

In section 3.3, the current advances devised in [39, 40, 41] to define a general method, to construct stable HTDt, were shown. It was there emphasized, that the main step of the stabilization procedure, is to couple the HTDt to another low-order derivative degree of freedom. Therefore, it is customary to address the quantization procedure keeping in mind, that the systems are by necessity constrained.

Now, given that the quantization scheme to be employed in this work is the canonical, the correct definition of the Hamiltonian in the constrained system, is of considerable importance. Thus, in this section, a brief review on the quantization of constrained systems is presented. The general purpose will be to give definitions and notation for further sections. The initial part will be devoted to general definition of singular theories to be used later, as well as theories with first and second class constraints, the Dirac program, and some theorems which will also be of use later in section 4, when the interpretation of HTDt is considered.

This section of the general review is based on [44, 45, 46, 47, 48, 49, 50, 51, 52], and is by no means complete.

3.4.1 Singular theories

Let us start by defining the Hessian, by means of which singular and non-singular theories are classified.

Definition 3.3 (Hessian Matrix - Hessian) *The Hessian or kinetic matrix is defined as [45]:*

$$M_{ij} = \frac{\partial^2 L}{\partial \dot{x}^i \partial \dot{x}^j} \tag{47}$$

Where L is the Lagrangian, that possibly depends on $x^i, i = 1, \ldots, n$ with $n \in \mathbb{N}$, and on its first time derivatives[12]

[12] This definition will be extended to higher derivatives.

The Hessian is the determinant of the Hessian matrix:

$$Hessian = det|M_{ij}| \tag{48}$$

Now, theories can be classified as singular and nonsingular.

Definition 3.4 (Singular theories) *The theory is singular when the Hessian is zero and non-singular otherwise [45].*

According to the stabilization procedure depicted in section 3.3, it is clear that stable HTDt are singular. This poses a major problem when "Hamiltonizing" the theory because the Legendre transform cannot be well defined in terms of only conjugate momenta and generalized variables. To see this, consider for instance the definition of conjugate momenta $p_i = \frac{\partial L}{\partial \dot{x}^i}$ and the Hamiltonian given by the Legendre transform. In order to have the Hamiltonian completely written in terms of canonical variables, it must be possible to invert the previous relation to give $\dot{x}^i$ in terms of the canonical variables. If the Hessian is zero, M_{ij} is not invertible, *i.e.* $\frac{\partial p_i}{\partial \dot{x}^j}$ is not invertible, which equals to the impossibility of uniquely expressing $\dot{x}_i$ in terms of the canonical variables. To put in clearer terms, let us consider a different form of the Euler-Lagrange equations:

$$\begin{aligned}
0 &= \frac{\partial L(x,\dot{x})}{\partial x^i} - \frac{d}{dt}\frac{\partial L(x,\dot{x})}{\partial \dot{x}^i} \\
0 &= \frac{\partial L(x,\dot{x})}{\partial x^i} - \frac{\partial}{\partial \dot{x}^j}\left(\frac{\partial L(x,\dot{x})}{\partial \dot{x}^i}\right)\ddot{x}^j - \frac{\partial}{\partial x^j}\left(\frac{\partial L(x,\dot{x})}{\partial \dot{x}^i}\right)\dot{x}^j \\
0 &= \frac{\partial L}{\partial x^i} - \left(\frac{\partial^2 L}{\partial \dot{x}^j \dot{x}^i}\right)\ddot{x}^j - \frac{\partial^2 L}{\partial x^j \dot{x}^i}\dot{x}^j \\
M_{ij}\ddot{x}^j &= \frac{\partial L}{\partial x^i} - \frac{\partial^2 L}{\partial x^j \dot{x}^i}\dot{x}^j
\end{aligned} \tag{49}$$

From equation (49), it is clear that if the Hessian is zero, *i.e.* M_{ij} is not invertible, the equations of motion for x^j cannot be uniquely expressed as $\ddot{x}^j = f(x^j, \dot{x}^j)$. To summarize the previous analysis, it is worth to put the consequences as pointed out in [45]:

Theorem 3.2 ("Hamiltonization" and Cauchy problem for Singular theories[45]) *- It is not possible to go to the Hamiltonian formalism in the standard way, when the theory is singular.*

- The Euler-Lagrange equations of motion for singular theories, provided with initial conditions, do not have unique solutions. The problem can be solved for certain class of constraints, by imposing their conservation. (The distinction of these cases is made explicit below.)

Now, it is evident that for canonical quantization, something must be done in order to correctly go from the Lagrangian to the Hamiltonian formalism. One possible way of doing this has been formally established in a simple way. The whole idea is to address the core of the problem, which goes back to the impossibility of inverting $\dot{x}^i$ in terms of x^i, p^i. Thus, the simple solution starts by allowing the existance of some "primarly unexpressible velocities"($v^i = \dot{x}^i$) which cannot be inverted in terms of x^i, p^i, in the Hamiltonian formalism. This must be understood as formal extension, and therefore, is usually called "extended Hamiltonian formalism" [45].

In order to cope with singular theories (which shall be at the main point of the stabilization of HTDt), the new formalism must include $3n$ equations that depend on the set of x^i, v^i, p^i, with $i = 1, \dots, 3n$ ($n \in \mathbb{N}$). The notation to be used in section 4 will be based on that followed in [45], and therefore, it is briefly developed in what follows.

Definition 3.5 (Extended Hamiltonian formalism [45]) *First consider a new Lagrangian L^v, which instead of depending on $\dot{x}^i$, depends on v^i, by means of the definition,*

$$v^i = \dot{x}^i \tag{50}$$

this is, $L^v = L^v(x, v)$[13]*. The corresponding Lagrange equations would be derived from the action $S = \int dt L^v + p_i(\dot{x}^i - v^i)$. The Hessian Matrix is automatically redefined to M^v*

The extended Hamiltonian will be defined by the function[14]

$$H^*(x, v, p) = p_i v^i - L^v(x, v) \tag{51}$$

With this definition, the Hamilton's equations are:

$$\dot{p}_i = -\frac{\partial H^*}{\partial x^i} \quad v^i = \frac{\partial H^*}{\partial p_i} \quad \frac{\partial H^*}{\partial v^i} = 0 \tag{52}$$

Now, just as in the case of HTDt (See the example about the Pais-Uhlenbeck oscillator, given in section 3.1), there are in general some primarly expressible and unexpressible velocities. It is important to denote them clearly, therefore, let us adopt the following notation.

Given the Hessian matrix of a singular theory, its rank will be denoted $R = rank|M_{ij}|$. As usual, the rank theorem gives $n = R + nul$, where n is the number of degrees of freedom as defined above, and Nul is the nullity. It must be noted that the formalism to be reviewed [45], is constructed upon the free part of the Lagrangian, *i.e.* the L_{int} part of $L = L_o + L_{int}$ is not considered in the Hessian matrix.

In order to clearly identify the group of primarly expressible and unexpressible velocities, let us order the variables in such a way that the minor of maximum rank, is on the top left corner of the Hessian matrix. In this way, the groups of variables are:

$$\begin{array}{cccc} Q^\mu = x^\mu & \Pi_\mu = p_\mu & \Lambda^\mu = v^\mu & \mu = 1, \dots, R \\ q^\alpha = x^{R+\alpha} & \pi_\alpha = p_{R+\alpha} & \lambda^\alpha = v^{R+\alpha} & \alpha = 1, \dots, Nul \end{array} \tag{53}$$

Where the determinant of the sub-Hessian matrix of the $\dot{Q}$ or Λ is $\neq 0$; i.e., the Λ are the primarly expressible velocities. It is possible, by definition, to invert $\Lambda = \Lambda(x, \Pi, \lambda)$, and whenever an expression is expressed on terms of Λ and Π, or equivalently, when the λ are expressed in terms of

[13] x, v, p, denote all the x^i, v^i, p^i, with $i = 1, \dots, 3n$ and $n \in \mathbb{N}$

[14] The usual summation convention over repated indices, will be used.

the latter, the expression will be distinguished with a bar upon it. Now, consider the third set of Hamilton's equation for the primarly unexpressible velocities (λ), given in equation (52) and let us employ the Λ to rewrite the λ.

$$\begin{aligned} \frac{\partial H^*}{\partial \lambda}|_{\Lambda=\Lambda(x,\Pi,\lambda)} &= 0 \\ \frac{\partial H^*}{\partial \lambda}|_{\Lambda=\Lambda(x,\Pi,\lambda)} = \pi_\alpha - \frac{\partial L}{\partial \lambda^\alpha}|_{\Lambda=\Lambda(x,\Pi,\lambda)} &= 0 \\ \Phi^{(1)}_\alpha = \pi_\alpha - \frac{\partial \bar{L}}{\partial \lambda^\alpha}(x,\Pi) &= 0 \end{aligned} \quad (54)$$

Where equation (54) establishes a relation between conjugate momenta π and generalized coordinates x, or in other words, (54) are constraints, which will be called $\Phi^{(1)}_\alpha$. The superscript (1) denotes that these are primary constraints, because they are implied by the extended Hamilton's equations of motion (52). It is important to note that there are as much constraints, as the dimension of the null space of the Hessian matrix. Equivalently, there are as much vectors in the basis of the null space, as constraints, and therefore, it would be easy to show that the constraints must be independent. This is why they are identified with the subscript $\alpha = 1, \ldots, Nul$.

Now, it is possible to eliminate all the primarly unexpressible velocities λ form the extended Hamiltonian (51), and therefore, it can be written as,

$$\begin{aligned} H^{(1)} &= H^*|_{\Lambda=\Lambda(x,\Pi,\lambda)} = H(x,p,\Lambda) + \left(v^i \frac{\partial H^*}{\partial v^i}\right)_{\Lambda=\Lambda(x,\Pi,\lambda)} \\ H^{(1)} &= H + \lambda^\alpha \Phi^{(1)}_\alpha \end{aligned} \quad (55)$$

where it is evident that the primarly unexpressible velocities are Lagrange multipliers as coefficients to the corresponding constraints $\Phi^{(1)}_\alpha$.The previous analysis can be summarized as follows:

Remarks 3.1 (Singular theories and constraints. Hamiltonian with constraints [45]) *- A singular theory is characterized by a Hessian matrix (M_{ij}) whose determinant is zero. The latter implies that there are exactly Nul (Nulllity) constraints $\Phi^{(1)}_\alpha$ in the extended Hamiltonian formalism, which are independent. In other words, constraints are included in the equations of motion of a singular theory (The remark is equally valid for all classes of constraints. These are defined in the next section).*

- The Hamiltonian system of equations with primary constraints, is[15]:

$$\begin{array}{cc} \dot{\eta} = \{\eta, H^{(1)}\} & \Phi(\eta) \approx 0 \\ H^{(1)} = H(\eta) + \lambda^\alpha \Phi_\alpha(\eta) & \alpha = 1, \ldots, Nullity \end{array} \quad (56)$$

Where η denotes all the possible canonical variables (x,p), Φ denotes all the possible α constraints as well as the different stage constraints (That so far are only primary-stage, but will be later expanded to secondary constraints in section 3.4.2), and the $\approx$ denotes weak equality, which stands for an equality that is only fulfilled in the corresponding hypersurface in phase space.

[15]Note that there is a big difference between primary and second class constraints, in comparison with primary and secondary constraints. They are going to be further specified in next section.

3.4.2 Dirac's programme

Based on the definitions and notation given in the previous section, it is straightforward to give a brief review of the Dirac's programme [44, 45, 46, 47]. This will be amply used when building some stable higher derivative field theories in section 4 and also to analyse the procedure. In particular, it will become apparent that once a HTDT is brought to the point where the Dirac's program can be applied, the theory is indeed completely healthy in the sense of Ostrogradsky.

In this work, the constraints that make a HTDT singular, are those which are time independent. With this on mind, we shall be restricted to the following:

$$\dot{\Phi}^{(1)}_{\alpha} = \{\Phi^{(1)}_{\alpha}, H^{(1)}\} = 0 \tag{57}$$

Substituting the definition for $H^{(1)}$ given in 3.1, recalling that the summations convention is being employed, the following is obatined:

$$\dot{\Phi}^{(1)}_{\alpha} = \{\Phi^{(1)}_{\alpha}, H^{(1)}\} + \{\Phi^{(1)}_{\alpha}, \Phi^{(1)}_{\beta}\}\lambda^{\beta} = 0 \tag{58}$$

From (58) it is evident, that if the matrix composed of the Poisson brackets between the primary constraints, $\{\Phi^{(1)}_{\alpha}, \Phi^{(1)}_{\beta}\}$, is invertible, the primary unexpressible velocities (that have become recognisable as Lagrange multipliers), can be solved in terms of the other canonical variables, and therefore, the equations of motion can be fully determined given the following Hamiltonian:

$$\begin{aligned} H^{(1)} &= H + \lambda^{\alpha}\Phi^{(1)}_{\alpha} \\ H^{(1)} &= H - \{\Phi^{(1)\alpha}, \Phi^{(1)}_{\beta}\}^{-1}\{\Phi^{(1)}_{\beta}, H^{(1)}\}\Phi^{(1)}_{\alpha} \end{aligned} \tag{59}$$

If the latter is not the case and $\{\Phi^{(1)}_{\alpha}, \Phi^{(1)}_{\beta}\}$ is not invertible, it is not possible to solve for λ^{β}, which indicates that the procedure is not complete, because the theory is not fully determined. In order to proceed, one must notice that $rank|\{\Phi^{(1)}_{\alpha}, \Phi^{(1)}_{\beta}\}|_{\Phi^{(1)}=0} = \rho \neq n$ and therefore, drawing upon the Rank theorem, there must exist as many linearly independent vectors $u^{\alpha}_{(k)}$, as $Nul_2 = n - \rho$, which is the nullity of the matrix being considered. These vectors belong to the null space of the matrix and therefore fulfill:

$$\{\Phi^{(1)}_{\alpha}, \Phi^{(1)}_{\beta}\}|_{\Phi^{(1)}=0} u^{\alpha}_{(k)} = 0 \tag{60}$$

This last equation, applied to (58), $\{\Phi^{(1)}_{\alpha}, H^{(1)}\}u^{\alpha}_{(k)} + \left(\{\Phi^{(1)}_{\alpha}, \Phi^{(1)}_{\beta}\}u^{\alpha}_{(k)}\right)\lambda^{\beta} = 0$, immediately gives:

$$\{\Phi^{(1)}_{\alpha}, H^{(1)}\}u^{\alpha}_{(k)} = 0 \tag{61}$$

which, in principle, give Nul_2 equations that relate x and p. As pointed out in [45], some of these equations may be identically satisfied, and therefore, do not provide new information. Nevertheless, there must be some relations that do impose new constraints between x and p. The resulting independent constraints are called secondary constraints and are denoted by $\Phi^{(2)}$. Then, the time independence $\dot{\Phi}^{(2)} = 0$ is again demanded, which can be enough to solve for the Lagrange multipliers λ. If again, the matrix of poisson brackets between the constraints is not invertible, some additional constraints result in exactly the same procedure described above. This procedure must be carried out up to the stage where the primarly unexpressible velocities (λ) are expressed in terms of x and p. All the new constraints that may arise if the procedure is continued, are equally called secondary constraints and denoted by $\Phi^{(2)}$. It is important to note that all the constraints that may result in this procedure, must be included in every step to the definition of the Hamiltonian given in equation (56). Furthermore, in order to be consistent with the notation, $H^{(1)} \to H^{(1,2)}$.

Before going on with further developments, some remarks are in order. First note that all the constraints are defined as $\Phi \approx 0$, which, as mentioned above, denotes that the constraints are only satisfied on the constraint hypersurface in phase space. When an expression is evaluated in $\Phi = 0$, it denotes that it is on the surface of constraints, or in other words, the expression is evaluated on a hypersurface in phase space, which is defined by the equations $\Phi = 0$ (Recalling that Φ denotes all the primary and secondary constraints).

3.4.3 First- and second-class constraints

Definition 3.6 (First- and second-class constraints [45, 44]) *- First-class constraints are those whose Poisson bracket with any other constraint, is zero on the constraint surface.*

- Second-class constraints are those which make part of a set of constraints, whose matrix $|\{\Phi, \Phi\}|_{\Phi=0}$ *is nonsingular*

It is evident that if a singular theory possesses only second-class constraints, all the primarly unexpressible velocities λ are uniquely defined. This can be easily seen from the fact, that at every stage, after demanding the conservation of constraints on time, the condition for the λ to be uniquely found, is that the matrix $\{\Phi_\alpha, \Phi_\beta\}$ is invertible. This is, all the constraints found until that stage, Φ, must be of second-class. If this is indeed the case, it can be shown that all the constraints can be directly added to the Hamiltonian $H^{(1,2)}$, because all the Lagrange multipliers λ are solvable. The Dirac matrix, which is the matrix whose entries are the Poisson brackets of all the constraints admissible by the theory, is obviously invertible in this case. The whole Hamiltonian in (56) is, therefore, defined as $H^{(1)} = H(\eta) + \{\Phi^\alpha, \Phi_\beta\}^{-1}\{\Phi_\beta, H^{(1)}\}\Phi_\alpha(\eta)$, and the equations of motion are given in terms of the Dirac Bracket $\dot{\eta} = \{\eta, H\}_{D(\Phi)}|_{\Phi=0}$.

It can be shown that a theory with only second-class constraints must exclusively contain primary second-class constraints. In other words, a theory with second-class constraints, only, can be completely defined by the Dirac's programme described above, because the primarly unexpressible velocities can be deduced through conservation of constraints in time. On the other hand, it is also possible to show that for this kind of theories there is the possibility to find a canonical transformation that changes the original canonical variables into another set (ω, Ω), which has two very well defined kinds of variables: physical ω and non-physical Ω, in the sense that the latter define the constraint surface by equations $\Omega = 0$. The latter group of canonical variables Ω, are the same in number as the original number of primary and secondary constraints (of all classes) for the original set of canonical variables. In summary, for theories with only second class constraints, it is possible to redefine the theory in such a way that the constraints simply reduce to demand, for some of them, to be strictly zero in phase space (This defines the constraint surface); i.e. they are non-dynamical.

In the next section, the previous review will be extended to the most interesting case of theories with first-class constraints, which is a possibility to make HTDt stable, but that has not been considered in this work.

3.4.4 Theories with first-class constraints

Consider the case where the Dirac matrix $|\{\Phi,\Phi\}|_{\Phi=0}$ is singular on the constraint surface. According to the definition given above (3.6), theories which admit the latter condition are defined as theories with first-class constraints. The study of the latter is quite pertinent to HTDt, because they must be singular and include this kind of constraints, in order to be stable, if higher than two derivatives in the Lagrangian, are included. This will also imply that these theories essentially contain a functional arbitrariness, *i.e.* they are not completely well defined, and therefore, the conditions derived from the definitions and theorems to be defined in brief, will be understood in section 4 as a motivation to write sensible HTDt.

For a theory of this kind, the nullity of the Dirac matrix will be defined as $Nul_3 = N_\Phi - R_\Phi$, where N_Φ is the combined number of primary and secondary constraints in the theory[16], or in other words, all the admissible constraints which result after demanding all the conservation of constraints in time. N_Φ is, therefore, the dimension of the Dirac Matrix. R_Φ is the rank of the Dirac matrix. Now, in order to continue with the analysis, let us consider the following statement, which is prooved in [45] and extends the possibilities of finding particuar sets of physical canonical variables for a theory with first-class constraints. This theorem will be of major importance for the results given in section (4), and is a generalization that also applies to theories with only second-class constraints (For further reference, see [44, 45, 46, 47].)

Remarks 3.2 (Classification of canonical variables [45]) *It can be shown that for a theory with its set $\{\Phi\}$ of all constraints, classified as primary $\Phi^{(1)}$ and secondary $\Phi^{(2)}$, the canonical variables $\eta = (x,p)$ can be transformed to the set $\eta = (\omega, Q, \Omega)$, with the following properties:*

- *There are as many canonical variables Ω, as number of constraints Φ.*
- *The set Ω, defines exactly the same constraint hypersurface in phase space, as $\{\Phi\}$ did in the original theory, $\eta = (x,p)$.*
- *The set of canonical variables Ω (which also define the constraint surface), can be further specified as $\Omega = (\mathcal{P}, \varphi)$, where $\mathcal{P}$ are the first-class constraints and φ are the second-class constraints.*

  Furthermore, the set $\eta = (\omega, Q, \Omega)$ can be endowed with more details, or equivalently, we reassure that there exist such a canonical transformation that also provide the following conditions upon the sets:

[16] Recall in the notation adopted here [45], this differs from first- and second-class constraints

- $\mathcal{P}$ *can be divided into two different sets of primary and secondary constraints* $\mathcal{P} = (\mathcal{P}^{(1)}, \mathcal{P}^{(2)})$. *All the variables* $\mathcal{P}$ *are the canonical momenta conjugated to* Q.
- φ *can also be divided into two different sets of primary and secondary constraints* $\varphi = (\varphi^{(1)}, \varphi^{(2)})$, *but they are not completely disconnected. The relation is the following:*
  * *The subsets* $\varphi^{(1)} = (v, \psi^{(1)})$ *and* $\varphi^{(2)} = (u, \psi^{(2)})$, *where* v *and* u *together, form a set of canonnically conjugated variables, while* $\psi^{(1)}$ *and* $\psi^{(2)}$ *form themselves, two sets of canonnically conjugated variables.*

- ω, *on itself, is a set of canonically conjugated variables.*

It is worth to emphasize that all the subdivisions of Ω *count as canonical variables and constraints as well, because all of them define the constraint surface. Therefore, the superscript* (1), (2) *only denotes that they are canonical variables that also define the constraint surface equivalent to that defined by* $\Phi^{(1)}$ *and* $\Phi^{(2)}$.*In other words,* $(\mathcal{P}^{(1)}, \varphi^{(1)}(v, \psi^{(1)}))$ *are equivalent to the primary constraints* $\Phi^{(1)}$ *and* $(\mathcal{P}^{(2)}, \varphi^{(2)}(u, \psi^{(2)}))$ *to the secondary constraints that appear at different stages after demanding the conservation of the primary constraints.*

Given the possibility to denote the canonical variables as stated in the remark (3.2), it is possible to write the Hamiltonian (56) in an useful way, by considering $\Omega^{(1)} = (\mathcal{P}^{(1)}, 0)$, adding the second constraints (which was shown possible in the previous section) and Taylor expanding it in the new canonical variables (ω, Q, Ω) around the constraint surface. Since the constraint surface is defined by $\Omega = 0$, the expansion to first order is:

$$\begin{aligned}
H^{(1)} &= H|_{\Phi=0} + \lambda_\alpha \Phi^\alpha|_{\Phi=0} + \mathcal{O}(\Phi^2) \\
&\downarrow \quad \eta(x,p) \to \eta(\omega, Q, \Omega) \\
H^{(1)} &= H|_{\Omega=0} + \lambda_{(1)}\Omega^{(1)}|_{\Omega=0} + \lambda_{(2)}\Phi^{(2)}|_{\Omega=0} + \mathcal{O}(\Omega^2) \\
H^{(1)} &= H|_{\Omega=0} + \lambda_{(1)}\Omega^{(1)}|_{\Omega=0} + A_{(2)}\mathcal{P}^{(2)}|_{\Omega=0} + B_{(2)}\varphi^{(2)}|_{\Omega=0} + \mathcal{O}(\Omega^2) \\
H^{(1)} &= H_{Phys} + \lambda_{(1)}\Omega^{(1)}|_{\Omega=0} + A_{(2)}\mathcal{P}^{(2)}|_{\Omega=0} + B_{(2)}C_{(1)}u|_{\Omega=0} + B_{(2)}C_{(2)}\psi^{(2)}|_{\Omega=0}
\end{aligned} \tag{62}$$

Where $H|_{\Omega=0}$, *i.e.* the original Hamiltonian of the theory without constraints, evaluated at the constraint surface, is defined as the *Physical Hamiltonian* H_{Phys}.

$$H_{Phys} = H|_{\Omega=0} \tag{63}$$

This is very similar to the procedure that was described in the previous section and therefore, the name given to this Hamiltonian, attends to important facts to be described below. Let us compute the following Poisson bracket, between the Hamiltonian expressed as in (62) and the First-class constraints $\mathcal{P}$, to clarify the last assertion:

$$
\begin{aligned}
\{H^{(1)}, \mathcal{P}\}|_{\Omega=0} &= \{H_{Phys} + \lambda_{(1)}\Phi^{(1)} + A_{(2)}\mathcal{P}^{(2)} + B_{(2)}C_{(1)}u + B_{(2)}C_{(2)}\psi^{(2)}, \mathcal{P}\}|_{\Omega=0} \\
\{H^{(1)}, \mathcal{P}\}|_{\Omega=0} &= \{H_{Phys}, \mathcal{P}\}|_{\Omega=0} + \{\lambda_{(1)}\Omega^{(1)} + A_{(2)}\mathcal{P}^{(2)} + B_{(2)}C_{(1)}u + B_{(2)}C_{(2)}\psi^{(2)}, \mathcal{P}\}|_{\Omega=0} \\
\{H^{(1)}, \mathcal{P}\}|_{\Omega=0} &= \{H_{Phys}, \mathcal{P}\}|_{\Omega=0} + \{\Omega\}|_{\Omega=0} = \{H_{Phys}, \mathcal{P}\} + \{\Omega\}|_{\Omega=0} \\
\{H^{(1)}, \mathcal{P}\}|_{\Omega=0} &= \frac{\partial H_{Phys}}{\partial Q}\frac{\partial \mathcal{P}}{\partial \mathcal{P}} - \frac{\partial H_{Phys}}{\partial P}\frac{\partial \mathcal{P}}{\partial \mathcal{Q}} = \frac{\partial H_{Phys}}{\partial Q} + \{\Omega\}|_{\Omega=0} \\
\{H^{(1)}, \mathcal{P}\}|_{\Omega=0} &= 0 \qquad (64)
\end{aligned}
$$

Where the following facts were used: from the remark (3.2), the canonical conjugated momenta to Q, are $\mathcal{P}$, the conservation of $\mathcal{P}$ constraints was demanded, and on the other hand, the physical Hamiltonian does not depend on the canonical variables Ω, because by definition, it is on the constraint surface. [45]

Some important conclusions needed in section (4), are given in the following theorem:

Theorem 3.3 (Results about theories with First-class and Second-class constraints [45]) *A theory with only primary first-class constraints, implies only secondary first-class constraints. Similarly, if the singular theory implies only primary second-class-constraints, there will appear only secondary Second-class constraints.*

- *It is always possible to find a canonical transformation for the canonical varaibles $\eta = (x, p)$ in a singular theory, to another set $\eta = (\omega, Q, \Omega)$, where the Ω define the constraint surface in phase space $\Omega = 0$, the ω evolve with the physical Hamiltonian $H_{Phys} = H|_{\Omega=0}$ and its solutions are independent of a set of unphysical canonical variables Q, whose evolution depend on the arbitrary functions $\lambda_{\mathcal{P}(1)}$, that cannot be computed by requiring time conservation of all the constraints admissible by the theory; i.e. the singular theory is not completely defined.*

- *A singular theory with $\mathcal{C} = Nullity(\{\Omega, \Omega\})$ (Nullity of the Dirac matrix, in any set of equivalent constraints, e.g. $\{\Omega\}$, or $\{\Phi\}$) First-class constraints, cointains exactly the same number of arbitrary functions $\lambda_{\mathcal{P}(1)}$ that enter the theory as primarly unexpressible velocities. The fixing of these latter functions, provides total control over the solutions for the dynamical variables ω of the theory, and this last procedure is considered as a Gauge fixing of what from now on, will be defined as a Gauge theory.*

As will be seen in the next section, many results of theories with constraints will be used. However, the introduction given above is by no means complete (See [43, 44, 45, 46, 47]).

4 Results

Despite some field theories involving higher derivatives had been found to be stable in the sense of Ostrogradsky [6, 7, 8, 9, 10, 11, 12, 13], no fundamental reason for such stability has been found. One of the problems is that the source of the stability is obscured by the complicated final form of the theory. In that sense, the conditions for the stability given in [39] in theories with a finite number of degrees of freedom (f *dof*s) are quite meaningful. They have isolated the source of the Ostrogradskian instability and have stated the way in which it can be removed. Furthermore, they turn out to be quite clear and easy to apply. Nevertheless, these stabilization conditions [39], were intended only for Lagrangians with finite degrees of freedom in classical theories. Accordingly, up until now, no model in field theory involving higher derivatives, had been built using this method.

The main contribution to be presented in this thesis, is the construction of a field theory for a real scalar field including higher derivatives, based on a naive extension of the conditions suggested for f *dof*s in [39]. While applying this methodology, some subtleties arise if explicit covariance of the theory is imposed. In particular, in section 4.2 the problems of the naive extension are addressed, and a simple suggestion to overcome the difficulties, is given. In fact, based on these analyses, in section 4.3.6, a continuum field theory of a higher derivative scalar field, coupled in a very specifical way with a stable, low derivative, vector field, is analysed.

With this toy model, the stability is proposed to be checked by writing the Hamiltonian density and examining its properties. This is, however, much more subtle than it seems, because the theory has constraints. In fact, the constraints lie in the core of the stabilization method, and therefore, we held the opinion that there is no way to avoid a quantization with constraints, if a quantum theory with nice properties is expected to be reached. By different reasons, some of them explained in due moment, canonical quantization is the preferred scheme in this work. The main reason is roughly, that writing the Hamiltonian, the Ostrogradsky's instability can be immediately identified, as was explained in section (3) (See also [2, 3, 39, 40]).

Recalling again that there are constraints in these theories, by theorems that were introduced in (3.4), the existance of primarily unexpressible velocities that cannot be inverted in terms of momenta, is also guaranteed. These velocities, being not completely defined in the first stage, must be defined in some way. For this reason, an extended Hamiltonian formalism is adopted, and now the unexpressible velocities have some meaning in the theory. They are in fact Lagrange multipliers of the primary constraints, which are fixed by demanding the conservation of constraints in time. For such systems, Dirac introduced what is now known as the Dirac's programme [43, 44, 45, 46, 47], where new Poisson brackets, also referred as Dirac brackets, are formulated.

Based on such construction, by theorems that were also cited in section 3.4, the existance of a set of variables that are canonical in the new Dirac brackets, is assured. This set of variables has nice properties to analyse the dynamics of the system. In particular, they are such that the new set of variables is in fact, the disjoint union of two sets of pairs of canonically conjugated variables. Furthermore, one of the sets is equivalent to the set of initial constraints, and the variables in this new set, vanish identically whenever they are demanded to obey the equations of motion. This means that the other pairs of canonical conjugated variables, in the other disjoint set, are dynamical in a trivial constraint hypersurface inside the phase space of field configurations, defined by the previous non-dynamical set. Here, the triviality is seen more precisely, in a Hamiltonian generating the evolution of the set of dynamical variables. Following a very suggestive notation of [45], this Hamiltonian is called the Physical Hamiltonian $\mathcal{H}_{phys}$.

This brief introduction to the main concepts to be used in this work, also motivates the latter as the right approach to study the Ostrogradsky's instability. Specifically, as was commented above, the difficulties when trying to identify the source of the stability in some higher-derivative field theories lie in the fact that the final form of the theory, usually does not allow to identify the key properties leading to the stability [6, 7, 8, 9, 10, 11, 12, 13]. Nevertheless, we count with a perfect point of view for the analysis of our proposed toy theories. The reason is that, even though we are not sure of the naive assumption of stability, we do know where to look: the constraint structure of the theory. This is, based on the procedure given in [39] for finite *dofs*, we postulate a higher-derivative field theory. We claim it to be stable by introducing a constraint, then we study the constraint structure, find the physical Hamiltonian and identify if the new set of physical variables evolve in a stable way. Since the model without stabilization conditions is manifestly unstable, as can be seen from a naive Hamiltonization without extended formalism, then, only the imposed conditions could have lead to the new nice properties of the theory.

These last theories have been brought to a sound place, where canonical quantization to a higher-derivative quantum (free) field theory, can be considered. Furthermore, it is interesting enough, that the only way for the field theory to be stable, turned out to put a condition on the coupling parameter of the higher-derivative term of the scalar field, relating the latter to the inverse of the mass parameter of the stabilizer vector field. The condition is, at least for this particular toy model, not the most interesting upper bound, but only a lower bound. However, the existance of such relation was unexpected and more meaningful for the physical interpretation of higher derivative terms. They could be interpreted, at least inside the very restricted panorama, allowed by this simple toy model, as a lower bound on energies at which a higher-derivative term, may indeed, appreciably describe some dynamics in a scattering process.

Finally, it is worth mentioning that not only in field theories, have the higher derivative models been studied. Also, some models, which are the finite f dofs counterparts of the field theory, are addressed at the end of this section. This is done in order to obtain a little more insight about the results found for field theory. Accordingly, at the end of section for field theory 4.3.6, the motivation will be further explained.

Let us start by giving a brief summary of the results.

4.1 Brief summary of the results

This summary is intended as a sketchy introduction to the results given in the main sections of this thesis (4.3, 4.4), in order to make the arguments clearer.

Only a short analysis will be given at the end of this summary, based on the most important results while analysing these theories. These will be simply stated as facts and equations to be justified later, along with a comprehensive analysis in (4.3, 4.4).

The comments given above, in the introduction to section (4), provide the ideas behind the following results and set a motivation for the used methodologies. The general conclusions are given in section (5).

Higher-derivative field theory

As has been remarked before, some field theories including higher derivatives have been found to be stable under certain conditions to their parameters [6, 7, 8, 9, 10, 11, 12, 13], however, the origin of their stability has not been analysed. This problem is in general difficult to address because the cited models are of remarkable difference, possibly making an unified approach at least very cumbersome. However, since the main intention is to devise at least one methodology to build stable higher derivative theories, the following straightforward approach is considered in this introductory work:

There is a widely known model, the Pais-Uhlenbeck oscillator, which in its extension to field theory, gives a generalization of the Klein-Gordon equation and is unstable in the sense of Ostrogradsky [2, 3, 4, 5, 39, 40, 41]. We consider this as the basis for a yet to be built toy model. The additional terms that we claim, are needed for the model to be stable, are introduced as a function $\mathcal{F}$ which depends on the Pais-Uhlenbeck (P-U) field φ, their derivatives, another field κ and their corresponding kinetic and mass terms.

$$\begin{aligned} \mathcal{L}(\varphi(x),\kappa(x)) &= \mathcal{L}_{P-U}(\varphi(x)) + \mathcal{F}(\varphi(x),\kappa(x)) \\ \mathcal{L}_{P-U}(\varphi(x)) &= -\frac{1}{2}\alpha(\partial_\mu\partial^\mu\varphi)(\partial_\nu\partial^\nu\varphi) + \frac{1}{2}(\partial_\mu\varphi)(\partial^\mu\varphi) - \frac{1}{2}w\varphi^2 \end{aligned} \tag{65}$$

The reason to consider such an additional field $\kappa(x)$, is that we are taking the minimal extension of the stabilization conditions recently proposed in [39], for classical models with finitely many degrees of freedom (See section 3), in which a primary constraint between the momenta of the unstable, higher derivative field φ and the momenta of a low derivative field, here κ, must be imposed in order to control the Ostrogradskian instability.

The form of the function $\mathcal{F}(\varphi(x),\kappa(x))$ is expected to be reached by means of the cited stabilization conditions [39], and at the end, we want to check whether the extension to field theory of such procedure, indeed works.

For matter of simplicity, let us introduce the following definition,

$$\psi_\mu(x) := \partial_\mu\varphi(x) \tag{66}$$

and impose the latter as constraints on $\mathcal{L}(\varphi(x),\kappa(x))$, with Lagrange multipliers λ_u ($\mu = 0,1,2,3$):

$$\mathcal{L}(\varphi(x),\kappa(x),\psi_\mu(x),\lambda_\mu(x)) = -\tfrac{1}{2}\alpha(\partial_\mu\psi^\mu)(\partial_\nu\psi^\nu) + \tfrac{1}{2}\psi_\mu\psi^\mu - \frac{1}{2}w\varphi^2 + \mathcal{F}(\varphi(x),\kappa(x),\psi_\mu(x)) + \lambda^\mu(\psi_\mu - \partial_\mu\varphi) \tag{67}$$

With the Lagrangian as in (67), the Hessian matrix (See 3.4) turns out to be degenerate. This degeneracy is just a result of the trivial constraints that have been introduced and therefore, have no physical implication.

To define the form of the stabilizer function $\mathcal{F}(\varphi(x),\kappa(x),\psi_\mu(x))$ it seems to be necessary, as with finitely many degrees of freedom, to demand an additional constraint as described above, between the momenta of the unstable, higher derivative field φ and the momenta of the low derivative field κ. Thus, defining the momenta as usual $P_{\Phi_a} := \dfrac{\partial\mathcal{L}}{\partial\dot{\Phi}_a}$ (with $\Phi_a \in \{\varphi(x),\kappa(x),\psi_\mu(x),\lambda_\mu(x)\}$), we consider the relavant sub-matrix of the Hessian,

$$\begin{pmatrix} \delta P_{\psi_0} \\ \delta P_\kappa \end{pmatrix} = M_* \begin{pmatrix} \delta\dot{\psi}^0 \\ \delta\dot{\kappa} \end{pmatrix} \tag{68}$$

The reason to demand the constraint of P_κ with P_{ψ_0} and not with P_{ψ_i} ($i = 1,2,3$) is widely explained in section 4.2 and 4.3, motivated by the discussion in section 3. The rough idea is that the Ostrogradskian instability is expected to arise from a linearity of the momentum P_φ in the Hamiltonian when taking the Legendre transform of $\mathcal{L}$, and therefore, we hope to fix the instability by having $P_\varphi\partial_0\varphi = P_\varphi\psi_0$ non-linear in P_φ[17]. Furthermore, the constraint is not imposed directly with P_φ, basically for the same reasons given above, but also because this momentum is already constrained to be $-\lambda_0$, given the convenient introduction of constraints and Lagrange multipliers λ_μ by means of which the Lagrangian has been rewritten as in (67).

For the sub-matrix of the Hessian (67) to be degenerate, only time derivatives of the stabilizer field κ are required in the Lagrangian. Since we are interested in toy models that may motivate extensions to relativistic QFT, we demand the Lagrangian density to be a Lorentz scalar and therefore, spatial derivatives of κ should also enter at least in its kinetic term. Furthermore, this stabilizer function should include a term of the form $\eta(\partial_\mu\psi^\mu)\partial_\nu\kappa$. This is the simplest structure that this term can take if we demand degeneracy (See section 4.2). But again, for this term to be Lorentz scalar, η should in fact be a four vector such that $\eta^\nu(\partial_\mu\psi^\mu)\partial_\nu\kappa$.

In a theory with just scalar fields, we would be obligued to include an additional η^ν in the theory, which in fact breaks the covariance that we were trying to impose by means of a preferred direction,

[17]This can already be seen by referring to the final constraint content of the toy model that is being motivated with the present discussion. Just for future reference and to make plausible this rough argument, that is deepened in sections 4.2 and 4.3, the constraint we refer to, is: $\Xi_1(x) := mA_0(x) + \dfrac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0$ in equation (72). Notice that on-shell, $\psi_0(x) := \partial_0\varphi(x)$ can be rewritten in terms of P_φ, which would make $P_\varphi\psi_0$ non-linear and possibly bounded from below in P_φ. The other arising linear terms in P_φ are of no real meaning in this out-of-context deiscussion. One must recall that this is just a constraint out of other constraints that altogether will give rise to secondary constraints. Further subtleties of this boundedness or unboundedness contribution to the Hamiltonian density, are discussed in 4.3 when restricting the parameter space for the theory to be stable.

specifically, a directional derivative in the Lagrangian: $\eta^\nu \partial_\nu \kappa$. Thus, after some subtleties and some additional arguments regarding the tensor rank of the momenta that should be constrained, which are further discussed in section 4.2, the way to achieve the constraint and keep Lorentz covariance is to demand κ to be a vector field. These subtleties do not arise in non-relativistic theories and therefore were not anticipated in [39, 40, 41].

Altogether, one finds that $\mathcal{F}(\varphi(x), \kappa(x), \psi_\mu(x))$ should in fact be $\mathcal{F}(A^\mu(x), \psi_\mu(x))$, including kinetic terms for the vector field and the stabilizer term $g(\partial_\mu \psi^\mu)\partial_\nu A^\nu$, with g as a coupling constant. Since the latter term would already break U(1) global symmetry, and this is just a toy model in which we desire to understand the basics of stabilization of Higher derivative theories, there is no fundamental reason to make things unnecessarily difficult by imposing, in some cumbersome way, the U(1) symmetry and promote it to local, to reach gauge invariance. Therefore, we can also include mass terms for the vector field, the Maxwell Lagrangian $\frac{-1}{4}F_{\mu\nu}F^{\mu\nu}$ (with $F_{\mu\nu} = \partial_{[\mu}A_{\nu]}$) and $(\partial_\mu A^\mu)^2$. This last term is evidently of major importance for the stabilization of this particular toy model, because $\frac{-1}{4}F_{\mu\nu}F^{\mu\nu}$ does not include $\dot{A}_0$, which would make impossible for the constraint between P_{ψ_0} and P_{A_0} (before analysed as P_κ) to exist, if ψ_0 is dynamical at all[18].

Finally, the Lagrangian in the initial fields, in which the higher derivatives are easy to identify, takes the form:

$$\mathcal{L} = -\tfrac{1}{2}\alpha(\partial_\mu\partial^\mu\phi)(\partial_\nu\partial^\nu\phi) + \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}\omega\phi^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}\beta\partial_\mu A^\mu\partial_\nu A^\nu + \frac{m}{2}A^2 - g\partial_\mu\partial^\mu\phi\partial_\nu A^\nu \tag{69}$$

or, with the convenient introduction of the Lagrange multipliers:

$$\mathcal{L} = -\tfrac{1}{2}\alpha(\partial_\mu\psi^\mu)(\partial_\nu\psi^\nu) + \tfrac{1}{2}\psi_\mu\psi^\mu - \frac{1}{2}\omega\phi^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}\beta\partial_\mu A^\mu\partial_\nu A^\nu + \frac{m}{2}A^2 - g\partial_\mu\psi^\mu\partial_\nu A^\nu + \lambda^\mu(\psi_\mu - \partial_\mu\varphi) \tag{70}$$

The signs for the kinetic and mass terms of the vector field are not really relevant at this point, because the regions in parameter space $(\alpha, \omega, \beta, m, g)$ in which this model[19] is stable, are yet to be defined.

Now, with the general form of the Lagrangian density (70) and demanding the Hessian of the sub-system (68) to be degenerate with the specific stabilizer function $g\partial_\mu\partial^\mu\phi\partial_\nu A^\nu$, one finds that:

if the naive extension to the realtivistic field theory (69) of the stabilization conditions proposed in [39] works, then, $g^2 \overset{!}{=} \alpha\beta$ *in (69) should give a healthy continuum field theory that could be quantized.*

Let us check:

[18]Let us recall that $-\frac{1}{4}(F^{\mu\nu})^2 = \frac{P_{A_i}^2}{2} - \frac{1}{4}(F^{ij})^2$

[19]As has been emphasised, this has been just a sketchy argument for the construction of the model. Some additional subtleties, e.g. the DOFs propagated by this vector field, are more thoroughly discussed in sections 4.2 and 4.3

Since the Ostrogradskian instability has always been analysed in the Hamiltonian formalism and specifically, since the instability of the basic model of P-U has been identified in this way, and we wish to stabilize such a model, we set to Hamiltonize the theory.

There are 9 constraints, 8 of them were imposed when introducing the Lagrange multipliers, from (69) to (70), and have no physical implications. The ninth constraint arises by demanding $g^2 \overset{!}{=} \alpha\beta$ and should have a physical implication: to control the Ostrogradskian instability. This constraint is:

$$\zeta_0(x) \quad := \quad P_{A_0}(x) - \frac{g}{\alpha} P_{\psi_0}(x) \approx 0 \tag{71}$$

Note that if $g = 0$, the stabilizer term $g\partial_\mu \partial^\mu \phi \partial_\nu A^\nu$ does not enter in the Lagrangian density and the two fields $A_\mu(x)$, $\varphi(x)$ would be completely independent. No kinetic constraint (71) would appear between them, and P_{A_0} would simply fulfill the definition $P_{A_0} = \dfrac{\partial \mathcal{L}}{\partial \dot{A}_0}$ ($\not\approx 0$ *i.e.* it is not taking $g \to 0$ in (71). This constraint simply does not appear). Also note that this is just a free field theory in the usual sense, because there are only bilinears of fields in the Lagrangian $\mathcal{O}(\varphi^2, A^2, \varphi A)$ and yet, we consider some kind of kinetic interaction or coupling by means of $\mathcal{O}(\varphi A)$. This is indeed unique in order to deal with the Ostrogradskian instability and can be understood by recalling some of the features of the latter as discussed in section 3, specifically, the characterization given in [6]: "*The Ostrogradskian instability is a* ***kinetic instability*** *with an arbitrarily fast time scale*".

The extended Hamiltonian formalism must be adopted because of the primary constraints that lead to primarily unexpressible velocities (See section 3 or [45], chapter 2), and just summarizing the results to be found in section 4.3 and that are widely analysed there, the constraint structure of the theory, including primary and second ones, all second-class (See section 3) $\{\Pi\} = \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}$, is:

$$\begin{aligned}
\zeta_0(x) &:= P_{A_0}(x) - \frac{g}{\alpha} P_{\psi_0}(x) \approx 0 \\
\zeta_i(x) &:= P_{\psi_i}(x) \approx 0 \\
\Delta_\mu(x) &:= P_{\lambda_\mu}(x) \approx 0 \\
\Lambda^0(x) &:= P_\varphi(x) + \lambda^0(x) \approx 0 \\
\Xi_1(x) &:= mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0 \\
\Xi_{2_i}(x) &:= -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \approx 0 \\
\Xi_{3_i}(x) &:= -\psi^i(x) - \partial_i \varphi(x) \approx 0
\end{aligned} \tag{72}$$

The Dirac brackets built with these constraints is given in equation (188), or in short,

$$\{\mathcal{F}(x), \mathcal{G}(y)\}_{D(\Pi)} = \{\mathcal{F}(x), \mathcal{G}(y)\} - \int d^3z d^3z' \left(\{\mathcal{F}(x), \Pi_a(z)\} \{\Pi(z), \Pi(z')\}^{-1}_{a,a'} \{\Pi_{a'}(z'), \mathcal{G}(y)\} \right) \tag{73}$$

The new set $\{\Upsilon\} = \{\{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\} \cup \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i},\ \zeta_0, \zeta_i, \Delta_0, \Delta_i\}\}$ of canonical fields in the Dirac brackets (Not Poisson brackets. See section 3), is the disjoint union of two sets, $\{\Upsilon\} : \{\{\omega\} \cup \{\Omega\}\}$. The motivation for such construction will become apparent in short, while analysing the stability of the toy model. First, the dynamical fields $\{\omega\}$ are:

$$\begin{aligned}
\eta(x) &=: \varphi(x) - P_{\psi_0}(x) \\
P_\eta(x) &=: P_\varphi(x) \\
\gamma(x) &=: A^0(x) - \frac{\alpha}{m\alpha + \beta}\partial_i P_{A_i}(x) \\
P_\gamma(x) &=: \frac{m\alpha + \beta}{g} P_{\psi_0} \\
\Theta^i(x) &=: A^i(x) \\
P_{\Theta_i}(x) &=: P_{A_i}(x)
\end{aligned} \tag{74}$$

And the identically vanishing, new, equivalent constraints $\{\Omega\}$ (Notice the absence of the weak equalities ≈ 0), are:

$$\begin{aligned}
\zeta_0(x) &=: P_{A_0}(x) - \frac{g}{\alpha} P_{\psi_0}(x) \\
\zeta_i(x) &=: P_{\psi_i}(x) \\
\Delta_\mu(x) &=: P_{\lambda_\mu}(x) \\
\Lambda^0(x) &=: P_\varphi(x) + \lambda^0(x) \\
\Xi_1(x) &=: mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \\
\Xi_{2_i}(x) &=: -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \\
\Xi_{3_i}(x) &=: -\psi^i(x) - \partial_i \varphi(x)
\end{aligned} \tag{75}$$

It is important to emphasize that $\{\Omega\} = \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}$ are no longer constraints between the old canonical fields, but now, they are to be considered **canonical variables on their own**, that however, vanish identically when the equations of motion are satisfied and therefore are not dynamical. For this reason, in (75) there are no weak equalities ≈ 0, in comparison with (72).

The equal-time Dirac brackets in the new set of variables, are:

$$\begin{aligned}
\{\eta(x), P_\eta(y)\}_{D(\Omega)} &= \delta^{(3)}(\vec{x} - \vec{y}) \\
\{\gamma(x), P_\gamma(y)\}_{D(\Omega)} &= \delta^{(3)}(\vec{x} - \vec{y}) \\
\{\Theta^i(x), P_{\Theta_j}(y)\}_{D(\Omega)} &= \delta^i_j \delta^{(3)}(\vec{x} - \vec{y})
\end{aligned}$$

And,

$$\{\Upsilon_a(x), \Upsilon_b(y)\}_{D(\Omega)} = 0 \tag{76}$$

for all the other Dirac brackets.

The physical Hamiltonian ($\mathcal{H}_{phys}$) in the continuum field theory, in the new dynamical variables, is constructed in section 4.3. Roughly, it is built by rewriting the extended Hamiltonian in terms of the new set of canonical variables $\{\Upsilon\} = \{\{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\} \cup \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}\}$ and by means of some theorems of theories with constraints, some of them cited in section 3, also in [45, 46, 47], it is possible to set the new variables that are equivalent to the previous constraints $(\Omega : \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\})$ to 0, which leads to $\mathcal{H}_{phys}$. In this sketchy introduction to the results, the physical Hamiltonian that may lead to a stable model, upon the conditions in parameter space $\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $\mathbb{R} \ni m \overset{!}{>} 0$, $\omega \overset{!}{=} 0$, is:

$$\begin{aligned}
\mathcal{H}_{phys} \quad (\omega) \quad &= \frac{|\beta|}{2}\frac{1}{(m\alpha - |\beta|)^2}\left(P_\gamma^2 + \alpha \nabla P_\gamma \cdot \nabla P_\gamma\right) + \frac{1}{4}(F^{ij})^2 + \frac{m}{2|\beta|}(m\alpha - |\beta|)\gamma^2 \\
&+ \quad \frac{1}{2}\left(P_{\Theta_j}^2 + \frac{\alpha}{m\alpha - |\beta|}\nabla P_{\Theta_j} \cdot \nabla P_{\Theta_j}\right) + \frac{m}{2}(\Theta^i)^2 + \frac{P_\eta^2}{2} + \frac{1}{2}\nabla\eta \cdot \nabla\eta \\
F_{ij} \quad &=: \quad \partial_i \Theta_j - \partial_j \Theta_i
\end{aligned} \tag{77}$$

This Hamiltonian generates the evolution of the **new dynamical variables** $\omega = \{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\}$ by means of the Dirac bracket, and is only positive definte under the condition $\alpha \overset{!}{>} \frac{|\beta|}{m}$. We therefore find that the latter relation between the higher derivative coupling α, the kinetic coupling of the vector field β and its mass m, must be satisfied, if higher derivative terms are to describe the dynamics of the field $\varphi(x)$. Only in this case, a quantum theory of this toy model can be written, because the vacuum can be well defined. Equivalently, only in this case ghosts are not propagated by φ.

In other words, the fact that this classical Hamiltonian $\mathcal{H}_{phys}$ can be positive definite and bounded from below has an important physical significance. It shows that the constraint ζ_0 imposed in the theory (69) by demanding the sub-system (68) to be degenerate and that at least imposes the inclusion of the term $g\partial_\mu\partial^\mu\varphi\partial_\nu A^\nu$ in the Lagrangian density, together with $g^2 = \alpha\beta$, successfully eliminates the Ostrogradskian instability and the theory is good enough to be quantized.

Note that this simple conclusion can only be obtained from this classical Physical Hamiltonian density in the new set of variables, because by the construction of the new sets $\{\Upsilon\} : \{\{\omega\} \cup \{\Omega\}\}$, we can assure that $\{\omega\}$ evolve freely by $\mathcal{H}_{phys}$, in some trivial constraint hypersurface in phase space of field configurations, which is defined by the identically vanishing fields $\{\Omega\} \ni \Omega_f \equiv 0 \ \forall f$ (provided the equations of motion are satisfied). On the other hand, with the Hamiltonian density in the original fields φ, A_μ (Not $\mathcal{H}_{phys}(\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i})$, to be constructed in section 4.3. See equation (123)) and 16 constraints $\{\Pi\}$ (72), the one for which the Poisson brackets (not Dirac brackets) must be used to compute the dynamics, it is not possible to make such a simple analysis. Positive definiteness of this classical Hamiltonian (not $\mathcal{H}_{phys}$) would not mean anything, because there are some complicated constraints making the evolution non-trivial. The constraints (72) must still be imposed.

The advantage of finding the new set of fields and the physical Hamiltonian is evidently a gain in physical insight, and we can assure that the toy model (69) written in terms of the fields $\varphi(x)$ and $A_\mu(x)$ that seemed to propagate ghosts because of the higher order derivatives, in fact freely propagates healthy DOFs $\omega = \{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\}$. These are the **physical** DOFs propagated by the theory, but the key point in all this analysis, is that without doing all the procedure to find the DOFs $\{\omega\}$ and the physical Hamiltonian, we know that (69) with the conditions $\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $\mathbb{R} \ni m \overset{!}{>} 0$, $\omega \overset{!}{=} 0$, $g^2 \overset{!}{=} \alpha\beta$ and $\alpha \overset{!}{>} \frac{|\beta|}{m}$, safely propagates $\varphi(x)$ and $A_\mu(x)$ in a constrained surface given by (72).

The canonical quantization is carried out by promotion of the initial fields φ, A_μ and momenta to field operators and the Dirac (Not Poisson) equal-time brackets, are promoted to commutators in the following prescription,

$$\{\cdot,\cdot\}_{D(\Omega)} \to \frac{1}{i}[\cdot,\cdot] \tag{78}$$

It is also possible to quantize directly the physical DOFs $\omega = \{\eta, P_\eta, \gamma, P_\gamma,\ \Theta^i, P_{\Theta_i}\}$ which are canonical in the Dirac brackets, and therefore, the promotion to commutators is simply taken from the Poisson brackets (or also from the Dirac brackets, which in these DOFs, by definition, trivially reduce to Poisson brackets).

Specifically, the equal-time commutators in the physical DOFs, are:

$$\begin{aligned} [\eta(x), P_\eta(y)] &= i\delta^{(3)}(\vec{x}-\vec{y}) \\ [\gamma(x), P_\gamma(y)] &= i\delta^{(3)}(\vec{x}-\vec{y}) \\ \left[\Theta^i(x), P_{\Theta_j}(y)\right] &= i\delta^i_j\delta^{(3)}(\vec{x}-\vec{y}) \end{aligned}$$

And,

$$[\Upsilon_a(x), \Upsilon_b(y)] = 0 \tag{79}$$

for all the other commutators (with $\{\Upsilon\} = \{\{\omega\}\cup\{\Omega\}\} = \{\{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\}\cup\{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}\}$).

Therefore, we have checked that the naive extension of the stabilization conditions proposed for finitely many DOFs in [39], systematically applied to the Higher derivative field theory of Pais-Uhlenbeck, does the job at least in certain regions of parameter space, which suggests that this method could be applied to physically interesting theories. However, as we have found by demanding Lorentz covariance, other subtleties may arise if one insists in other symmetry groups for the theory, and it may be that the extension of the methodology requires modifications, or additional requirements, as here, or may even not go through. (For a detailed analysis, see sections 4.3.6 and 5).

The fact that we found additional conditions in parameter space for the theory, would not be ideal if one hopes that the conditions given in [39] work flawlessly upon extensions, but one can understand this issue as arising from other instabilities not related with Ostrogradsky's; something common that a model can show if their parameters are not restricted. The fundamental difference with the stable models analysed in [6, 7, 8, 9, 10, 11, 12, 13], is that now we know, by construction, the source of the stability. Furthermore, particularly interesting is the condition relating $\alpha \overset{!}{>} \frac{|\beta|}{m}$. This will be briefly discussed at the end of this section and in sections 4.3.6, 5.

The analysis has been done for a free theory, but it may be that this is the building block for a healthy higher derivative interacting theory. The main reason for such expectation is that a HTDT also shows its instability in a Feynman propagator of the higher derivative field, in QFT, that can be splitted in such a way that ghosts come up as propagating with the wrong sign. Since the propagator always is computed as the inverse of the differential operator derived from the free theory, bilinears of the form $\mathcal{O}(\varphi^2, A^2, \varphi A)$, if the ghosts are eliminated at this level, it is plausible that they are controled even in the perturbative expansion of a higher derivative interacting theory. The latter, at least in the case where no additional interaction terms involving higher derivatives enter the theory.

The model with a finite number of degrees of freedom

As will become apparent at the end of section (4.3.6) there are some reasons to motivate the discussion with the simpler case of finitely many *dof*s. Besides being easier to deal with, it is possible to build a model which can be interpreted as a limiting case of the field theory just analysed, after analysis and re-interpretation of the objects under consideration (fields $\leftrightarrow$ f *dof* s). In this sense, this subsection mainly serves as a check for consistency of the results obtained for field theory.

The general Lagrangian in the case of finitely many *dof*s, is:

$$L(x, y) = -\frac{1}{2}\alpha\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 - \frac{\beta}{2}\dot{x}^2 + \frac{m}{2}x^2 - g\ddot{y}\dot{x} \tag{80}$$

Again, the condition $g^2 \overset{!}{=} \alpha\beta$ needs to be imposed in order for the theory to be stable.

For this case, the basic results are very similar to those found for field theory (For a complete analysis see section 4.4). This is another advantage of taking this Hamiltonian with constraints formalism. The basic results are, as before, the following:

The constraint content of the theory (Again, all second-class. Now, only 4 constraints), with similar interpretation as with field theory, in regards of their physical implications, after defining for convenience $Q := \dot{y}$ and introducing the lagrange multiplier λ, is:

$$\begin{aligned}
\zeta &=: P_x - \frac{g}{\alpha} P_Q \approx 0 \\
\Lambda &=: P_y + \lambda \approx 0 \\
\Delta &=: P_\lambda \approx 0 \\
\Xi &=: mx + \frac{g}{\alpha}(P_y - Q) \approx 0
\end{aligned} \tag{81}$$

The Dirac bracket is built based on these constraints (See appendix 6.2), and the new set of variables, canonical in the Dirac brackets is in vector notation:

$$\begin{pmatrix} Q \\ y \\ x \\ \lambda \\ P_Q \\ P_y \\ P_x \\ P_\lambda \end{pmatrix} \rightarrow \begin{pmatrix} (Q - P_y - \frac{\alpha}{g} mx) \\ y - P_Q \\ x \\ \lambda + P_y \\ \frac{m\alpha + \beta}{g} P_Q \\ P_y \\ P_x - \frac{g}{\alpha} P_Q \\ P_\lambda \end{pmatrix} =: \begin{pmatrix} \Xi' \\ y' \\ x' \\ \Lambda \\ P_{x'} \\ P_{y'} \\ \zeta \\ \Delta \end{pmatrix} := \Upsilon \tag{82}$$

Again, the new set is the disjoint union of two sets $\{\omega\} = \{x', P_{x'}, y', P_{y'}\}$ dynamical, and $\{\Omega\} = \{\Xi', \Lambda, \Delta, \zeta\}$ non dynamical, that are equivalent to the previous set of constraints, although this time, are canonical variables on their own, vanishing identically when the equations of motion are satisfied. The Dirac brackets between variables are:

$$\begin{aligned}
\{y'(t), P_{y'}(t)\}_{D(\Omega)} &= 1 \\
\{x'(t), P_{x'}(t)\}_{D(\Omega)} &= 1 \\
&\text{And,} \\
\{\Upsilon_a(t), \Upsilon_b(t)\}_{D(\Omega)} &= 0
\end{aligned} \tag{83}$$

for all the other Dirac brackets, where, $\{\Upsilon\} : \{\{\omega\} \cup \{\Omega\}\}$.

The most general phsyical Hamiltonian in classical theory, in terms of the dynamical variables $\{\omega\}$, no yet positive definite, is:

$$H_{phys} = -\frac{\beta}{2} \frac{1}{(m\alpha + \beta)^2} P_{x'}^2 + \frac{w}{2} \left(\frac{g}{m\alpha + \beta} P_{x'} + y' \right)^2 + \frac{P_{y'}^2}{2} - \frac{m}{2\beta} x'^2 (m\alpha + \beta) \tag{84}$$

It turns out that in this easier case, there are more possibilities to eliminate the Ostrogradsky's instability. All of them are examined in section 4.4. We cite here only the results for three of them:

- If the parameters are restricted to $\mathbb{R} \ni \alpha \overset{!}{<} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{=} 0$, $w \overset{!}{\geq} 0$, $\mathbb{R} \ni g \overset{!}{=} \pm\sqrt{|\alpha||\beta|}$. With the new set of canonical variables in the Dirac brackets, the physical Hamiltonian reads:

$$H_{phys} = \frac{1}{2|\beta|}P_{x'}^2 + \frac{w}{2}\left(\pm\sqrt{\frac{|\alpha|}{|\beta|}}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} \tag{85}$$

By canonical quantization as described in (4.4), by promoting the canonical variables in the new set, to operators, and the Dirac brackets to commutators as usual,

$$\begin{aligned} \{x', P_{x'}\}_{D(\Omega)} = 1 \to [x', P_{x'}] &= i \\ \{y', P_{y'}\}_{D(\Omega)} = 1 \to [y', P_{y'}] &= i \end{aligned} \tag{86}$$

under the assumption of complete bases, and after stating the time independent Schrödinger equation (See section 4.4) for states $|\Psi\rangle \overset{\text{Pick convenient basis}}{\longrightarrow} \langle P_{x'}, y''|\Psi\rangle = \Psi(P_{x'}, y'')$ in Hilbert space, the energy spectrum can be found as[20]

$$\begin{aligned} E_n(P_x) &= \sqrt{w}\left(n + \frac{1}{2}\right) + \frac{1}{2|\beta|}P_{x'}^2 \\ n &\in \mathbb{N} \cup \{0\} \end{aligned} \tag{87}$$

Being $\{|P_{x'}\rangle\}$ the basis for $P_{x'}$ ($\int(d^3P_{x'})|P_{x'}\rangle\langle P_{x'}| = I$), from (87) the energy spectrum is continuum. It can be seen from (87) that the spectrum is also real, positive and bounded from below. This result, yet simple, is very important. The boundedness from below of the spectrum, assures that there is a lowest-energy state, or unique vacuum in Hilbert space, whose non-existance signaled the presence of the Ostrogradsky's instability in the theory. We can finally assure that this model with a finite number of degrees of freedom and with higher derivatives, can indeed be stabilized by the method proposed by [39] (i.e. demanding the existance of a particular kind of constraint. *i.e.* ζ in (81) or in (204) in the appendix 6.2). And furthermore, the stabilization does not get lost when quantizing the theory, which had not been exemplified before.

- For the cases with $m \neq 0$ and $w \overset{!}{=} 0$, very similar results to those found for fields, arise:

  If the parameters are restricted to $\mathbb{R} \ni \alpha \overset{!}{<} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{<} 0$, $w \overset{!}{\geq} 0$, $\mathbb{R} \ni g \overset{!}{=} \pm\sqrt{\alpha|\beta|}$. The physical Hamiltonian in the new set of canonical variables is:

$$H_{phys} = \frac{|\beta|}{2}\frac{1}{(|m||\alpha| - |\beta|)^2}P_{x'}^2 + \frac{w}{2}\left(\pm\frac{\sqrt{|\alpha||\beta|}}{(|\alpha||m| - |\beta|)}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} + \frac{|m|}{2|\beta|}x'^2(|\beta| - |\alpha||m|) \tag{88}$$

[20] For all the details, see section 4.4.

Now α has an upper limit defined by the mass of the stabilizer $|\alpha| < \frac{|\beta|}{|m|}$ in order for the model to be free of Ostrogradsky's instability. Some comments on this result are given below.

- Finally if the parameters are restricted to $\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{>} 0$, $w \overset{!}{\geq} 0$, $\mathbb{C} \ni g \overset{!}{=} \pm i\sqrt{\alpha|\beta|}$. The Lagrangian is:

$$L(x, y) = -\frac{1}{2}\alpha\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 + \frac{|\beta|}{2}\dot{x}^2 + \frac{m}{2}x^2 \mp i\sqrt{\alpha|\beta|}\ddot{y}\dot{x} \tag{89}$$

The first thing to note is the resemblance with the Lagrangian density for field theory after identifiying the completely different elements $x \leftrightarrow A^0$, $y \leftrightarrow \varphi$, and noting that no degree of freedom in the current case was introduced as analog to the other components of the vector field A^i, which were introduced just for explicit covariance of the theory, as is explained in 4.2. Even though the interpretation of these objects is completely different, it is useful to check consistency of the results.

Let us make a little detour to the field theory, in order to motivate the comparison: Consider the mode expansion for the initial fields in the continuum field theory $\{\Phi\} = \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$. The latter define the fields $\{\Upsilon\} = \{\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i\}$ in terms of modes, by means of their definition (74), (75), but now, let us take only one mode of the expansion, in particular, the 0-momentum ($\vec{p} = \vec{0}$) mode. Then, every term in the physical Hamiltonian (77), in the new set of fields expanded in these modes, including spatial derivatives, vanishes, because the derivatives of the exponentials put a $\vec{p}$ in front of these terms. Finally, we end up with a function, which is not the Hamiltonian $\mathcal{H}_{phys}(x)$, but only a term of the sum of its mode expansion, readily, the 0-momentum contribution to the energy function at every space-time point where $\mathcal{H}_{phys}(x) \to h_{phys_{\vec{p}=\vec{0}}}(x)$ can be evaluated. The vanishing of the spatial derivatives for this 0-momentum function makes sense, because we are "downgrading" from a covariant field theory to a finite *dof*s case.

This is a very rough approach but it is useful to gain some insight of the field theory. Upon re-interpretation, demoting each of the fields $\{\eta, \gamma, \Theta^i, P_\eta, P_\gamma, P_{\Theta_i}\}$ valued at every space-time point, to simple degrees of freedom, the 0-momentum function of the "physical Hamiltonian" described above (Upper function in (90)), takes a form that can be compared to the physical Hamiltonian for the finite *dof*s case, that is currently analysed (Lower function in (90)):

$$\begin{aligned} h_{phys_{\vec{p}=\vec{0}}} &\approx \frac{|\beta|}{2}\frac{1}{(m\alpha - |\beta|)^2}P_\gamma^2 + \frac{P_\eta^2}{2} + \frac{m}{2|\beta|}\gamma^2(m\alpha - |\beta|) + \frac{1}{2}P_{\Theta_j}^2 + \frac{m}{2}(\Theta^i)^2 \\ &\updownarrow \\ H_{phys} &= \frac{|\beta|}{2}\frac{1}{(m\alpha - |\beta|)^2}P_{x'}^2 + \frac{P_{y'}^2}{2} + \frac{m}{2|\beta|}x'^2(\alpha m - |\beta|) \end{aligned} \tag{90}$$

As commented before, there are only two terms not possible to identify (upon re-interpretation of the objects) in (90), these are the ones containing $\Theta^i =: A^i$ and $P_{\Theta_i} =: P_{A_i}$. Let us recall that the following was a possible assignment: $\varphi \leftrightarrow y$ and $A^0 \leftrightarrow x$. Leaving A^i without counterpart in the current model with only finite degrees of freedom.

Furthermore, let us note that H_{phys} in (90) is again positive define only when $\alpha > \dfrac{|\beta|}{m}$. All in all, this is also a very meaningful result to check consistency, because we analysed completely different models by means of two equivalent formalisms in field theory and finite degrees of freedom. Then, at the end, upon approximations and a very plausible analysis, we reach as the "limiting case" of field theory, what we had found for finite degrees of freedom.

To sum up:

We have proposed a covariant continuum field theory with higher derivatives (69). Then, based on the stabilization procedure given by [39], that was only written for theories with finite degrees of freedom, we have naively applied the latter to the proposed field toy model, expecting some sort of stabilization. Finally, we have verified that after some subtleties regarding the construction of the model and the transformation properties of the stabilizer field, it is possible to build a classical continuum field theory that can be, in principle - *disregarding problems of other origin, like dynamical instabilities not related with that of Ostrogradsky*- be brought to the starting point of a higher-derivative quantum field theory.

Furthermore, the only way to stabilize the higher-derivative scalar field in the sense of Ostrogradsky for this particular toy model, forces a relation between the coupling parameter of the higher-derivative term of the scalar field, $-\frac{1}{2}\alpha\partial_\mu\partial^\mu\varphi\partial_\nu\partial^\nu\varphi$ and the mass parameter of the stabilizer vector field $+\dfrac{m}{2}A_\mu A^\mu$ (as well, but less important, to $\dfrac{1}{2}|\beta|\partial_\mu A^\mu\partial_\nu A^\nu$. $|\beta| \to 1$ is possible without further interpretation in comparison with a mass term). The condition is, at least for this particular field toy model, not the most interesting upper bound, but only a lower bound $\alpha > \dfrac{|\beta|}{m}$. However, the existance of such relation was unexpected and more meaningful for the physical interpretation of higher-derivative terms. They could be interpreted, at least inside the very restricted panorama allowed by this simple toy model, as a lower bound on the energies at which a higher-derivative term, may appreciably describe some dynamics in a scattering process, based on for instance, the possibly known mass of the vector field (which in this case is the stabilizer with the usual low derivative terms).

Nevertheless, we do not claim that this kind of relation should appear in every possible stable higher-derivative field theory. We have only found a case in which this happens, the sole fact that this toy model with these interesting properties exists, speaks about possibly physically interesting higher-derivative models that include such kind of relations. In any case, if some higher-derivatives are needed to describe some dynamics in a physical phenomena, that such relation appears is much more interesting than a completely "disconnected" term, just added by hand to a low-derivative theory (*i.e.* no relation at all between parameters of high and low derivative terms). It may also be the case that even an upper bound, on the higher-derivative term parameter, arises in field theory, as we found for a model with finite degrees of freedom.

Now, let us proceed to a deeper discussion and deduction of the above-mentioned results.

4.2 Stabilization of a higher-time derivative scalar field

The stabilization procedure given in [39] was intended only for Lagrangians with finite degrees of freedom in classical theories. In particular, as was noted in section 3, the condition is simply satisfied by the singularity of the Hessian (kinetic) matrix in the Lagrangian formalism. In the Hamiltonian formalism that singularity of the Hessian implies the existance of primary constraints between momenta of the stable and unstable degrees of freedom.

In this work, we set to build models with higher-time derivatives in continuum field theory, *i.e.* including infinite degrees of freedom *dofs*. To begin with, we have naively extended the conditions proposed in [39], for a finite number of *dofs*, by demanding the singularity of the Hessian matrix. In principle, the simpler model consists of a scalar field ($\varphi(x)$) whose dynamics, we claim, could be described by terms involving up to second-time derivatives in the Lagrangian density. Furthermore, to write a theory susceptible of describing physical phenomena, we want it to be at least, in a very weak sense, mathematically stable. In the present case we are only building a toy model that allows us to see, how the already discussed stabilization condition, behaves. Therefore, following the argument in [39] for finite *dofs*, we also claim that, if we are to introduce higher derivatives for the field $\varphi(x)$, we also should include another field, which we will refer to, as the stabilizer $\kappa(x)$. The simplest case is for the stabilizer to be a scalar field, therefore, we begin with this model, but we will briefly show that it is not possible, in general, to stabilize in this way. Since this is not going to be the main model in this work, we give only a sketchy description of why it does not work.

For the simplest case, the Lagrangian density would be[21]

$$\begin{aligned}\mathcal{L}(\varphi(x),\kappa(x)) &= -\frac{1}{2}\alpha(\partial_\mu\partial^\mu\varphi)(\partial_\nu\partial^\nu\varphi)+\frac{1}{2}(\partial_\mu\varphi)(\partial^\mu\varphi)-\frac{1}{2}\omega\varphi^2-\frac{1}{2}\beta\partial_\mu\kappa\partial^\mu\kappa\\ &+ \frac{m}{2}\kappa^2+\mathcal{F}(\varphi,\kappa)\end{aligned} \tag{91}$$

where $\mathcal{F}(\varphi,\kappa)$ is some function depending on the fields, that we are to determine in order to make the theory stable. For this purpose we simply demand the Hessian to be singular, but before, let us write an equivalent Lagrangian that allows to work in a much more familiar way, with up to one derivative in the Lagrangian.

Let us define:

$$\psi_\mu(x) := \partial_\mu\varphi(x) \tag{92}$$

[21]The signs of the terms containing κ were chosen in this way for future reasons, but it will turn out to be the usual way, *i.e.* $\frac{1}{2}\beta\partial_\mu\kappa\partial^\mu\kappa-\frac{m}{2}\kappa^2$

and introducing four Lagrange Multipliers λ^μ, with $\mu = 0, 1, 2, 3$.

$$\mathcal{L}_{equivalent} = -\frac{1}{2}\alpha(\partial_\mu\psi^\mu)^2 + \frac{1}{2}\psi^2 - \frac{1}{2}\omega\varphi^2 - \frac{1}{2}\beta(\partial_\mu\kappa)^2 + \frac{m}{2}\kappa^2 + \lambda^\mu(\psi_\mu - \partial_\mu\varphi) + \mathcal{F}(\psi^\mu, \kappa) \quad (93)$$

As will be widely explained in the next section and was briefly commented in 4.1, this Hessian is already singular due to the introduction of $\psi_\mu := \partial_\mu\varphi$ as constraints with their respective Lagrange multipliers, which is not related to the stabilization and in fact is just a matter of convenience. Therefore, demanding a submatrix of the Hessian to be degenerate, it is easily found that $\mathcal{F}(\psi^\mu, \kappa)$ should take the following form, in the simplest of the possible cases:

$$\mathcal{F}(\psi^\mu, \kappa) = \eta\partial_\mu\psi^\mu\partial_\nu\kappa \quad (94)$$

As commented above, if we are to demand $\mathcal{L}$ to be explicitly covariant, the factor η should in fact, be a 4-vector ($\mathcal{F}(\psi^\mu, \kappa) = \eta^\nu\partial_\mu\psi^\mu\partial_\nu\kappa$). Then, being this a parameter in the theory, once we fix a 4-vector η^ν, covariance is not really there, we have just picked a preferred direction by imposing upon $\partial_\nu\kappa$ to be in fact a directional derivative precisely in the direction fixed by η^ν. However, let us not stop here. If the momenta are computed, evidently, P_{ψ_μ} must be a vector field, but P_κ is a scalar field. Therefore, a relation between these two objects, which transform under Lorentz transformations in completely different way, cannot be established. In short, if we demand covariance and **if** we consider only the simplest term for the stabilizer function $\mathcal{F}(\psi^\mu, \kappa)$ in $\mathcal{L}$, the stabilizer cannot be a scalar field.

There are two possible arguments against the previous discussion: first, it could be said that the difficulty arises by introducing $\psi_\mu(x) := \partial_\mu\varphi(x)$, but the fact is that we did this only for convenience, because at the end what is needed is a coupling, that will be called dynamical, between an unstable part of the higher derivative field φ and the momenta of the stabilizer κ. But it turns out that if one computes the momenta as defined by Ostrogradsky for theories with higher derivatives, after suitable extension to fields (See section 3 or [2, 3, 42]) exactly the same problem arises. Specifically, if one computes the momenta for $\mathcal{L}$ from the corresponding definition given in section 3, one obtains two different momenta for φ, $P_{1_{\alpha\beta}}$, P_{2_α}. One of them is a covariant tensor field of rank 2: $P_{1_{\alpha\beta}} = (-\alpha\Box\varphi - \eta^\mu\partial_\mu\kappa)\eta_{\alpha\beta}$ and the other, a vector field $P_{2_\sigma} = \partial_\sigma\varphi - \partial^\beta P_{1_{\sigma\beta}}$ $= \partial_\sigma\varphi - \partial_\sigma(-\alpha\Box\varphi - \eta^\mu\partial_\mu\kappa)$, these are the equivalent momenta, in a completely different fomalism, to P_{ψ_μ} and P_φ with the previous $\mathcal{L}_{equivalent}$. However, the momentum for κ is just a vector field $P_{\kappa_\alpha} = -\frac{\beta}{2}\partial_\alpha\kappa + \eta_\alpha\Box\varphi$. The most similiar relation between momenta is at first sight between $P_{1_{\alpha\beta}}$ and P_{κ_α}, however, they are, again, different objects (different tensor rank) and would not give the desired constraint. For P_{2_σ} and P_{κ_α}, the relation is by derivatives. Again, not a constraint between momenta. The conclusion from this short analysis is that the problem described above arises in both approaches when the Hamiltonian is going to be built, therefore, it seems to be not "approach-dependent".

There is yet another definition of momenta given also in [42] (See section 3). These momenta do not have the same problem, because all the momenta are defined as scalars, but it is explicitly explained that these definitions do not give the conventional meaning of "mechanical impulses" which one expects when building a Hamiltonian related to the energy of the system [22]. In short, a Hamiltonian built by Legendre transform with other definition of momenta, from the first two already tried, do not give a function related to the energy of the system. However, for our present purposes, we are interested in a Hamiltonian whose spectrum in a quantum theory is in fact the energy spectrum, because only with such a function instability in the sense of Ostrogradsky can be checked[23].

Finally, we can say that if we want a Hamiltonian with the stated property, the problem arising above is not approach dependent, but emerges in both formalisms. Furthermore, we have the additional problem of a fixed 4-vector η^ν which spoils covariance by fixing a direction, in particular, fixing the stabilizer term $\mathcal{F}(\varphi,\kappa) = \eta^\nu \partial_\mu \partial^\mu \varphi \partial_\nu \kappa$ to include a directional derivative, precisely in the direction given by η^ν.

The second argument against the previous discussion is that we could have considered interaction terms in the stabilizer function $\mathcal{F}(\varphi,\kappa)$, in order for a scalar stabilizer field to be good enough to eliminate the Ostrogradskian instability, but, if we are to control the instability at all orders in perturbation theory, it should be necessary to control all the propagating ghosts in free theory, without the "help" of interaction terms $\mathcal{O}(\varphi^3, \varphi^2 A, \dots)$. This is, it should not be possible to split the propagator of a higher derivative field into different propagators, some of them with "wrong" signs. Since the propagator always is computed as the inverse of the differential operator derived from the free theory, the argument follows. This, we expect, could lead to a healthy higher derivative interacting theory. However, many subtleties should be checked before claiming it to be truth.

Now, we proceed to the main results, by taking a vector field as stabilizer.

4.3 Higher derivative scalar free field theory stabilized by a kinetic coupling to a vector field

Based on the previous discussion it is now clear that a covariant, higher derivative scalar field theory, cannot in general be systematically stabilized by means of a dynamical coupling with another scalar field. The requirement of a constraint between the unstable degree of freedom of the scalar field and the momentum conjugate to the stabilizer field, imposes a restriction on the structure of the latter. To be more precise, for a scalar field theory involving at most second derivatives in the Lagrangian, a vector field as stabilizer is necessary, in order for the momentum of the former and the momentum of the latter, to have the same tensor rank and to allow for a constraint between them to even exist. Therefore, following the arguments given in 4.2 and the brief summary to these results given in the introduction to this section (4), the following model is proposed:

$$\begin{aligned} \mathcal{L}(\varphi(x), A^\mu(x)) &= -\frac{1}{2}\alpha(\partial_\mu\partial^\mu\varphi)(\partial_\nu\partial^\nu\varphi) + \frac{1}{2}(\partial_\mu\varphi)(\partial^\mu\varphi) - \frac{1}{2}\omega\varphi^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}\beta\partial_\mu A^\mu \partial_\nu A^\nu \\ &+ \frac{m}{2}A^2 - g\partial_\mu\partial^\mu\varphi\partial_\nu A^\nu \end{aligned} \tag{95}$$

[22] There are, however, difficulties when quantizing. See [53]

[23] In fact, if one computes the momenta in the third formalism in [42], no relation appears between these "non-mechanical momenta"

where α, w, β and m are, in principle, free parameters which we allow to $\in \mathbb{R}$. g, as was pointed out in previous sections, turns out to depend on α and β as $g^2 = \alpha\beta$, and this time, we allow $g \in \mathbb{C}$. From now on, the space-time dependence of the fields will be dropped, but it should be understood as implicit in the definition of the fields. However, when pertinent, *e.g.* Poisson and Dirac brackets, they will be included.

It is worth to bring to attention that this particular form, specifically $-g\partial_\mu\partial^\mu\varphi\partial_\nu A^\nu$, is considered as a naive extension of the stabilization procedure given in [39], which different to this case, only applied to classical models with only finite degrees of freedom. It is precisely the intention with this work, to show that a model for infinite degrees of freedom and higher derivatives, can indeed be constructed using the same criterion of stabilization, while giving extra attention to the subtleties involved when expecting the theory to be covariant.

Let us identify the parts that make up the Lagrangian. It can be written as:

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}_{P-U} + \mathcal{L}_A + \mathcal{L}_{dc} \\
\mathcal{L}_{P-U}(\varphi) &= -\frac{1}{2}\alpha(\partial_\mu\partial^\mu\varphi)(\partial_\nu\partial^\nu\varphi) + \frac{1}{2}(\partial_\mu\varphi)(\partial^\mu\varphi) - \frac{1}{2}\omega\varphi^2 \\
\mathcal{L}_A(A^\mu) &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{m}{2}A^2 - \frac{1}{2}\beta\partial_\mu A^\mu\partial^\mu A_\mu \\
\mathcal{L}_{dc} &= -g\partial_\mu\partial^\mu\varphi\partial_\nu A^\nu
\end{aligned} \tag{96}$$

where $\mathcal{L}_{P-U}(\varphi)$ is almost the **Pais-Uhlenbeck field Lagrangian**, with a difference from the latter only in the domain to which the parameters belong in the parameter space $\equiv (\mathbb{R} - \{0\}) \times \mathbb{R}^2 \ni D \ni (\alpha, w, m)$. It contains higher-time derivatives of the field φ, which would be referred to, as the unstable field from now on.

$\mathcal{L}_A$ is the sum of the **Proca Lagrangian** and a term $\beta(\partial_\mu A^\mu)^2$, which is reminiscent of that included for the Gupta-Bleuler quantization of the electromagnetic field. This term has an important function and its apperance in the Lagrangian is forced by the stabilization method exposed in 4.2, because besides the **stabilization term**, $\mathcal{L}_{dc}$, only in $\beta(\partial_\mu A^\mu)^2$ there are time derivatives of the vector field A^μ, whose conjugated momenta is precisely the one that controlls the instability of φ. To make this explicit, let us recall that:

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}(\dot{A}^i + \partial_i A^0)^2 - \frac{1}{4}(F^{ij})^2 \tag{97}$$

Furthermore, as previously found for the finite *dof* case, the relation $g^2 = \alpha\beta$ must be satisfied, which implies that $\beta \neq 0$, given that the term $\mathcal{L}_{dc} = -g\partial_\mu\partial^\mu\varphi\partial_\nu A^\nu$ plays a crucial role in the stabilization procedure proposed in [39]. This term is the one that includes the existance of a primary, second class constraint in the theory, that allows the existance of further constraints that eliminate the instability. This will be verified below when computing again the Hessian matrix.

Now, let us begin with the construction of the Hamiltonian in the sense of the extended formalism (See section 3) by writing, as before, an equivalent Lagrangian. Nevertheless, in this case there are two options as oposed to the finite *dof* case. Readily, the spatial components could be expanded by Fourier transform and the stabilization only verified through the temporal dependence[24], or make the field extension of the Dirac's Programme, that was followed before, for finite *dof*. The latter approach is followed for two main reasons: it is the natural extension to the previous approach, and is explicitly covariant (at least in the beginning) and therefore allows to clearly see how the instability is controlled in field theory, while simultaneously allowing for a comprehensive construction of the model. Besides, it turns out to be that this model may have a dynamical instability not related to the Ostrogradsky's instability, and that it is easier to see and control in this latter approach.

Let us define:

$$\psi_\mu(x) := \partial_\mu \varphi(x) \tag{98}$$

and introducing four Lagrange Multipliers λ^μ, with $\mu = 0, 1, 2, 3$.

$$\begin{aligned}\mathcal{L}_{equivalent} = \mathcal{L}_{eq} = -\frac{1}{2}\alpha(\partial_\mu\psi^\mu)^2 + \frac{1}{2}\psi^2 - \frac{1}{2}\omega\varphi^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \beta\frac{1}{2}(\partial_\mu A^\mu)^2 \\ +\frac{m}{2}A^2 - g\partial_\mu\psi^\mu\partial_\nu A^\nu + \lambda^\mu(\psi_\mu - \partial_\mu\varphi)\end{aligned} \tag{99}$$

To change to the Hamiltonian formalism, the following momenta are obtained (i=1,2,3):

$$\begin{aligned}P_{\psi_\mu} &:= \frac{\partial\mathcal{L}_{eq}}{\partial\dot{\psi}^\mu} = -\delta^0_\mu(\alpha\partial_\nu\psi^\nu + g\partial_\nu A^\nu) \\ P_{A_\mu} &:= \frac{\partial\mathcal{L}_{eq}}{\partial\dot{A}^\mu} = \delta^i_\mu(\dot{A}^i + \partial_i A^0) - \delta^0_\mu(\beta\partial_\nu A^\nu + g\partial_\nu\psi^\nu) \\ P_{\lambda_\mu} &:= \frac{\partial\mathcal{L}_{eq}}{\partial\dot{\lambda}^\mu} = 0 \\ P_\varphi &:= \frac{\partial\mathcal{L}_{eq}}{\partial\dot{\varphi}} = -\lambda^0\end{aligned} \tag{100}$$

The basic Poisson brackets are[25]

$$\begin{aligned}\{\psi^\mu(x), P_{\psi_\nu}(y)\} &= \delta^\nu_\mu\delta^3(\vec{x}-\vec{y}) \\ \{A^\mu(x), P_{A_\nu}(y)\} &= \delta^\nu_\mu\delta^3(\vec{x}-\vec{y}) \\ \{\lambda^\mu(x), P_{\lambda_\nu}(y)\} &= \delta^\nu_\mu\delta^3(\vec{x}-\vec{y}) \\ \{\varphi(x), P_\varphi(y)\} &= \delta^3(\vec{x}-\vec{y})\end{aligned} \tag{101}$$

[24] As in section 3, the Ostrogradsky's instability is easily seen in the non-positive definiteness of the Hamiltonian. Specifically when making the Legendre transform, by the definition of conjugated momenta. Since the Hamiltonian formalism takes in special consideration the time, aside from position, it is clear that the stabilization of such instability must take place in the temporal *dof*. However, the precise way to see this, will be explained in the analysis to be presented below.

[25] All brackets of functions of space-time will be treated as equal-time (*i.e.* $x^0 = y^0$). In any case, for the canonical quantization, the equal time commutation relations will be imposed to the Dirac brackets, and causality does not need to be checked by means of a vanishing commutator between fields at space-like distances[50], because the Lagrangian is a Lorentz scalar.

As before with the finite *dof* case, there are some primary constraints in the theory, and therefore, the Legendre transform cannot be carried out in the usual way. First, the complete constraint content of the theory must be known. Because the velocity of the field conjugate to the momenta does not appear explicitly in some of the definitions (100), the velocities $\{\dot{\lambda}^\mu, \dot{\psi}^i, \dot{\varphi}\}$ in a Legendre transform could not be inverted. These are the ones that avoid the Hamiltonization of the theory in the present stage.

Furthermore, if one considers the Hessian matrix, as defined in section 3, it turns out to be singular. In particular, the nullity is exactly 8. However, the null space of this matrix is generated by constraints that come from the specific way in which this Lagrangian $\mathcal{L}_{eq}$ was constructed and not the theory $\mathcal{L}$ itself. In particular, with the definition given above, $\psi_\mu := \partial_\mu \varphi$ and the introduction of the λ^μ Lagrange multipliers, the 8 linearly dependent vectors in the column (or row) space of the Hessian matrix were introduced. Specifically, these come from the fact that there are 8 unexpressible velocities $\{\dot{\lambda}^\mu, \dot{\psi}^i, \dot{\varphi}\}$ and therefore ψ^i, λ^μ are not dynamical (e.g. their momenta do not appear in the Lagrangian, when expressing it in terms of canonical conjugated fields) and for φ, there is a constraint restricting its momentum, P_φ with λ^0. This means that none of these 8 vectors in the null space is really fundamental in the theory $\mathcal{L}$. However, we want to demand a fundamental constraint in the theory, which can be achieved by imposing the Hessian matrix to be singular [26]. Now, if it is already singular for other, non relevant reasons (*i.e.* the form of $\mathcal{L}_{eq}$), one must be careful to identify the important part of the singularity of the Hessian. Other option would be to focus on the Hessian Matrix of $\mathcal{L}$ instead of that for $\mathcal{L}_{eq}$, but with the latter, the common definition of Hessian matrix that has also been used before, and that was cited in section 3, can be used after suitable but straightforward extension to fields. This is:

$$M_{ab} = \frac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{\Phi}^a \partial \dot{\Phi}^b} \tag{102}$$

where the Φ^a are the fields in the theory $\{\psi^\mu, A^\mu, \varphi, \lambda^\mu\}$ and the index a, identifies each one of them in this set.

In particular, we are interested on analysing the following sub-system:

$$\begin{pmatrix} \delta P_{\psi_0} \\ \delta P_{A_0} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{\psi}^0 \partial \dot{\psi}^0} & \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{\psi}^0 \partial \dot{A}^0} \\ \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{A}^0 \partial \dot{\psi}^0} & \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{A}^0 \partial \dot{A}^0} \end{pmatrix} \begin{pmatrix} \delta \dot{\psi}^0 \\ \delta \dot{A}^0 \end{pmatrix} \tag{103}$$

therefore, for our purposes, it is possible to restrict the analysis to a submatrix of the Hessian Matrix defined above, in particular:

$$M_* = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{\psi}^0 \partial \dot{\psi}^0} & \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{\psi}^0 \partial \dot{A}^0} \\ \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{A}^0 \partial \dot{\psi}^0} & \dfrac{\partial^2 \mathcal{L}_{eq}}{\partial \dot{A}^0 \partial \dot{A}^0} \end{pmatrix} \tag{104}$$

[26] Another possibility is to identify the constraints and impose new ones at the level of the equations of motion, but this approach is not useful for the present intentions, because we intend to check the stability of the theory at the level of the Hamiltonian and some basic properties of its domain in the Hilbert space, after quantization.

From (100) it is immediate to see that

$$M_* = \begin{pmatrix} -\alpha & -g \\ -g & -\beta \end{pmatrix} \tag{105}$$

and the determinant is $det(M_*) = \alpha\beta - g^2$.

According to the stabilization procedure [39], presented in section 3, it is desired that this subsytem contains a constraint, which implies that we must impose $det(M_*) = \alpha\beta - g^2 = 0$, which again gives:

$$g^2 \stackrel{!}{=} \alpha\beta \tag{106}$$

From the requirement $g^2 \stackrel{!}{=} \alpha\beta$ it follows that there is a quite important primary constraint in the theory:

$$P_{A_0} \stackrel{!}{=} \frac{g}{\alpha} P_{\psi_0} \tag{107}$$

From the previous discussion, regarding the 8 primarly unexpressible velocities $\{\dot{\lambda}^\mu, \dot{\psi}^i, \dot{\varphi}\}$ in (100) and the imposed constraint $P_{A_0} = \frac{g}{\alpha} P_{\psi_0}$ (107), there are exactly nine primary constraints in the theory, which are:

$$\begin{aligned} \zeta_0 &:= P_{A_0} - \frac{g}{\alpha} P_{\psi_0} \approx 0 \\ \zeta_i &:= P_{\psi_i} \approx 0 \\ \Delta_\mu &:= P_{\lambda_\mu} \approx 0 \\ \Lambda^0 &:= P_\varphi + \lambda^0 \approx 0 \end{aligned} \tag{108}$$

It must be noted, as before, that the $\approx$ symbol, means in this context that only the solution of the equations of motion implied by the complete theory (which involves the knowledge of all the constraints present in the theory), satisfy these constraints, or in other words, the canonical fields evolve in phase space, of all possible field configurations $\mathcal{D}(\Phi)$, constrained to the surface defined by all the equations between the fields (108). It is then clear using $\approx$ that (108) are not identically satisfied for every field configuration.

However, by the theorem cited in (3.2) it is possible to Hamiltonize $\mathcal{L}$, in a certain set of equivalent canonically conjugated fields, which will allow for an easier interpretation. Because the equivalent new canonical fields can be built to evolve with certain redefined Poisson brackets (Dirac brackets) on a new phase space of possible field configurations $\mathcal{D}(\Phi)_D$, it is possible to make it, such that the new fields evolve apparently not constrained *i.e.* the equivalent constraints are satisfied identically at every point, $\approx \to \equiv_D$. For this reason, this Hamiltonian is usually called the physical Hamiltonian and is what we set to build in what follows.

4.3.1 Hamiltonization of the field theory in the extended formalism

Let us define the following function with some domain in the phase space of field configurations $\mathcal{D}(\Phi)$:

$$f(\Phi) = P_\varphi \dot{\varphi} + P_{\psi_\mu} \dot{\psi}^\mu + P_{A_\mu} \dot{A}^\mu + P_{\lambda_\mu} \dot{\lambda}^\mu - \mathcal{L}_{eq} \tag{109}$$

Now, let us make use of the definition of canonical momenta (100) being careful to identify $\{\dot{\lambda}^\mu, \dot{\psi}^i, \dot{\varphi}\}$ as not properly defined, because in the present stage, these velocities cannot be inverted in terms of canonical variables. Furthermore, it must be considered that the constraint given by $P_{A_0} = \frac{g}{\alpha} P_{\psi_0}$, which was inserted in the theory $\mathcal{L}$ by demanding $g^2 = \alpha\beta$, makes $\dot{A}_0$ and $\dot{\psi}_0$ linearly dependent. Thus, only one of the latter can be inverted in terms of either P_{A_0} or P_{ψ_0} in (109). In this case, $\dot{A}_0$ is selected as primarily unexpressible velocity and the $\dot{\psi}_0$ is inverted by the definition of momenta[27]

$$\begin{aligned} f(\Phi) = \frac{1}{2} P_{A_i} P_{A_i} - \frac{1}{2\alpha} P_{\psi_0}^2 + \frac{1}{4} F_{ij} F^{ij} - \frac{m}{2} A_\mu A^\mu - \frac{1}{2} \psi_\mu \psi^\mu + \frac{1}{2} w \varphi^2 + P_\varphi \psi_0 - P_{A_i} \partial_i A_0 \\ - P_{\psi_0} (\partial_i \psi^i + \frac{g}{\alpha} \partial_i A^i) - \lambda^i (\psi_i - \partial_i \varphi) + \dot{A}^0 (P_{A_0} - \frac{g}{\alpha} P_{\psi_0}) + \dot{\varphi} (P_\varphi + \lambda^0) + \dot{\lambda}^\mu (P_{\lambda_\mu}) + \dot{\psi}^i (P_{\psi_i}) \end{aligned} \tag{110}$$

This function can be identified with a Hamiltonian only in the extended formalism (See section 3) by the definition of the primarily unexpressible velocities:

$$\begin{aligned} V_\varphi &:= \dot{\varphi} \\ V_A^0 &:= \dot{A}^0 \\ V_\lambda^\mu &:= \dot{\lambda}^\mu \\ V_\psi^i &:= \dot{\psi}^i \end{aligned} \tag{111}$$

$$\begin{aligned} \mathcal{H}^{(1)}(\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi) = \frac{1}{2} P_{A_i} P_{A_i} - \frac{1}{2\alpha} P_{\psi_0}^2 + \frac{1}{4} F_{ij} F^{ij} - \frac{m}{2} A_\mu A^\mu \\ - \frac{1}{2} \psi_\mu \psi^\mu + \frac{1}{2} w \varphi^2 + P_\varphi \psi_0 - P_{A_i} \partial_i A_0 - P_{\psi_0} (\partial_i \psi^i + \frac{g}{\alpha} \partial_i A^i) - \lambda^i (\psi_i - \partial_i \varphi) \\ + V_A^0 (P_{A_0} - \frac{g}{\alpha} P_{\psi_0}) + V_\varphi (P_\varphi + \lambda^0) + V_\lambda^\mu (P_{\lambda_\mu}) + V_\psi^i (P_{\psi_i}) \end{aligned}$$

Now, it is clear that in (110), the unexpressible velocities $V_\varphi, V_A^0, V_\lambda^\mu, V_\psi^i$, which multiply the constraints that can be identified from (108), are in fact Lagrange multipliers of the primary constraints of the theory[28]. The latter is in accordance with the fact that the unexpressible velocities are, meanwhile, undefined functions of the other canonical fields of the theory. For this reason, the index (1) is introduced to the extended Hamiltonian, to denote that this is the first stage Hamiltonian. It is

[27] This independence of the result upon the selection of primarily unexpressible velocities is expected, however it may not be easily seen. Fortunately, there is a theorem supporting this claim. One of many references is chapter 2 of [45].

[28] The introduction of this interpretation as Lagrange multipliers is due to P.M. Dirac [44, 45, 46, 47]

therefore mandatory to define the functions $V_\varphi, V_A^0, V_\lambda^\mu, V_\psi^i$, to have a well defined Hamiltonian as a basis for the canonical quantization. This is achieved by demanding the temporal stability of the constraints, which also leads to find the complete constraint structure of the theory.

The following is the Hamiltonian in first stage, with the corresponding Poisson brackets between constraints [29].

$$
\begin{aligned}
\mathcal{H}^{(1)}(\Phi) &= \mathcal{H}_0 + V_A^0\zeta_0 + V_\varphi\Lambda^0 + V_\lambda^\mu\Delta_\mu + V_\psi^i\zeta_i \\
\mathcal{H}_0(\Phi) &= \frac{1}{2}P_{A_i}P_{A_i} - \frac{1}{2\alpha}P_{\psi_0}^2 + \frac{1}{4}F_{ij}F^{ij} - \frac{m}{2}A_\mu A^\mu - \frac{1}{2}\psi_\mu\psi^\mu + \frac{1}{2}w\varphi^2 + P_\varphi\psi_0 - P_{A_i}\partial_i A_0 \\
&- P_{\psi_0}(\partial_i\psi^i + \frac{g}{\alpha}\partial_i A^i) - \lambda^i(\psi_i - \partial_i\varphi) \\
\zeta_0 &:= P_{A_0} - \frac{g}{\alpha}P_{\psi_0} \approx 0 \\
\zeta_i &:= P_{\psi_i} \approx 0 \\
\Delta_\mu &:= P_{\lambda_\mu} \approx 0 \\
\Lambda^0 &:= P_\varphi + \lambda^0 \approx 0
\end{aligned}
$$

$$
\begin{array}{lll}
\{\zeta_0(x), \zeta_i(y)\} = 0 & \{\zeta_0(x), \Delta_\mu(y)\} = 0 & \{\zeta_0(x), \Lambda^0(y)\} = 0 \\
\{\zeta_i(x), \Delta_\mu(y)\} = 0 & \{\zeta_i(x), \Lambda^0(y)\} = 0 & \{\Delta_\mu(x), \Lambda^0(y)\} = -\delta_\mu^0\delta^{(3)}(\vec{x}-\vec{y})
\end{array}
\tag{112}
$$

It is worth noting the fields on which the Hamiltonian depends $\mathcal{H}^{(1)}(\Phi) := \mathcal{H}^{(1)}(\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi)$, but the real number of degrees of freedom (extending the concept to fields), cannot yet be known, until the other constraints in the theory become evident.

4.3.2 Constraint content of the theory

In order to determine the unexpressible velocities and completely define the theory, we impose ($\overset{!}{=}$) that the constraints (Π) are not dynamical, *i.e.* their Poisson bracket with the Hamiltonian in first stage is set to zero:

$$\dot{\Pi}(x) = \int d^3y\{\Pi(x), \mathcal{H}^{(1)}(y)\} \overset{!}{=} 0 \tag{113}$$

The evolution of each of the primary constraints imply, if the system is consistent, one of three possibilities. Readily, they can fix one of the Lagrange multipliers $V_\varphi, V_A^0, V_\lambda^\mu, V_\psi^i$, or can be satisfied identically by a known result (such as another constraint already known at this stage), or finally, it can imply another weak equation defined by the symbol: $(\overset{!}{=} 0) \to (\overset{!}{\approx} 0) \to \approx 0$, in the sense that it is satisfied by the field configurations that are solution to the equations of motion, and are not identically satisfied in every point of the phase space of fields configurations.

[29] It is worth to note that in the computations to be considered, a very reasonable assumption will be taken. When integrating by parts, the canonical fields accompanying the derivatives of the deltas will be assumed to be appropiate test functions, *i.e.* we assume they vanish outside some finite interval and are smooth or at least $\mathcal{C}^2$. Both assumptions are in fact considered since the very beginning, when the Lagrangian $\mathcal{L}$ was written with up to two derivatives, and when we claimed that the equations of motion are reached by means of a stationary action principle, with vanishing fields at infinity, allowing boundary terms to vanish.

These new constraints are secondary and only differ from the primary by the fact that these are implied by the momenta, while the former arise in the equations of motion. As pointed out in [44], the difference among these constraints is not relevant, because another equivalent Lagrangian could be written such that another primary-secondary classification could be reached. The results of the evolution of the constraints are the following:

$$\dot{\zeta}_0(x) = \int d^3y\{\zeta_0(x), \mathcal{H}^{(1)}(y)\} \stackrel{!}{=} 0$$

$$\dot{\zeta}_0(x) = mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \stackrel{!}{=} 0$$

$\Rightarrow$ Define it as a secondary constraint:

$$\Xi_1 := mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0 \tag{114}$$

It is worth noting that the constraint Ξ_1 includes a relation with the secondary constraint that appears in the Proca Lagrangian, and it turns out to be, that this constraint is of major importance to control the Ostrogradsky's instability.

$$\dot{\zeta}_i(x) = \int d^3y\{\zeta_i(x), \mathcal{H}^{(1)}(y)\} \stackrel{!}{=} 0$$

$$\dot{\zeta}_i(x) = -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \stackrel{!}{=} 0$$

$\Rightarrow$ Define the following three secondary constraints:

$$\Xi_{2_i} := -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \approx 0 \tag{115}$$

$$\dot{\Delta}_\mu(x) = \int d^3y\{\Delta_\mu(x), \mathcal{H}^{(1)}(y)\} \stackrel{!}{=} 0$$

$$\dot{\Delta}_\mu(x) = -\delta^0_\mu V_\varphi + \delta^i_\mu(-\psi^i - \partial_i\varphi) \stackrel{!}{=} 0$$

$\Rightarrow$ The following unexpressible velocity is defined:

$$V_\varphi(x) = 0$$

And define the following three secondary constraints:

$$\Xi_{3_i} := -\psi^i(x) - \partial_i\varphi(x) \approx 0 \tag{116}$$

$$\begin{aligned}
\dot{\Lambda}^0(x) &= \int d^3y\{\Lambda^0(x),\mathcal{H}^{(1)}(y)\} \stackrel{!}{=} 0 \\
\dot{\Lambda}^0(x) &= -w\varphi(x) + \partial_i\lambda^i(x) + V^0_\lambda \stackrel{!}{=} 0 \\
&\Rightarrow \text{ The following unexpressible velocity is defined:} \\
V^0_\lambda(x) &= w\varphi(x) - \partial_i\lambda^i(x) \qquad (117)
\end{aligned}$$

Therefore, after imposing the stability of the 9 primary constraints, another 7 secondary constraints arise and two primarly unexpressible velocities, or Lagrange multipliers, are fixed. It is therefore clear that the system is not yet well defined, because there still remain Lagrange multipliers to be found. The solution comes again from the imposition of temporal stability (or constraint conservation in time) of the 7 secondary constraints. This is, let us impose again (113), but this time the Hamiltonian must be written including the new constraints with additional Lagrange multipliers and making explicit the known dependence of the Lagrange multipliers which have been already found by the previous procedure.

The Hamiltonian on second stage ($\mathcal{H}^{(2)}$), is:

$$\mathcal{H}^{(2)}(\Phi) = \mathcal{H}_0 + V^0_A\zeta_0 + V_\varphi(\Phi)\Lambda^0 + V^0_\lambda(\Phi)\Delta_0 + V^i_\lambda\Delta_i + V^i_\psi\zeta_i + \chi_1\Xi_1 + \chi^i_2\Xi_{2_i} + \chi^i_3\Xi_{3_i} \qquad (118)$$

With $V_\varphi(\Phi) = 0$ and $V^0_\lambda(\Phi) = w\varphi(x) - \partial_i\lambda^i(x)$, as was found before and the constraints as previously defined. It is important to note that new Lagrangian multipliers have been introduced $\{\chi_1, \chi^i_2, \chi^i_3\}$ and they must be possible to be written in terms of the canonical variables, in order for the theory to be consistent. The outcome of this computation leads again to three possibilities as described above.

The equal-time Poisson brackets including the new constraints are:

$$\begin{array}{ccc}
\{\zeta_0(x),\zeta_i(y)\} = 0 & \{\zeta_0(x),\Delta_\mu(y)\} = 0 & \{\zeta_0(x),\Lambda^0(y)\} = 0 \\
\{\zeta_0(x),\Xi_1(y)\} = -(m+\frac{g^2}{\alpha^2})\delta^{(3)}(\vec{x}-\vec{y}) & \{\zeta_0(x),\Xi_{2_i}(y)\} = 0 & \{\zeta_0(x),\Xi_{3_i}(y)\} = 0 \\
\{\zeta_i(x),\Delta_\mu(y)\} = 0 & \{\zeta_i(x),\Lambda^0(y)\} = 0 & \{\zeta_i(x),\Xi_1(y)\} = 0 \\
\{\zeta_i(x),\Xi_{2_j}(y)\} = \delta_{ij}\delta^{(3)}(\vec{x}-\vec{y}) & \{\zeta_i(x),\Xi_{2_j}(y)\} = \delta_{ij}\delta^{(3)}(\vec{x}-\vec{y}) & \{\Delta_\mu(x),\Lambda^0(y)\} = -\delta^0_\mu\delta^{(3)}(\vec{x}-\vec{y}) \\
\{\Delta_\mu(x),\Xi_1(y)\} = 0 & \{\Delta_\mu(x),\Xi_{2_i}(y)\} = \delta_{ij}\delta^{(3)}(\vec{x}-\vec{y}) & \{\Delta_\mu(x),\Xi_{3_i}(y)\} = 0 \\
\{\Lambda^0(x),\Xi_1(y)\} = 0 & \{\Lambda^0(x),\Xi_{2_i}(y)\} = 0 & \{\Lambda^0(x),\Xi_{3_i}(y)\} = \partial^{(y)}_i(\delta^{(3)}(\vec{y}-\vec{x})) \\
\{\Xi_1(x),\Xi_{2_i}(y)\} = \frac{g}{\alpha}\partial^{(y)}_i(\delta^{(3)}(\vec{y}-\vec{x})) & \{\Xi_1(x),\Xi_{3_i}(y)\} = \frac{g}{\alpha}\partial^{(y)}_i(\delta^{(3)}(\vec{y}-\vec{x})) & \{\Xi_{2_i}(x),\Xi_{3_j}(y)\} = 0
\end{array} \qquad (119)$$

Where $\partial^{(y)}_i(\delta^{(3)}(\vec{y}-\vec{x}))$ will make sense after considering the Poisson brackets integrated over $\{\vec{x}\in\mathbb{R}^3\}$ or $\{\vec{y}\ \in\mathbb{R}^3\}$, and $\partial^{(x)}_i$ denotes $\frac{\partial}{\partial x^i}$.

With the equations (119) it is easy to compute the evolution of the new constraints and demand their conservation. This is:

$$\begin{aligned}
\dot{\Xi}_1(x) &= \int d^3y\{\Xi_1(x), \mathcal{H}^{(2)}(y)\} \overset{!}{=} 0 \\
\dot{\Xi}_1(x) &= \frac{g}{\alpha}(-w\varphi(x) + \partial_i^{(x)}\Lambda^i + \frac{1}{\alpha}P_{\psi_0}(x) + \partial_i\psi^i(x) + \frac{g}{\alpha} + \frac{\alpha}{g}m)\partial_i A^i(x)) + \partial_i\partial_i P_{\psi_0} \\
&+ \int d^3y\left(V_A^0\left(m + \frac{g^2}{\alpha^2}\right)\delta^{(3)}(\vec{x}-\vec{y}) + \frac{g}{\alpha}(\Xi_{2_i} + \Xi_{3_i})\partial_i^{(y)}\delta^{(3)}(\vec{y}-\vec{x})\right) \overset{!}{=} 0
\end{aligned} \tag{120}$$

$$\begin{aligned}
\dot{\Xi}_{2_i}(x) &= \int d^3y\{\Xi_{2_i}(x), \mathcal{H}^{(2)}(y)\} \overset{!}{=} 0 \\
\dot{\Xi}_{2_i}(x) &= V_\lambda^i + V_\psi^i + \frac{g}{\alpha}\int d^3y\Xi_1\partial_i^{(x)}(\delta^{(3)}(\vec{x}-\vec{y})) + \partial_i(\psi_0(x) - P_\varphi(x)) \overset{!}{=} 0
\end{aligned} \tag{121}$$

$$\begin{aligned}
\dot{\Xi}_{3_i}(x) &= \int d^3y\{\Xi_{3_i}(x), \mathcal{H}^{(2)}(y)\} \overset{!}{=} 0 \\
\dot{\Xi}_{3_i}(x) &= V_\psi^i + \int d^3y\left(V_\varphi + \frac{g}{\alpha}\Xi_1 - \psi_0\right)\partial_i^{(x)}(\delta^{(3)}(\vec{x}-\vec{y})) \overset{!}{=} 0
\end{aligned} \tag{122}$$

Equations (120, 121, 122) let us know that by fixing the Lagrange multipliers that they involve, the constraints are conserved. Therefore, the process has ended because no more constraints appear and in fact, all the constraints in the system have been written in equation (119).

From (119) it is also possible to identify a very important fact. There is no constraint that has vanishing Poisson bracket with all the other constraints in the theory. Hence, all the set of constraints is of second-class and the dynamics of the system is well defined[30]. This also allows to address the question if all the Lagrange multipliers are defined by the process above. The answer is, yes, but from the previous expressions they can be very difficult to find. This answer will be given in a better approach using the inverse of the Dirac matrix, which will be built in short.

From the previous computations it can be ascertained that the following is the Hamiltonian of the theory. The primarily unexpressible velocities no longer appear, because they have been already defined by means of the imposition of conservation of constraints in time. This is emphasised by making explicit their dependence on Φ, which denotes all the canonical fields.

[30]With well defined it is meant that there are no Gauge symmetries that must be fixed, in order to obtain equations of motion with unique solution to an initial value-problem. If the set were of first class, then, the latter would not be the case [44, 46, 45].

$$
\begin{aligned}
\mathcal{H}(\Phi) &= \mathcal{H}_0 + V_A^0(\Phi)\zeta_0 + V_\varphi(\Phi)\Lambda^0 + V_\lambda^0(\Phi)\Delta_0 + V_\lambda^i(\Phi)\Delta_i + V_\psi^i(\Phi)\zeta_i + \chi_1(\Phi)\Xi_1 + \chi_2^i(\Phi)\Xi_{2_i} \\
&+ \chi_3^i(\Phi)\Xi_{3_i} \\
\mathcal{H}_0(\Phi) &= \frac{1}{2}P_{A_i}P_{A_i} - \frac{1}{2\alpha}P_{\psi_0}^2 + \frac{1}{4}F_{ij}F^{ij} - \frac{m}{2}A_\mu A^\mu - \frac{1}{2}\psi_\mu\psi^\mu + \frac{1}{2}w\varphi^2 + P_\varphi\psi_0 - P_{A_i}\partial_i A_0 \\
&- P_{\psi_0}(\partial_i\psi^i + \frac{g}{\alpha}\partial_i A^i) - \lambda^i(\psi_i - \partial_i\varphi) \\
\zeta_0(x) &:= P_{A_0}(x) - \frac{g}{\alpha}P_{\psi_0}(x) \approx 0 \\
\zeta_i(x) &:= P_{\psi_i}(x) \approx 0 \\
\Delta_\mu(x) &:= P_{\lambda_\mu}(x) \approx 0 \\
\Lambda^0(x) &:= P_\varphi(x) + \lambda^0(x) \approx 0 \\
\Xi_1(x) &:= mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0 \\
\Xi_{2_i}(x) &:= -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \approx 0 \\
\Xi_{3_i}(x) &:= -\psi^i(x) - \partial_i\varphi(x) \approx 0
\end{aligned}
$$

$\Rightarrow$ Equations of motion, or evolution of the field configurations are given by:

$$
\dot{\Phi}(x) = \int d^3y\{\Phi(x), \mathcal{H}(\Phi(y))\}\ ,\ \Phi(x) =: \Phi \in \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\} \tag{123}
$$

Where the weak equalities $\approx$ cannot be taken as strong or satisfied at every point in phase space, but only by those fields satisfying the equations of motion [44, 45, 46, 47]. This is, the fields written in (123) are functions representing all possible configurations in phase space. Only after the Poisson brackets have been calculated with a Hamiltonian including the weakly vanishing constraints ($\mathcal{H}$), and after we explicitly demand that the fields Φ are no longer general fields in configuration space, but restrict them to satisfy the equations of motion, is that the equalities ($\approx$) can be considered in the strong sense ($\approx\rightarrow=$)[31].

Now that all the constraints in the theory have been determined, it is important to count how many really dynamical variables has the system. From $\mathcal{H}(\Phi) := \mathcal{H}(\psi^\mu, P_{\psi_\mu}, A^\mu,\ P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi)$, it seems that there are 26 canonical fields, and from (123) it can be seen that there are 16 constraints among them, then, the number of degrees of freedom of the theory is 10. Furthermore, these 10 *dof*s are canonical fields in phase space of possible fields configurations, thus, there are 5 "fields" (not momenta), and this is in accordance with the original formulation of the model (95), also with 5 fields $\mathcal{L} = \mathcal{L}(\varphi, A^\mu)$.

Nevertheless, this is a very rough approach, because the 5 *dof*s we are referring to, in the current Hamiltonian form of the theory, are not really distinguishable from those in the set $\{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$. This is due to the fact that they are constrained by the 16 relations $\{\zeta_0, \zeta_i, \Delta_\mu, \Lambda^0, \Xi_1,$

[31] The last statement is in fact reduntant, but helps for clarity, because once one imposes the equations of motion in its full structure, *i.e.* containing a evolution restricted to the constraints, the constraints are identically satisfied. They are solutions to the constraints.

$\Xi_{2_i}, \Xi_{3_i}\}$ defining a dynamical hypersurface in phase space of field configurations. It is precisely on this surface that some relation between the 26 fields commented above, give rise to 10 fields that truly evolve freely by some Hamiltonian function not yet known but equivalent to that in (123). Furthermore, these 10 fields must be canonical conjugated pairs, and is from this set that the dofs counting make sense, or is at least much more intuitive. The remaining 16 dofs in (123) take the form of the new, equivalent, equations of constraints. This is, in a new set of fields satisfying the properties just described, there are exactly 16 canonical variables which are completely independent of the other 10 that evolve freely on the mentioned hypersurface. In fact, this hypersurface in phase space is defined by these 16 fields (Ω_f, $\mathbb{N} \ni f = 1, \ldots, 16$) which are equations of constraint vanishing identically over the surface that they define (Phase space of possible field configuartions $\supset$ Hypersurface=: $\{\Omega_f \equiv 0, \forall f\}$). This description, and the existance of such equivalent system of fields, Hamiltonian and constraints, is guaranteed by a theorem that was cited in (3.4), remark (3.2)[44, 45, 46, 47].

The construction of such an equivalent system is important to truly understand the dynamics implied by this higher derivative theory, because in a complicated system like (123), it is non-trivial to identify the properties in which we are interested,*i.e.* to verify the presence or not of the Ostrogradsky's instability. With this on mind, we set to build the Dirac brackets (See 3.4).

4.3.3 Dirac Brackets of the theory

The reason to approach the search for the equivalent system precisely described in section (3.2) (See also section 3.4) by means of Dirac Brackets is the following: We are set to build an equivalent Hamiltonian that determines the evolution of 10 canonically conjugated fields, yet to be defined, on a surface defined by 16 fields, whose form must also be found. The way to construct them, comes from the properties they have, which can be expressed in terms of their Dirac brackets.

Some facts are known beforehand because of theorem (3.2): the variables must be separated into two disjoint sets of pairs of canonically conjugated variables $\{\omega_{g=1,\ldots,10}, \Omega_{f=1,\ldots,16}\}$. The set $\{\Omega_f\}_{f=1,\ldots,16}$ is such that $\Omega_f \equiv 0, \forall f$ and the Dirac bracket built with these constraints $\{\Omega_{f=1,\ldots,16}\}$ ($\in \{\zeta_0, \zeta_i, \Delta_\mu, \Lambda^0, \Xi_1, \Xi_{2_i}, \Xi_{3_i}\}$) equals a re-defined Poisson bracket only in the set $\{\omega_{g=1,\ldots,10}\}$, *i.e.*:

$$\{\cdot,\cdot\}^{\omega,\Omega}_{DiracB(\Omega)} = \{\cdot,\cdot\}^{\omega}_{PoissonB} \tag{124}$$

The last equation also implies that $\{\omega_g\}$ is a set of pairs of canonnically conjugated fields, independent of the sector $\{\Omega_f\}$. Finally, as was cited in (3.4) from [45], the problem can be stated as the following change:

$$\begin{array}{rcl}
\mathcal{H}(x) & = & \mathcal{H}(\Phi(x)) \\
\dot{\Phi}(x) & = & \int d^3y\{\Phi(x),\mathcal{H}(y)\} \ , \Pi \approx 0 \\
\Phi & \in & \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\} \text{ are canonical fields.} \\
\Pi & \in & \{\zeta_0, \zeta_i, \Delta_\mu, \Lambda^0, \Xi_1, \Xi_{2_i}, \Xi_{3_i}\} \text{ are relations between } \Phi \text{ constraining the evolution.} \\
 & \downarrow & \\
\tilde{\mathcal{H}}(x) & = & \tilde{\mathcal{H}}(\omega(x),\Omega(x)) \\
\dot{\omega}(x) & = & \int d^3y\{\omega(x),\tilde{\mathcal{H}}(y)\}^{\omega,\Omega}_{DiracB(\Omega)} \ , \Omega(x) \equiv 0 \\
\{\omega,\Omega\} & & \text{are canonical fields, but only } \{\omega\} \text{ are dynamical.}
\end{array} \tag{125}$$

From the last equation it becomes evident that being $\Omega \equiv 0$, only ω are dynamical, therefore, they are treated as physical. Hence, following the notation in [45], and as was defined in section 3.4, we are interested in the physical Hamiltonian:

$$\mathcal{H}_{phys}(\omega) = \tilde{\mathcal{H}}|_{\Omega=0} \tag{126}$$

Such that

$$\begin{array}{rcl}
\dot{\omega}(x) & = & \int d^3y\{\omega(x),\tilde{\mathcal{H}}(y)\}^{\omega,\Omega}_{DiracB(\Omega)} \ , \Omega(x) \equiv 0 \\
 & \downarrow & \\
\dot{\omega}(x) & = & \int d^3y\{\omega(x),\mathcal{H}_{phys}(y)\}_{PoissonB} \ , \Omega(x) \equiv 0
\end{array} \tag{127}$$

The latter is justified by the following theorem, that allows to make constraints identically zero inside the Dirac brackets[32] which is justified in the Appendix E of [45]:

$$\{\cdot,\{\Omega\}\}_{DiracB(\Omega)} = \{\Omega\} \tag{128}$$

To put this discussion in context, let us identify the upper section of (125) with the complete theory written in (123). We are therefore, set to find the equivalent lower section of (125), which is better suited to analyse the properties of the classical Hamiltonian, as was described in the introduction of section (4).

The equal-time Dirac brackets for any pair of functions $\mathcal{F}(x), \mathcal{G}(y)$ of space-time, are (See 3.4, or [43, 44, 45, 46, 47]):

$$\{\mathcal{F}(x),\mathcal{G}(y)\}_{D(\Pi)} = \{\mathcal{F}(x),\mathcal{G}(y)\} - \int d^3z d^3z' \left(\{\mathcal{F}(x),\Pi_a(z)\}\{\Pi(z),\Pi(z')\}^{-1}_{a,a'}\{\Pi_{a'}(z'),\mathcal{G}(y)\}\right) \tag{129}$$

[32] This is in fact the way to build the Dirac brackets.

Where $\Pi_a(x)$ denotes the constraints between fields $\{\zeta_0, \zeta_i, \Delta_\mu, \Lambda^0, \Xi_1, \Xi_{2_i}, \Xi_{3_i}\}$, with a running through all of them and $\{\Pi(z), \Pi(z')\}^{-1}_{a,a'}$ is the component (a, a') of the inverse of the Dirac matrix.

The Dirac matrix is defined as a matrix whose entries are the Poisson brackets between constraints, for all the constraints in the theory. As was pointed out before, the set of constraints is of second-class, therefore, the Dirac matrix is non-singular and the inverse can be computed. Since $\{\Pi(z), \Pi(z')\}^{-1}_{a,a'}$ acts like an integration kernel in (129), we must only invert the operators inside the matrix inverse (See 3.4, or [44, 45, 46, 47]). Dropping the $D(\Omega)$ notation for the Dirac brackets, the Dirac matrix is:

$$\mathcal{D}(x,y) = \begin{pmatrix}
\{\zeta_0(x),\zeta_0(y)\} & \{\zeta_0(x),\zeta_j(y)\} & \{\zeta_0(x),\Delta_0(y)\} & \{\zeta_0(x),\Delta_j(y)\} & \{\zeta_0(x),\Lambda^0(y)\} & \{\zeta_0(x),\Xi_1(y)\} & \{\zeta_0(x),\Xi_{2_j}(y)\} & \{\zeta_0,\Xi_{3_j}(y)\} \\
\{\zeta_i(x),\zeta_0(y)\} & \{\zeta_i(x),\zeta_j(y)\} & \{\zeta_i(x),\Delta_0(y)\} & \{\zeta_i(x),\Delta_j(y)\} & \{\zeta_i(x),\Lambda^0(y)\} & \{\zeta_i(x),\Xi_1(y)\} & \{\zeta_i(x),\Xi_{2_j}(y)\} & \{\zeta_i(x),\Xi_{3_j}(y)\} \\
\{\Delta_0(x),\zeta_0(y)\} & \{\Delta_0(x),\zeta_j(y)\} & \{\Delta_0(x),\Delta_0(y)\} & \{\Delta_0(x),\Delta_j(y)\} & \{\Delta_0(x),\Lambda^0(y)\} & \{\Delta_0(x),\Xi_1(y)\} & \{\Delta_0(x),\Xi_{2_j}(y)\} & \{\Delta_0(x),\Xi_{3_j}(y)\} \\
\{\Delta_i(x),\zeta_0(y)\} & \{\Delta_i(x),\zeta_j(y)\} & \{\Delta_i(x),\Delta_0(y)\} & \{\Delta_i(x),\Delta_j(y)\} & \{\Delta_i(x),\Lambda^0(y)\} & \{\Delta_i(x),\Xi_1(y)\} & \{\Delta_i(x),\Xi_{2_j}(y)\} & \{\Delta_i(x),\Xi_{3_j}(y)\} \\
\{\Lambda^0(x),\zeta_0(y)\} & \{\Lambda^0(x),\zeta_j(y)\} & \{\Lambda^0(x),\Delta_0(y)\} & \{\Lambda^0(x),\Delta_j(y)\} & \{\Lambda^0(x),\Lambda^0(y)\} & \{\Lambda^0(x),\Xi_1(y)\} & \{\Lambda^0(x),\Xi_{2_j}(y)\} & \{\Lambda^0(x),\Xi_{3_j}(y)\} \\
\{\Xi_1(x),\zeta_0(y)\} & \{\Xi_1(x),\zeta_j(y)\} & \{\Xi_1(x),\Delta_0(y)\} & \{\Xi_1(x),\Delta_j(y)\} & \{\Xi_1(x),\Lambda^0(y)\} & \{\Xi_1(x),\Xi_1(y)\} & \{\Xi_1(x),\Xi_{2_j}(y)\} & \{\Xi_1(x),\Xi_{3_j}(y)\} \\
\{\Xi_{2_i}(x),\zeta_0(y)\} & \{\Xi_{2_i}(x),\zeta_j(y)\} & \{\Xi_{2_i}(x),\Delta_0(y)\} & \{\Xi_{2_i}(x),\Delta_j(y)\} & \{\Xi_{2_i}(x),\Lambda^0(y)\} & \{\Xi_{2_i}(x),\Xi_1(y)\} & \{\Xi_{2_i}(x),\Xi_{2_j}(y)\} & \{\Xi_{2_i}(x),\Xi_{3_j}(y)\} \\
\{\Xi_{3_i}(x),\zeta_0(y)\} & \{\Xi_{3_i}(x),\zeta_j(y)\} & \{\Xi_{3_i}(x),\Delta_0(y)\} & \{\Xi_{3_i}(x),\Delta_j(y)\} & \{\Xi_{3_i}(x),\Lambda^0(y)\} & \{\Xi_{3_i}(x),\Xi_1(y)\} & \{\Xi_{3_i}(x),\Xi_{2_j}(y)\} & \{\Xi_{3_i}(x),\Xi_{3_j}(y)\}
\end{pmatrix} \quad (130)$$

Note that this is a 16×16 matrix, but the spatial indices have been denoted by $i, j = 1, 2, 3$. Thus, (130) is in fact built by blocks.

With the Poisson brackets between constraints given in equation (119), replacing in (123), the Dirac Matrix for the system is obtained:

$$\mathcal{D}(x,y) = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & -\left(m+\frac{g^2}{\alpha^2}\right)\delta^{(3)}(\vec{x}-\vec{y}) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \delta_{ij}\delta^{(3)}(\vec{x}-\vec{y}) & \delta_{ij}\delta^{(3)}(\vec{x}-\vec{y}) \\
0 & 0 & 0 & 0 & -\delta^{(3)}(\vec{x}-\vec{y}) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \delta_{ij}\delta^{(3)}(\vec{x}-\vec{y}) & 0 \\
0 & 0 & \delta^{(3)}(\vec{x}-\vec{y}) & 0 & 0 & 0 & 0 & \partial_j^{(y)}(\delta^{(3)}(\vec{y}-\vec{x})) \\
\left(m+\frac{g^2}{\alpha^2}\right)\delta^{(3)}(\vec{x}-\vec{y}) & 0 & 0 & 0 & 0 & 0 & \frac{g}{\alpha}\partial_j^{(y)}(\delta^{(3)}(\vec{y}-\vec{x})) & \frac{g}{\alpha}\partial_j^{(y)}(\delta^{(3)}(\vec{y}-\vec{x})) \\
0 & -\delta_{ij}\delta^{(3)}(\vec{y}-\vec{x}) & 0 & -\delta_{ij}\delta^{(3)}(\vec{y}-\vec{x}) & 0 & -\frac{g}{\alpha}\partial_i^{(x)}(\delta^{(3)}(\vec{x}-\vec{y})) & 0 & 0 \\
0 & -\delta_{ij}\delta^{(3)}(\vec{y}-\vec{x}) & 0 & 0 & \partial_i^{(x)}(\delta^{(3)}(\vec{x}-\vec{y})) & -\frac{g}{\alpha}\partial_i^{(x)}(\delta^{(3)}(\vec{x}-\vec{y})) & 0 & 0
\end{pmatrix} \quad (131)$$

With the Dirac matrix (130), (129) and the constraints,

$$\begin{aligned}
\zeta_0(x) &:= P_{A_0}(x) - \frac{g}{\alpha}P_{\psi_0}(x) \approx 0 \\
\zeta_i(x) &:= P_{\psi_i}(x) \approx 0 \\
\Delta_\mu(x) &:= P_{\lambda_\mu}(x) \approx 0 \\
\Lambda^0(x) &:= P_\varphi(x) + \lambda^0(x) \approx 0 \\
\Xi_1(x) &:= mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0 \\
\Xi_{2_i}(x) &:= -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \approx 0 \\
\Xi_{3_i}(x) &:= -\psi^i(x) - \partial_i\varphi(x) \approx 0
\end{aligned}$$

the following are the Dirac brackets for the theory (123):

$$
\begin{aligned}
\{\mathcal{F}(x),\mathcal{G}(y)\}_{D(\text{II})} = \{\mathcal{F}(x),\mathcal{G}(y)\} \quad - & \int d^3z d^3z' \Big(-\frac{1}{m+\frac{g^2}{\alpha^2}} \frac{g}{\alpha}\delta_{ij}\{\mathcal{F}(x),\zeta_0(z)\}\{\partial_i^{(z')}\zeta_j(z'),\mathcal{G}(y)\} \\
+ & \frac{1}{m+\frac{g^2}{\alpha^2}}\{\mathcal{F}(x),\zeta_0(z)\}\{\Xi_1(z'),\mathcal{G}(y)\} \\
+ & \frac{1}{m+\frac{g^2}{\alpha^2}}\frac{g}{\alpha}\delta_{ij}\{\mathcal{F}(x),\partial_j^{(z)}\zeta_i(z)\}\{\zeta_0(z'),\mathcal{G}(y)\} \\
+ & \delta_{ij}\{\mathcal{F}(x),\partial_j^{(z)}\zeta_i(z)\}\{\Delta_0(z'),\mathcal{G}(y)\} \\
- & \delta_{ij}\{\mathcal{F}(x),\zeta_i(z)\}\{\Xi_{3_j}(z'),\mathcal{G}(y)\} \\
- & \delta_{ij}\{\mathcal{F}(x),\Delta_0(z)\}\{\partial_i^{(z')}\zeta_j(z'),\mathcal{G}(y)\} \\
+ & \delta_{ij}\{\mathcal{F}(x),\Delta_0(z)\}\{\partial_i^{(z')}\Delta_j(z'),\mathcal{G}(y)\} \\
+ & \{\mathcal{F}(x),\Delta_0(z)\}\{\Lambda^0(z'),\mathcal{G}(y)\} \\
- & \delta_{ij}\{\mathcal{F}(x),\partial_j^{(z)}\Delta_i(z)\}\{\Delta_0(z'),\mathcal{G}(y)\} \\
- & \delta_{ij}\{\mathcal{F}(x),\Delta_i(z)\}\{\Xi_{2_j}(z'),\mathcal{G}(y)\} \\
+ & \delta_{ij}\{\mathcal{F}(x),\Delta_i(z)\}\{\Xi_{3_j}(z'),\mathcal{G}(y)\} \\
- & \{\mathcal{F}(x),\Lambda^0(z)\}\{\Delta_0(z'),\mathcal{G}(y)\} \\
- & \frac{1}{m+\frac{g^2}{\alpha^2}}\{\mathcal{F}(x),\Xi_1(z)\}\{\zeta_0(z'),\mathcal{G}(y)\} \\
+ & \delta_{ij}\{\mathcal{F}(x),\Xi_{2_i}(z)\}\{\Delta_j(z'),\mathcal{G}(y)\} \\
+ & \delta_{ij}\{\mathcal{F}(x),\Xi_{3_i}(z)\}\{\zeta_j(z'),\mathcal{G}(y)\} \\
- & \delta_{ij}\{\mathcal{F}(x),\Xi_{3_i}(z)\}\{\Delta_j(z'),\mathcal{G}(y)\}\Big)\delta^{(3)}(\vec{z}'-\vec{z})
\end{aligned} \tag{132}
$$

An alternative form of these brackets, already developed in terms of the canonical fields (after using the fundamental Poisson brackets (101)), is presented in the Appendix (6.1). Those Dirac brackets, are in fact ready to use for the computations below.

4.3.4 New set of canonical fields in the Dirac brackets

Now that the Dirac brackets (132), (189) have been explicitly computed for the theory (123), it is possible to compute the sets of new canonical variables $\{\omega_g\}_{g=1,\ldots,10}$, $\{\Omega_f\}_{f=1,\ldots,16}$ that were described above. Nevertheless, as it is pointed out by [45], the problem is non-trivial, because there is no standard approach to find such sets of canonical fields. To deal with this problem, a very simple yet useful approach is taken: we construct the equivalent to the "*square matrix of Poisson brackets*" by the new Dirac brackets (189). Then, we extend this matrix, adding the rows and columns composed of the Dirac brackets of the canonical fields, with the derivatives of the fields that appear on the constraint equations (123). The reason for the latter is that we expect a set of constraints $\{\Omega_f\}_f$, $\Omega_f \equiv 0\ \forall f$, such that it is equivalent to $\{\Pi\}$ (See equation (135)) and in principle, one expects a linear combination of $\{\Pi\}$ to form $\{\Omega_f\}_f$, because it is the simpler way in which (124) is satisfied. Then, with the described extended matrix of Dirac brackets between canonical fields and pertinent derivatives, the problem reduces to find fields whose matrix of Dirac brackets (not the extended one) takes the usual form (J) in the symplectic notation, *i.e.*:

$$\dot{\boldsymbol{\Upsilon}} = J\frac{\partial\tilde{\mathcal{H}}}{\partial\boldsymbol{\Upsilon}} \tag{133}$$

$$J = \begin{pmatrix} 0_n & \tilde{I} \\ -\tilde{I} & 0_n \end{pmatrix} \tag{134}$$

where $\boldsymbol{\Upsilon}$ is a vector of all the fields in the sets $\{\omega_g\}_{g=1,\ldots,10}$, $\{\Omega_f\}_{f=1,\ldots,16}$, 0_n is a 13×13 zero matrix and $\tilde{I}$ is a diagonal 13×13 matrix, with entries 0 or 1 in the diagonal. The reason for such a form of the $\tilde{I}$ matrix is that in the new variables, the set $\{\Omega_f\}$ is demanded to be identically 0. Thus, only the equations corresponding to dynamical degrees of freedom ($\{\omega_g\}$) have a "1" entry in the diagonal of $\tilde{I}$. The latter can be better understood by noting that equation (133) is in fact the system of Hamilton equations for the new fields ω, Ω.

It is important to note that this is not a canonical transformation between (Φ) and (ω, Ω), because by searching these sets, such that (124) is satisfied, we are making the following change:

$$\begin{aligned}
\{\cdot,\cdot\}^{\Phi}_{PoissonB(\Pi)} \quad &\rightarrow \quad \{\cdot,\cdot\}^{\omega,\Omega}_{DiracB(\Omega)} = \{\cdot,\cdot\}^{\omega}_{PoissonB} && (135)\\
\Phi \quad &\in \quad \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\} \text{ is the initial set of canonical fields.}\\
\Pi \quad &\in \quad \{\zeta_0, \zeta_i, \Delta_\mu, \Lambda^0, \Xi_1, \Xi_{2_i}, \Xi_{3_i}\} \text{ are the constraints for the set of fields } \{\Phi\}\\
\text{and} \quad &\quad \{\omega_g\}_{g=1,\ldots,10}\ ,\ \{\Omega_f\}_{f=1,\ldots,16} \text{ are the new set of canonical fields.}\\
&\quad \Omega_f \equiv 0\ \forall f \text{ are the new equivalent constraints in the new set.}
\end{aligned}$$

Based on the previous discussion, the following equal-time Dirac brackets between fields $\{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$, are computed using equation (189). The lenghty results are only an introduction to a very lenghty computation. They are presented in the appendix (6.1). These results are very important, because allow to form the extended matrix of Dirac brackets between the initial set of fields, and this is the first step in the search for the new set of canonical variables.

With the Dirac brackets between fields, given in equation (190) the matrix of these brackets, extended with those with the pertinent[33] derivatives of the fields, is written below. In particular, the entries of the matrix of Dirac brackets, where $\Phi_a \in \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$, with the index a running through all of them, are:

$$J_{ab} = \{\Phi_a, \Phi_b\}_{D(\Pi)} \tag{136}$$

[33] As was described above, with "pertinent" we refer to those derivatives of fields explicitly appearing on the constraints (123), because in principle, only these will appear in the new sets of canonical fields satisfying (124). For a more extensive discussion, see the beginning of this subsection.

And we define the extended matrix of Dirac brackets by adding $\{\partial_i P_{\psi_0}, \partial_i P_{A_i}, \partial_i \varphi\}$ to the set, above. This is, for $\tilde{\Phi} \in \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi, \partial_i P_{\psi_0}, \partial_i P_{A_i}, \partial_i \varphi\}$. The entries of the extended matrix of Dirac brackets are:

$$\tilde{J}_{ab} = \{\tilde{\Phi}_a, \tilde{\Phi}_b\}_{D(\Pi)} \tag{137}$$

To make this extension clearer, let us recall the secondary constraints of the theory:

$$\begin{aligned}
\Xi_1(x) &:= mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0 \\
\Xi_{2_i}(x) &:= -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \approx 0 \\
\Xi_{3_i}(x) &:= -\psi^i(x) - \partial_i \varphi(x) \approx 0
\end{aligned}$$

These are the only constraints involving derivatives of the fields, and are precisely these derivatives that we are considering for the extended matrix of Dirac brackets.

Based on (190), in appendix 6.1, it is evident that for the matrix of Dirac brackets, it is important to write separately the 0 and j indices. Hence, for the vector $\tilde{\Phi}$ of fields,

$$\tilde{\Phi}^T =: \left(\psi^0, \psi^i, \varphi, A^0, A^j, \lambda^0, \lambda^j, P_{\psi_0}, P_{\psi_i}, P_\varphi, P_{A_0}, P_{A_i}, P_{\lambda_0}, P_{\lambda_i}, \partial_i P_{\psi_0}, \partial_i P_{A_j}, \partial\varphi\right) \tag{138}$$

the extended matrix of Dirac brackets (33×33) is schematically presented in the appendix (6.1), and is built using the results (189). Now, since there is no standard way[34] to find the desired new set of canonical fields described above, we propose to use the previous matrix in the following way:

First, keeping in mind that this is a matrix of bilinear expressions but also that it is antisymmetric, it is easy to see that through linear combinations between columns,

$$\{\cdot, \tilde{\Phi}_{a_k}\}_{D(\Pi)} + \{\cdot, \tilde{\Phi}_{a_l}\}_{D(\Pi)} = \{\cdot, \tilde{\Phi}_{a_k} + \tilde{\Phi}_{a_l}\}_{D(\Pi)} \tag{139}$$

we are only making use of the distributive property of Dirac brackets. By the fact that the entries are bilinear expressions and by the antisymmetry of the matrix $\tilde{J}$, we are forced to do exactly the same combination with the rows if we are to recover at the end, some matrix susceptible to be identified as a "matrix of Dirac brackets", between some field combinations.

$$\{\tilde{\Phi}_{a_k}, \cdot\}_{D(\Pi)} + \{\tilde{\Phi}_{a_l}, \cdot\}_{D(\Pi)} = \{\tilde{\Phi}_{a_k} + \tilde{\Phi}_{a_l}, \cdot\}_{D(\Pi)} \tag{140}$$

Furthermore, since the diagonal entries of $\tilde{J}$ are always 0, the process just described is not problematic.

However, it is very important to clearly define what we are doing: This is not a formal operation in which one obtains different $\tilde{J}$ matrices in between, but only a useful mnemotechnic way to keep track of the values of the Dirac brackets, between some combination of the initial canonical fields.

[34] See [45] chapter 2, for the related discussion.

Put in other words, one could consider this procedure as a guide to operate with the fields, then redefine them accordingly, and then compute again a matrix of Dirac brackets between the redefined field combinations. Use the newly computed matrix to find again the most suitable combination of fields, in order to reach a set which satisfies the conditions one desires. Finally iterate this construction process[35].

Specifically, as has been extensively described, we are interested in some variables that are canonical in the Dirac brackets $\{\cdot,\cdot\}_{D(\Omega_f)}$, and the variables $\{\Omega_f\}_{f=1,\dots,16}$, vanish identically $\Omega_f \equiv 0\ \forall f$. In other words, we expect them to be such that the final J has the form described for (134).

Following the mentioned procedure with the extended matrix (192), a $\tilde{J}$ matrix, whose entries are the Dirac brackets between the fields $\{\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i, \partial_i P_{\psi_0}, \partial_i P_{A_i}, \partial_i\varphi\}$ is finally found to have the form (134) that was described above, implying that the conditions are fulfilled. This will be verified, but first, let us state the change of fields:

$$\Phi =: \begin{pmatrix} \psi 0 \\ \psi^i \\ \varphi \\ A^0 \\ A^j \\ \lambda^0 \\ \lambda^j \\ P_{\psi_0} \\ P_{\psi_i} \\ P_\varphi \\ P_{A_0} \\ P_{A_i} \\ P_{\lambda_0} \\ P_{\lambda_i} \end{pmatrix} \rightarrow \begin{pmatrix} mA_0 + \frac{g}{\alpha}(P_\varphi - \psi_0) - \partial_i P_{A_i} \\ -\psi^i - \partial_i\varphi \\ \varphi - P_{\psi_0} \\ A^0 - \frac{1}{m + \frac{g^2}{\alpha^2}}\partial_i P_{A_i} \\ A^i \\ P_\varphi + \lambda^0 \\ -\psi^i - \lambda^i - \partial_i P_{\psi_0} \\ P_{A_0} - \frac{g}{\alpha}P_{\psi_0} \\ P_{\psi_i} \\ P_\varphi \\ \frac{\alpha}{g}\left(m + \frac{g^2}{\alpha^2}\right)P_{\psi_0} \\ P_{A_i} \\ P_{\lambda_0} \\ P_{\lambda_i} \end{pmatrix} =: \begin{pmatrix} \Xi_1 \\ \Xi_{3_i} \\ \eta \\ \gamma \\ \Theta^i \\ \Lambda^0 \\ \Xi_{2_i} \\ \zeta_0 \\ \zeta_i \\ P_\eta \\ P_\gamma \\ P_{\Theta_i} \\ \Delta_0 \\ \Delta_i \end{pmatrix} := \Upsilon \tag{141}$$

where by the previous expression, the vector $\Upsilon^T = (\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i)$ has been defined to the left. The explicit dependence on the space-time point is not written to avoid very cumbersome expressions, but it must be noted that the field combinations like $\varphi - P_{\psi_0}$ are in fact defined at certain (x), *i.e.* $\varphi(x) - P_{\psi_0}(x)$.

The new fields $\{\Upsilon\}$ are defined to be canonical in the new Dirac brackets $\{\cdot,\cdot\}_{D(\Omega)}$. The latter was suggested by the notation of the new fields in (141), and in fact is verified by computing their Dirac brackets in the new constraints $\{\Omega\} = \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}$ which are in fact the same constraints initially found, for the previous set of fields. Put in other words, in accordance with the previous discussions, the set of new fields contains a subset $\{\Omega\}$ which vanish identically and defines the new constraints. These are linear combinations of the initial constraints. In this case,

[35] For the outcome of the last procedure, is important to be aware that different outcomes related by canonical transformation, can be reached. This assertion comes from the fact that the separation of canonical variables, described in the cited theorem (3.2), is not unique, but all of the possible sets are related by canonical transformation. The theorem related to the last assertion is proved in chapter 2 of [45].

they were chosen to be the same, because the model is simple enough to do so. However, the set $\{\omega\} = \{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\}$ is not trivial at all, and are to be considered as the physical dynamical variables on the constraint hypersurface defined by $\{\Omega\}$, that evolve by some Hamiltonian $\tilde{\mathcal{H}}$ that we now set to build.

Under a explicit computation, the following are the Dirac brackets on the constraints $\{\Omega\}$, for the new fields $\Upsilon_a \in \{\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i\}$:

$$\begin{aligned}
\{\eta(x), P_\eta(y)\}_{D(\Omega)} &= \delta^{(3)}(\vec{x}-\vec{y}) \\
\{\gamma(x), P_\gamma(y)\}_{D(\Omega)} &= \delta^{(3)}(\vec{x}-\vec{y}) \\
\{\Theta^i(x), P_{\Theta_j}(y)\}_{D(\Omega)} &= \delta^i_j\delta^{(3)}(\vec{x}-\vec{y})
\end{aligned}$$

And,

$$\{\Upsilon_a(x), \Upsilon_b(y)\}_{D(\Omega)} = 0 \tag{142}$$

for all the other Dirac brackets.

With (142) it is verified that the Dirac brackets in the canonical variables $\{\Omega\}$, which vanish identically and are also the new constraints, can be considered as Poisson brackets only in the fields $\{\omega\} = \{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\}$. This is, the equation (124) $\{\cdot,\cdot\}^{\omega,\Omega}_{DiracB(\Omega)} = \{\cdot,\cdot\}^{\omega}_{PoissonB}$ has been satisfied.

Hamiltonian in the new canonical fields

Now that the canonical fields have been found, what remains is to write the Hamiltonian (123) in terms of these new fields. Let us recall the Hamiltonian (123) and express it in a convenient way for the following discussion:

$$\begin{aligned}
\mathcal{H}(\Phi) &= \mathcal{H}_0 + V^0_A(\Phi)\zeta_0 + V_\varphi(\Phi)\Lambda^0 + V^0_\lambda(\Phi)\Delta_0 + V^i_\lambda(\Phi)\Delta_i + V^i_\psi(\Phi)\zeta_i + \chi_1(\Phi)\Xi_1 + \chi^i_2(\Phi)\Xi_{2_i} \\
&+ \chi^i_3(\Phi)\Xi_{3_i} \\
\mathcal{H}_0(\Phi) &= \frac{1}{2}P_{A_i}P_{A_i} - \frac{1}{2\alpha}P^2_{\psi_0} + \frac{1}{4}F_{ij}F^{ij} - \frac{m}{2}A_\mu A^\mu - \frac{1}{2}\psi_\mu\psi^\mu + \frac{1}{2}w\varphi^2 + P_\varphi\psi_0 - P_{A_i}\partial_i A_0 \\
&- P_{\psi_0}(\partial_i\psi^i + \frac{g}{\alpha}\partial_i A^i) - \lambda^i(\psi_i - \partial_i\varphi)
\end{aligned}$$

$$\begin{aligned}
\mathcal{H}(\Phi) &= \mathcal{H}_0 + \sum_a V_{\Phi_a}\Pi_a(\Phi) \\
H(t) &= \int d^3x\mathcal{H}(\Phi(x)) \\
\dot{\Phi}(x) &= \{\Phi(x), H(t)\}
\end{aligned} \tag{143}$$

Before going on with the computation, there is an important fact that greatly simplifies this task. First, let us recall that the Lagrange multipliers, which we identified as primarly unexpressible velocities, were fixed by imposing the conservation of constraints in time. However, some of these, specifically those found in the last stage, were not explicitly written. The reason for that is supported by two arguments: firstable, we are sure that they can be found and fixed because the Dirac matrix of Poisson brackets between constraints, is invertible. This is a simple result that was stated in section 3.4, and reads [46, 45]:

$$V_{\Pi_a}(\Phi) \quad = \quad -\int d^3z d^3z' \{\Pi, \Pi\}^{-1}_{a,a'} \{\Pi_{a'}, H_0(t)\} + \mathcal{O}(\Pi) \tag{144}$$

where sum convention in a has been used, $\{\Pi, \Pi\}^{-1}_{a,a'}$ is the inverse of the Dirac matrix, and $\mathcal{O}(\Pi)$ is some combination of constraints. Therefore, all the primarily unexpressible velocities are fixed, because the constraint content of the theory has been defined and there are no first-class constraints.

Second, there is also a general result [45], that was presented in 3.4, where it is stated that only $\mathcal{H}_0$ need to be transformed to the new fields. For the latter, the general argument may be difficult to follow. However, for the present case it is much simpler: The set of variables $\{\Omega\}$ that vanish identically, can in fact be written as the same combination of initial fields $\{\Phi\}$, that were considered as constraints $\{\Pi\}$, *i.e.* the variables $\{\Omega\}$ coincide with the constraints for the initial set of canonical fields $\Phi \in \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$. Therefore, as was presented in 3.4, and also mentioned in the discussion above, by the following property of Dirac brackets,

$$\{\cdot, \{\Omega\}\}_{D(\Omega)} = \mathcal{O}(\Omega)$$

follows:

$$\begin{aligned}
\{\cdot, \mathcal{H}(\Phi)\}_{D(\Omega)} = \{\cdot, \mathcal{H}_0(\Phi) + \sum_a V_{\Pi_a}\Pi_a(\Phi)\}_{D(\Omega)} \quad &= \quad \{\cdot, \mathcal{H}_0(\Phi)\} + \{\cdot, \sum_a V_{\Phi_a}\Pi_a(\Phi)\}_{D(\Omega)} \\
&= \quad \{\cdot, \mathcal{H}_0(\Phi)\} + \{\cdot, \sum_a V_{\Phi_a}\Omega_a(\Phi)\}_{D(\Omega)} \\
&= \quad \{\cdot, \mathcal{H}_0(\Phi)\} + \{\cdot, \{\Omega\}\}_{D(\Omega)} \\
&= \quad \{\cdot, \mathcal{H}_0(\Phi)\} + \mathcal{O}(\Omega)
\end{aligned}$$

Which implies that whenever is possible to consider weak equalities as strong, for instance, if the field configurations are demanded to follow the equations of motion,

$$\{\cdot, \mathcal{H}(\Phi)\}_{D(\Omega)} \to \{\cdot, \mathcal{H}_0(\Phi)\}_{D(\Omega)} \tag{145}$$

Then, the set of dynamical fields $\{\omega\}$ evolves with the Hamiltonian that is transformed from $\mathcal{H}_0(\Phi)$, *i.e.* we must only care about rewriting the latter in terms of the new canonical fields:

$$\mathcal{H}_0(\Phi) \to \tilde{\mathcal{H}}(\omega, \Omega) \tag{146}$$

The latter is just a particular analysis of the general result found in [45] (Chapter 2).

With the condition (106), $g^2 \stackrel{!}{=} \alpha\beta$, the Hamiltonian in the new set of fields $\{\Upsilon\} = \{\omega, \Omega\}$ is:

$$\begin{aligned}
\tilde{\mathcal{H}}(\omega,\Omega) &= -\frac{\beta}{2}\frac{1}{(m\alpha+\beta)^2}\left(P_\gamma^2+\alpha\nabla P_\gamma\cdot\nabla P_\gamma\right)+\frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_\gamma+\eta\right)^2+\frac{1}{4}(F^{ij})^2 \\
&- \frac{m}{2\beta}(m\alpha+\beta)\gamma^2+\frac{1}{2}\left(P_{\Theta_j}^2+\frac{\alpha}{m\alpha+\beta}(\nabla P_{\Theta_j}\cdot\nabla P_{\Theta_j}\right)+\frac{m}{2}(\Theta^i)^2+\frac{P_\eta^2}{2}+\frac{1}{2}\nabla\eta\cdot\nabla\eta \\
&+ \mathcal{O}(\Omega^2) \\
F_{ij} &=: \partial_i\Theta_j-\partial_j\Theta_i \qquad (147)
\end{aligned}$$

4.3.5 Physical Hamiltonian

The final form of the Hamiltonian density in terms of the new canonical fields is (147), but we already now that the dynamical variables $\{\omega\}$ evolve on the hypersurface defined by the identically vanishing fields Ω, which implies the already mentioned definition (See 3.4 and previous subsection on Dirac brackets for a broader discussion):

$$\begin{aligned}
\mathcal{H}_{phys} &= \tilde{\mathcal{H}}(\omega,\Omega)|_{\Omega\equiv 0} \\
&\downarrow \\
\mathcal{H}_{phys}(\omega) &= -\frac{\beta}{2}\frac{1}{(m\alpha+\beta)^2}\left(P_\gamma^2+\alpha\nabla P_\gamma\cdot\nabla P_\gamma\right)+\frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_\gamma+\eta\right)^2+\frac{1}{4}(F^{ij})^2 \\
&- \frac{m}{2\beta}(m\alpha+\beta)\gamma^2+\frac{1}{2}\left(P_{\Theta_j}^2+\frac{\alpha}{m\alpha+\beta}\nabla P_{\Theta_j}\cdot\nabla P_{\Theta_j}\right)+\frac{m}{2}(\Theta^i)^2+\frac{P_\eta^2}{2}+\frac{1}{2}\nabla\eta\cdot\nabla\eta \\
F_{ij} &=: \partial_i\Theta_j-\partial_j\Theta_i
\end{aligned}$$

And the equations of motion for the fields ω, are given by:

$$\begin{aligned}
\dot{\omega}_a(x) &= \int d^3y\{\omega_a(x),\mathcal{H}_{phys}(y)\} \\
\text{With} \quad & \omega_a\in\{\eta,P_\eta,\gamma,P_\gamma,\Theta^i,P_{\Theta_i}\} \qquad (148)
\end{aligned}$$

Where terms like $\nabla P_{\Theta_j}\cdot\nabla P_{\Theta_j}$ can be also written as $\partial_i P_{\Theta_j}\partial_i P_{\Theta_j} = (\partial_i P_{\Theta_j})^2$, or $(\Theta^i)^2 = \Theta^i\Theta^i$.

Even though the new canonical fields were defined by (141), it is worth recalling the definition of the dynamical fields $\{\omega\}$, using the fact that $g^2 \stackrel{!}{=} \alpha\beta$, in terms of the previous fields $\{\Phi\}$.

$$\begin{aligned}
\eta(x) &=: \varphi(x) - P_{\psi_0}(x) \\
P_\eta(x) &=: P_\varphi(x) \\
\gamma(x) &=: A^0(x) - \frac{\alpha}{m\alpha+\beta}\partial_i P_{A_i}(x) \\
P_\gamma(x) &=: \frac{m\alpha+\beta}{g} P_{\psi_0} \\
\Theta^i(x) &=: A^i(x) \\
P_{\Theta_i}(x) &=: P_{A_i}(x)
\end{aligned} \tag{149}$$

Also, it is important to emphasize that $\{\Omega\} = \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}$ are no longer constraints between canonical fields $\{\Phi\}$, but now, they are to be considered **canonical variables on their own**, that however, vanish identically when the equations of motion are satisfied, and therefore are not dynamical[36]. On the other hand, we have reached a definition of the non-dynamical fields $\{\Omega\}$, which confusingly enough takes the same form as the constraints before, but one must be careful to recognize, that the fields $\{\Phi\}$, in terms of which $\{\Omega\}$ are now expressed, are **not** canonical in the Dirac brackets. Let us recall the definitions:

$$\begin{aligned}
\zeta_0(x) &=: P_{A_0}(x) - \frac{g}{\alpha} P_{\psi_0}(x) \\
\zeta_i(x) &=: P_{\psi_i}(x) \\
\Delta_\mu(x) &=: P_{\lambda_\mu}(x) \\
\Lambda^0(x) &=: P_\varphi(x) + \lambda^0(x) \\
\Xi_1(x) &=: mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \\
\Xi_{2_i}(x) &=: -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \\
\Xi_{3_i}(x) &=: -\psi^i(x) - \partial_i \varphi(x)
\end{aligned} \tag{150}$$

Finally, it is worth recalling that these new canonical fields (149), (150), are canonical in the new structure defined by the Dirac brackets (142).

In a strict sense, we could carry on with the analysis of the Ostrogradsky's instability with the classical form of the theory. However, the motivation to eliminate the instability can be better understood when writing the corresponding quantum theory. Therefore, let us continue with this analysis in the following section.

4.3.6 Canonical quantization of the stable field theory

The motivation up until now, has been to put the theory $\mathcal{L}$ in a form susceptible to be quantized. Since the theory with higher derivatives must contain constraints to be stable, the procedure was

[36] This can be seen by the previous discussion, when taking $\Omega \equiv 0$ in the physical Hamiltonian, which explains why they no longer appear in the physical Hamiltonian.

not straightforward and the Dirac's programme had been followed. Now, having defined a set of canonical fields $\{w\}$ that freely evolve with $\mathcal{H}_{phys}$, by means of the Dirac brackets $\{\cdot,\cdot\}_{D(\Omega)}$, we can almost soundly use a quantization scheme, as a first approach to a quantum field theory involving higher derivatives.

A functional approach may be better justified for certain purposes, like computations of scattering processes, but for the present analysis we are interested in the form of the Hamiltonian field operator and in the case with only one degree of freedom, also the spectrum. The reason for the latter is that the difficulties with these theories have been historically analysed in the Hamiltonian formalism and the most straightforward quantization with the Hamiltonian, is canonical.

Because the Ostrogradsky's instability can be identified with the unboundedness from below of the spectrum and the non-positive definiteness of the Hamiltonian, we are now in a position to analyse the sector in the space of parameters (α, β, m, w) in order to avoid at least the last problem, in order to finally find the physical theory that do not include the Ostrgradsky's instability and can indeed, be quantized.

What we mean by the latter, is that an unbounded from below energy operator does not lead to a so called "stable vacuum", and in some approaches to QFT, it is highly important to demand that the domain of the operators in the Hilbert space, be such that an unique vacuum, or lowest energy state, is defined. In other words, the vacuum should be stable or at least, should not decay with the "infinite time scale"[6] that it takes in the Ostrogradsky's instability. One example of such formalism is the Wightman axiomatic quantum field theory, which is a very promising approximation to give a sound mathematical foundation to QFT. To be more precise, in this formalism, it is explicitly stated that the domain of the Hermitian valued operators, in the space of Schwartz-class functions, must contain an unique vacuum state.

Based on the previous discussion, let us find the parameters for which the classical field theory $\mathcal{L}$, with Hamiltonian (148), could, in principle, be quantized. Keeping in mind that $g^2 \stackrel{!}{=} \alpha\beta$, it may be the possible that $g \in \mathbb{C}$. In fact, in order to avoid the Ostrogradsky's instability for this particular field theory, this must be the case, and taking $\mathbb{R} \ni \alpha \stackrel{!}{>} 0$, $\mathbb{R} \ni \beta \stackrel{!}{<} 0$, and $|\beta|$ as the absolute value of β, the coupling constant g, takes the form:

$$g \stackrel{!}{=} \mp i\sqrt{\alpha|\beta|} \tag{151}$$

This allows the physical Hamiltonian (148) to have the desired properties. However, this implies the appearence of an imaginary coupling constant in the Lagrangian. As will be shown now, this may indeed make sense. Let us recall the Lagrangian in the initial fields:

$$\mathcal{L} = -\tfrac{1}{2}\alpha(\partial_\mu\partial^\mu\phi)(\partial_\nu\partial^\nu\phi) + \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}\omega\phi^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}\beta\partial_\mu A^\mu\partial_\nu A^\nu + \frac{m}{2}A^2 - g\partial_\mu\partial^\mu\phi\partial_\nu A^\nu$$

It is important to note that the term with the coupling g,

$$-g\partial_\mu\partial^\mu\phi\partial_\nu A^\nu \tag{152}$$

includes three derivatives. In a very rough approach, terms with an odd number of derivatives could lead to a non-hermitian term. This is readily seen at the end form of the Hamiltonian (148).

With $\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $g \overset{!}{=} \pm i\sqrt{\alpha|\beta|}$, the physical Hamiltonian takes the following form:

$$\begin{aligned} \mathcal{H}_{phys} \quad (\omega) \quad &= \frac{|\beta|}{2}\frac{1}{(m\alpha-|\beta|)^2}\left(P_\gamma^2+\alpha\nabla P_\gamma\cdot\nabla P_\gamma\right)+\frac{w}{2}\left(\frac{\mp i\sqrt{\alpha|\beta|}}{m\alpha-|\beta|}P_\gamma+\eta\right)^2+\frac{1}{4}(F^{ij})^2 \\ &+ \quad \frac{m}{2|\beta|}(m\alpha-|\beta|)\gamma^2+\frac{1}{2}\left(P_{\Theta_j}^2+\frac{\alpha}{m\alpha-|\beta|}\nabla P_{\Theta_j}\cdot\nabla P_{\Theta_j}\right)+\frac{m}{2}(\Theta^i)^2+\frac{P_\eta^2}{2}+\frac{1}{2}\nabla\eta\cdot\nabla\eta \end{aligned} \tag{153}$$

The question regarding if this Hamiltonian would lead to an Hermitian operator or not, is in general difficult to address, but the following discussion may be a first approach from a much simpler and practical perspective: In (153), the term $+\frac{w}{2}\left(\frac{\mp i\sqrt{\alpha|\beta|}}{m\alpha-|\beta|}P_\gamma+\eta\right)^2$ appears to be complex, but it must be considered that i is multiplying P_γ and therefore, if it is complex or not, could be seen explicitly after taking a mode expansion[37] of the initial fields $\{\Phi\} = \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$. The latter define the fields $\{\Upsilon\} = \{\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i},\ \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i\}$ in terms of modes, by means of their definition (149), (150) and only then, the term $iP_\gamma(\hat{a}_i, \hat{a}_i^*)$ could be seen to be real or complex. However, this would be a lenghty computation for a simple toy model, where the only intention has been to see how the classical conditions on theories with only one degree of freedom, could be extended to covariant field theories. Therefore, some simpler approaches could be taken: the first one and more related to the present discussion, is simply to see that no difficulty arises when considering field toy models contained in the current set, by taking the special case of massless higher derivative scalar field $\varphi(x)$. This is, by demanding that $w \overset{!}{=} 0$ in the Lagrangian, we get the final physical Hamiltonian:

$$\begin{aligned} \mathcal{H}_{phys} \quad (\omega) \quad &= \frac{|\beta|}{2}\frac{1}{(m\alpha-|\beta|)^2}\left(P_\gamma^2+\alpha\nabla P_\gamma\cdot\nabla P_\gamma\right)+\frac{1}{4}(F^{ij})^2+\frac{m}{2|\beta|}(m\alpha-|\beta|)\gamma^2 \\ &+ \quad \frac{1}{2}\left(P_{\Theta_j}^2+\frac{\alpha}{m\alpha-|\beta|}\nabla P_{\Theta_j}\cdot\nabla P_{\Theta_j}\right)+\frac{m}{2}(\Theta^i)^2+\frac{P_\eta^2}{2}+\frac{1}{2}\nabla\eta\cdot\nabla\eta \end{aligned} \tag{154}$$

From (154), it is easy to see that a condition for positive definiteness of the Hamiltonian, after demanding $w \overset{!}{=} 0$, is that

$$\alpha > \frac{|\beta|}{m} \tag{155}$$

Now, finally, after starting with a Lagrangian (95) that could or could not lead to a correct quantum theory, and after imposing restrictions on the parameters ($\alpha, m \in \mathbb{R}^+$, $\beta \in \mathbb{R}^-$, $g^2 \overset{!}{=} \alpha\beta$,

[37] This would be an expansion with ladder operators in a quantum theory.

$w \stackrel{!}{=} 0$, and more interesting $\alpha > \frac{|\beta|}{m}$), we can, on a sound basis, proceed with the canonical quantization of the corresponding physical Hamiltonian (154). By promoting the continuum fields $\{\Upsilon\} = \{\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i\}$ to operators, and consequently also promoting the physical Hamiltonian to a field operator, and the **Dirac brackets** to commutators in the following prescription,

$$\{\cdot, \cdot\}_{D(\Omega)} \rightarrow \frac{1}{i}[\cdot, \cdot] \tag{156}$$

we obtain a quantum field theory without the Ostrogradsky's instability.

Specifically, the commutators are:

$$\begin{aligned}
[\eta(x), P_\eta(y)] &= i\delta^{(3)}(\vec{x} - \vec{y}) \\
[\gamma(x), P_\gamma(y)] &= i\delta^{(3)}(\vec{x} - \vec{y}) \\
\left[\Theta^i(x), P_{\Theta_j}(y)\right] &= i\delta^i_j\delta^{(3)}(\vec{x} - \vec{y})
\end{aligned}$$

And,

$$[\Upsilon_a(x), \Upsilon_b(y)] = 0 \tag{157}$$

for all the other commutators promoted from the Dirac brackets.

This result seems quite feeble after so much work, but the importance relies on the fact that we have proposed a covariant continuum field model, based on the results in [39] for a theory with finite degrees of freedom, and have verified that after some subtleties regarding the construction and the transformation properties of the stabilizer field, it is possible to build a classical continuum field theory that can be, in principle - *disregarding problems of other origins, like dynamical instabilities not related with that of Ostrogradsky*- be brought to the starting point of a higher derivative quantum field theory.

It must be noted that the brackets that were quantized, are not the Poisson but the Dirac brackets. This fundamental difference is what forced us to carry on with such cumbersome procedure of finding the constraint structure in the theory, then the Dirac brackets and finally the new canonical fields $\{\Upsilon\}$ (141) evolving freely, without apparent constraint by the physical Hamiltonian (154). When we say "freely", we refer to the fact that the set $\{\eta, \gamma, \Theta^i, P_\eta, P_\gamma, P_{\Theta_i}\}$ evolve in a "trivial" hypersurface in the phase space of field configurations, defined by the "planes" $\Omega_f \equiv 0$ for $\Omega_f \in \{\Xi_1, \Xi_{3_i}, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, \Delta_0, \Delta_i\}$ $\forall f$.

Further comments on this result, as well as additional approaches to understand its meaning are given in section 4.1. The ideas written there, are complementary and offer other perspectives to the more extensive analyses shown above.

Other approaches: Introduction to the next sections.

Now that the approach by simplifying (153) through $w \stackrel{!}{=} 0$ has been taken, we motivate the following section. Let us recall that (153) has the possible problem of a term that could or could not be complex, and even though this is not conclusive about the hermiticiy of a corresponding Hamiltonian operator in a quantum theory, we are set with a weaker analysis regarding the possibility of a positively defined Hamiltonian function in the classical continuum theory, which is a sufficient condition to guarantee the disappearence of Ostrogradsky's instability.

As was described above, an approach different to setting $w \stackrel{!}{=} 0$ would be to expand the fields in modes (ladder operators in a quantum theory). But there is a third approach which is a simplification of the previous one, much weaker but easier for interpretation. This is, consider the same mode expansion described above, for the initial fields $\{\Phi\} = \{\psi^\mu, P_{\psi_\mu}, A^\mu, P_{A_\mu}, \lambda^\mu, P_{\lambda_\mu}, \varphi, P_\varphi\}$. The latter define the fields $\{\Upsilon\} = \{\Xi_1, \Xi_{3_i}, \eta, \gamma, \Theta^i, \Lambda^0, \Xi_{2_i}, \zeta_0, \zeta_i, P_\eta, P_\gamma, P_{\Theta_i}, \Delta_0, \Delta_i\}$ in terms of modes, by means of their definition (149), (150), but now, let us take only one mode of the expansion, in particular, the momentum $\vec{p} = \vec{0}$ mode. Then, every term in the physical Hamiltonian (148) including spatial derivatives, vanishes, because the derivatives of the exponentials put a $\vec{p}$ in front of these terms. Finally, we end up with a function, which is not the Hamiltonian $\mathcal{H}_{phys}(x)$, but only a term of the sum of its mode expansion, readily, the 0 momentum contribution to the energy function at every space-time point where $\mathcal{H}_{phys}(x) \to h_{phys_{\vec{p}=\vec{0}}}(x)$ can be evaluated. This is a very rough approach but it can be interpreted in the following familiar way:

- We get, upon re-interpretation, a similar model to that of a classical theory with finite degrees of freedom. In particular, if we demote the fields $\{\eta, \gamma, \Theta^i, P_\eta, P_\gamma, P_{\Theta_i}\}$ valued at every space-time point, to simple degrees of freedom, the "physical Hamiltonian" for one, higher-time derivative degree of freedom (playing the role of φ), coupled to another stable degree of freedom (which plays the same role as the 0-component of the vector field A^0 in field theory), would take the following form:

$$\begin{aligned} h_{phys_{\vec{p}=\vec{0}}} &\overset{\rightsquigarrow}{\approx} -\frac{\beta}{2}\frac{1}{(m\alpha+\beta)^2}P_\gamma^2 + \frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_\gamma + \eta\right)^2 - \frac{m}{2\beta}(m\alpha+\beta)\gamma^2 + \frac{1}{2}P_{\Theta_j}^2 \\ &+ \frac{m}{2}(\Theta^i)^2 + \frac{P_\eta^2}{2} \end{aligned} \tag{158}$$

this function can also be reached, upon the previously discussed re-interpretation, by taking out all the terms containing spatial derivatives, and only leaving the temporal part of the general (not yet parameter constrained) physical Hamiltonian (148). This makes sense, because we are "downgrading" from a covariant theory of continuum fields to a theory with finite degrees of freedom and therefore, only with temporal derivatives. It is important to note that this is not a formal procedure, however, this will prove to be a way to check consistency of the previous results, because we are going to reach equivalent results from a different computation in classical mechanics with finitely many degrees of freedom.

With this motivation, we will follow the same procedure as above, though a little simpler, and then quantize it in the canonical scheme. It will be found to be simple enough to easily compute the energy spectrum in certain specific cases, and this will be of help to understand the problematic term described above, for field theory, when $w \neq 0$. Furthermore, this approach is interesting on its own, because it is the first step to test the stabilization extension from classical to quantum theory.

But before going on, let us continue with some remarks regarding the previous results:

Remarks:

- It is evident from the final form of the Hamiltonian, that the equations of motion in the new canonical fields Υ, will be of first order, as is usual in the Hamiltonian formalism. However, this seems not quite right. We started with a higher derivative Lagrangian, therefore, we naturally expect higher order equations of motion. To be more precise, we expected fourth order equations of motion in the Lagrangian formalism. However, the constraint that was imposed through $g^2 \stackrel{!}{=} \alpha\beta$ eliminated one *dof* and in fact, this is a general result in [39] for models with one *dof* and up to second time derivatives[38].

  Then, with this model for infinite degrees of freedom, such result was confirmed, and at the same time, it poses a question: Why to bother with such complicated theories, including two time derivatives in the Lagrangian, if after all, the only devised method to stabilize them, eliminates the higher derivatives in the equations of motion? One possible reason may be that, it is not true that all possible equations of motion can be reached by a minimum action principle, which is the main assumption when one writes a Lagrangian. Therefore, not all equations of motion can be reached by means of explicitly covariant low-derivative lagrangians. The physical Hamiltonian density is not explicitly covariant because of the properties of the Hamiltonian formalism, but was derived without spoiling covariance, starting from a higher-derivative and explicitly covariant Lagrangian density. Hence, the equations of motion, may lead to new dynamics. However, there may be much better and well developed arguments and the question is still open.

- It is worth noting that the new dynamical variable $\gamma(x) = A^0(x) - \dfrac{\alpha}{m\alpha+\beta}\partial_i P_{A_i}(x)$, is highly related to the secondary constraint that is found for the Proca Lagrangian alone. In that case, the new set of variables defines that constraint as a non-dynamical field (See chapter 2, [45]), but here, the instability of the higher-derivative field φ, turns that canonical field into a dynamical one, which evolves with $\mathcal{H}_{phys}$.

- Even though the limiting case $\alpha \to 0$ restores the higher derivative field φ to a low derivative field, this analysis does not go through after such limit, because from the very beginning the constraint ζ_0 includes a term $\propto \dfrac{1}{\alpha}$.

[38]This was mentioned in section 3

- We ended up with 5 degrees of freedom in a certain configuration space, or equivalently, 10 canonical variables in a Hamiltonian formalism. These are $\{\eta, P_\eta, \gamma, P_\gamma, \Theta^i, P_{\Theta_i}\}$. On the other hand, we started with 4 degrees of freedom of the vector field A^μ and a scalar field with up to two derivatives in the lagrangian φ. As was shown in section (3) the unstable higher derivative degree of freedom can be written as one stable and other unstable degree of freedom. Therefore, φ gives in reality two degrees of freedom by means of its higher derivatives[39], therefore summing up 6 initial fields. The consistency check is that we demanded a new constraint between one of the DOFs propagated by the higher derivative field and A_0, in order to control the Ostrogradskian instability, thus reducing to 5 DOFs.

4.4 A stable model with higher derivatives and finite degrees of freedom

Based on the discussions of the last section, there are now two motivations to explore the properties of a much simpler model with finite degrees of freedom. Firstable, the instability of Ostrogradsky originates at the classical level, even for one degree of freedom, such as the widely discussed Pais-Uhlenbeck oscillator. Therefore, the easier the model that includes the instability, the most tractable would be the stabilization procedure. In fact, this do happen here, and at the end, we will see that in order to stabilize a Pais-Uhlenbeck oscillator, there are more ways than in the "equivalent" field theory (99). Therefore, the particular model for fields, introduces more conditions, that do not relate with the Ostrogradsky's instability.

Second, we are interested in understanding the apparently problematic term,

$$+\frac{w}{2}\left(\frac{\mp i\sqrt{\alpha|\beta|}}{m\alpha - |\beta|}P_\gamma + \eta\right)^2$$

in the general Hamiltonian (153). With this on mind, we will develop a simple model that contains exactly the same term, and upon canonical quantization, we will find the energy spectrum, turning out to be positive and bounded from below. Altough this result is strictly pertinent to finite degrees of freedom, at the end of the last section, we gave a brief discussion of how we can reach a comparable model on fields, by taking only the $\vec{p} = \vec{0}$ mode of the expansion in momenta of the physical Hamiltonian, and finally obtaining the corresponding function $h_{phys_{\vec{p}=\vec{0}}}(x)$ (158), which upon-reinterpretation, demoting fields to simple degrees of freedom, will be exactly the model in the present discussion. Therefore, even though we do not expect to give a definite answer to this problem in fields, we do expect to gain a little more insight into the properties of the field theory including this term. Nevertheless, it is worth recalling that a "problem free" higher-derivative field theory that could be quantized, was reached by setting the higher derivative continuum field to be massless, *i.e.* $w \stackrel{!}{=} 0$ in (159).

[39] This can also be seen from the equivalent Lagrangian, where $\psi_\mu =: \partial_\mu\varphi$. From this relation two canonical momenta were derived P_φ and P_{ψ_0} (The other P_{ψ_i} are non-dynamical, in fact, constraints). Furthermore, if one computes the propagator of the higher derivative field, one finds that it can be decomposed into two propagators, as for two scalars degrees of freedom.

Let us now begin with the model with finite degrees of freedom by re-writting the original Lagrangian density for field theory:

$$\mathcal{L}(\varphi(x), A^\mu(x)) = -\tfrac{1}{2}\alpha(\partial_\mu\partial^\mu\phi)(\partial_\nu\partial^\nu\phi) + \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}\omega\phi^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}\beta\partial_\mu A^\mu\partial^\mu A_\mu + \frac{m}{2}A^2 - g\partial_\mu\partial^\mu\phi\partial_\nu A^\nu \tag{159}$$

We are now interested in the following model with finite degrees of freedom, whose resemblance to (159) is suggestive enough, by $\varphi \leftrightarrow y$, $A^0 \leftrightarrow x$. It is worth to note that the term $-\frac{1}{4}(F^{\mu\nu})^2 = \frac{1}{2}(\partial_0 A^i + \partial_i A^0)^2 - \frac{1}{4}(F^{ij})^2$ does not contain time derivatives of A^0, making the analogy even clearer.

$$L(x,y) = -\frac{1}{2}\alpha\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 - \frac{\beta}{2}\dot{x}^2 + \frac{m}{2}x^2 - g\ddot{y}\dot{x} \tag{160}$$

Now, let us briefly sketch the procedure to follow:

The canonical conjugated momenta will be found, and demanding the stability in the sense of Ostrogradsky, we will encounter constraints between momenta. This gives rise to the same problem already encountered with fields. This is, at least in canonical quantization, the canonical variables in phase space do not evolve freely, but constrainted to a hypersurface. Therefore, the canonical quantization cannot be carried out by simply promoting degrees of freedom to operators and Poisson brackets to commutators. In fact, it is not possible at all, to build the Hamiltonian without knowing the constraint content of the theory. Therefore, we follow exactly the same procedure as before, in section (4.3), beginning with an extended Hamiltonian formalism, including primarily unexpressible velocities which cannot be inverted in terms of the momenta, but rather appear in this extended Hamiltonian function, as Lagrange multipliers to be fixed by imposing the conservation of constraints in time.

After such procedure, knowing all the second-class constraints in the theory, we will find the Dirac matrix, its inverse and the corresponding Dirac brackets which roughly define a new symplectic structure on the phase space manifold. With these new brackets, we set to find the new set of canonical variables, and finally, to express the Hamiltonian in terms of these new variables. With this final Hamiltonian function, depending on some canonical variables that evolve freely on some lower dimensional constraint surface, which is trivially defined by some non-dynamical canonical variables of the new set, we can analyse in which regions of parameter space, (α, β, w, m) and g, can the Hamiltonian be positively defined, being this a sufficient condition to eliminate the Ostrogradsky instability. Then, in those particular cases, we can promote the new canonical variables to operators and Dirac brackets to commutators. Only then, we can introduce the suitable position or momentum basis (of the new canonical variables) to project the states of Hilbert space, and find a second order Schrödinger equation, which allows us to find the energy spectrum in some particular cases.

As it is evident from the previous description, the procedure will be very similar to the one followed in section (4.3), therefore, the main objective in this section will not be to repeat the justification of the procedure, but rather to quickly find the result leading to the relevant discussion that had not been addressed in the field theory model. However, all the procedure to find the new set of variables, although with less comments, and much less broader discussion, is given in the appendix (6.2). This will be put in constant reference to the development of section (4.3). It must be noted that even though this computations are similar in form, the objects being tretated are completely different. Therefore, the computations were completely independent.

Let us define an equivalent lagrangian to (160) by introducing $Q =: \dot{y}$ and one Lagrange multiplier λ. (See (99)):

$$L(x,y,Q,\lambda)_{eq} = -\frac{1}{2}\alpha\dot{Q}^2 + \frac{1}{2}Q^2 - \frac{1}{2}wy^2 - \frac{\beta}{2}\dot{x}^2 + \frac{m}{2}x^2 - g\dot{Q}\dot{x} + \lambda(Q - \dot{y}) \tag{161}$$

After computing the momenta, as is explained in detail in the appendix (6.2), the change between the set $\{\Phi\}$ to the new set of variables $\{\Upsilon\}$ is:

$$\Phi =: \begin{pmatrix} Q \\ y \\ x \\ \lambda \\ P_Q \\ P_y \\ P_x \\ P_\lambda \end{pmatrix} \rightarrow \begin{pmatrix} (Q - P_y - \frac{\alpha}{g}mx) \\ y - P_Q \\ x \\ \lambda + P_y \\ \frac{m\alpha + \beta}{g}P_Q \\ P_y \\ P_x - \frac{g}{\alpha}P_Q \\ P_\lambda \end{pmatrix} =: \begin{pmatrix} \Xi' \\ y' \\ x' \\ \Lambda \\ P_{x'} \\ P_{y'} \\ \zeta \\ \Delta \end{pmatrix} := \Upsilon \tag{162}$$

Now, let us briefly sketch what is being done with this search for new set of canonical variables (A complete discussion is presented in (4.3)):

$$\begin{aligned} \{\cdot,\cdot\}^{\Phi}_{PoissonB(\Pi)} &\rightarrow \{\cdot,\cdot\}^{\omega,\Omega}_{DiracB(\Omega)} = \{\cdot,\cdot\}^{\omega}_{PoissonB} \qquad (163)\\ \Phi &\in \{x, y, Q, \lambda, P_x, P_y, P_Q, P_\lambda\} \text{ is the initial set of canonical variables.}\\ \Pi &\in \{\zeta, \Delta, \Lambda, \Xi\} \text{ are the constraints among the variables } \{\Phi\}\\ \text{and } &\{\omega_g\}_{g=1,\dots,4} \in \{y', x', P_{y'}, P_{x'}\}\\ &\{\Omega_f\}_{f=1,\dots,4} \in \{\Xi', \Lambda, \zeta, \Delta\} \text{ are the new set of canonical variables.}\\ &\Omega_f \equiv 0\ \forall f \text{ are the new equivalent constraints in the new set.} \end{aligned}$$

The Dirac brackets between the variables in the new set, defined in (215), are:

$$\begin{aligned} \{y'(t), P_{y'}(t)\}_{D(\Omega)} &= 1\\ \{x'(t), P_{x'(t)}\}_{D(\Omega)} &= 1 \end{aligned}$$

And,

$$\{\Upsilon_a(x), \Upsilon_b(y)\}_{D(\Omega)} = 0 \tag{164}$$

for all the other Dirac brackets.

As was checked with the field theory, with (164) it is verified that the Dirac brackets in the canonical variables $\{\Omega\}$, can be considered as Poisson brackets only in the fields $\{\omega\} = \{y', x', P_{y'}, P_{x'}\}$. This is, the equation (124), that was discussed in the field theory, is also satisfied here $\{\cdot,\cdot\}^{\omega,\Omega}_{DiracB(\Omega)} = \{\cdot,\cdot\}^{\omega}_{PoissonB}$.

As was noted for the field theory, it is important to emphasize that $\{\Omega\} = \{\Xi, \zeta, \Delta, \Lambda\}$ are no longer constraints between canonical fields $\{\Phi\}$, but now, they are to be considered **canonical variables on their own**, that however, vanish identically when the equations of motion are satisfied, and therefore are not dynamical.

Now, the Hamiltonian (204) can be easily written in terms of the new set $\{\Upsilon\}$. Nevertheless, before making this computation, let us recall that still remains one Langrange multiplier, χ, to be defined. As was widely discussed in the section for field theory, this is not strictly necessary, because making use of some general theorems (see section 4.3.3), only the H_0 in the Hamiltonian must be transformed to the new set of variables. However, in this case, the Langrange multipliers are so simple, that it is worth to define the latter, write the complete Hamiltonian H and H_0, and verify that both Hamiltonians, written in the set $\{\Upsilon\}$, are the same.

The complete set of Lagrange multipliers, can be found by the inverse of the Dirac matrix (209) (See section 3), as:

$$\begin{pmatrix} \chi \\ V_x \\ V_y \\ V_\lambda \end{pmatrix} =: -\mathcal{D}^{-1} \begin{pmatrix} \{\Xi, H_0\} \\ \{\zeta, H_0\} \\ \{\Lambda, H_0\} \\ \{\Delta, H_0\} \end{pmatrix} = -\mathcal{D}^{-1} \begin{pmatrix} \frac{g}{\alpha}(-wy + \frac{P_Q}{\alpha}) \\ mx + \frac{g}{\alpha}(P_y - Q) \\ -wy \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{m\alpha+\beta}\Xi \\ \frac{g}{m\alpha+\beta}(wy - \frac{1}{\alpha}P_Q) \\ 0 \\ wy \end{pmatrix} \tag{165}$$

Therefore, we have the completeley defined Hamiltonian:

$$\begin{aligned} H &= H_0 + \left(\frac{g}{m\alpha+\beta}(wy - \frac{1}{\alpha}P_Q)\right)\zeta + wy\Delta + \left(\frac{\alpha}{m\alpha+\beta}\Xi\right)\Xi \\ H_0 &= -\frac{1}{2\alpha}P_Q^2 - \frac{1}{2}Q^2 + \frac{w}{2}y^2 - \frac{m}{2}x^2 + P_yQ \\ \zeta &=: P_x - \frac{g}{\alpha}P_Q \approx 0 \\ \Lambda &=: P_y + \lambda \approx 0 \\ \Delta &=: P_\lambda \approx 0 \\ \Xi &=: mx(t) + \frac{g}{\alpha}(P_y(t) - Q(t)) \approx 0 \end{aligned} \tag{166}$$

After writing both H and H_0, in terms of $\{\omega\} = \{y', x', P_{y'}, P_{x'}\}$ and the new constraints $\{\Omega\} = \{\zeta, \Lambda, \Delta, \Xi'\}$, which in fact coincide with the previous set $\{\Pi\}$ up to a multiplicative constant, we obtain exactly the same result up to terms containing the variables $\{\Omega\}$. However, as was extensively discussed in section (4.3.3), for the variables satisfying the equations of motion, weak equalities turn to be strong inside the Dirac brackets (not inside Poisson B.), *i.e.* $\Omega_f \equiv 0 \; \forall f$, and the physical Hamiltonian is,

$$H_{phys}(\omega) = \tilde{H}(\omega, \Omega)|_{\Omega \equiv 0} \tag{167}$$

Therefore, transforming H and H_0 we reach exactly the same result. With the condition (106), $g^2 \stackrel{!}{=} \alpha\beta$, the physical Hamiltonian in the new set of variables $\{\omega\}$ is:

$$H_{phys} = -\frac{\beta}{2}\frac{1}{(m\alpha+\beta)^2}P_{x'}^2 + \frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} - \frac{m}{2\beta}x'^2\,(m\alpha+\beta) \tag{168}$$

Let us stop the current analysis and come back to the function contributing as the zero momentum in the mode expansion, of the Hamiltonian density in field theory (158):

$$\begin{aligned} h_{phys_{\vec{p}=\vec{0}}} \;\approx\; & -\frac{\beta}{2}\frac{1}{(m\alpha+\beta)^2}P_\gamma^2 + \frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_\gamma + \eta\right)^2 - \frac{m}{2\beta}(m\alpha+\beta)\gamma^2 + \frac{1}{2}P_{\Theta_j}^2 \\ & + \frac{m}{2}(\Theta^i)^2 + \frac{P_\eta^2}{2} \end{aligned} \tag{169}$$

Now, let us point out that there are only two terms not possible to identify (upon re-interpretation of the objects) in (168), these are the ones containing $\Theta^i =: A^i$ and $P_{Theta_i} =: P_{A_i}$. The reason is clear if we recall the suggestive assignment when writing the Lagrangian for the classical model with a finite number of degrees of freedom. This is:

$$\begin{aligned} \varphi &\leftrightarrow y \\ A^0 &\leftrightarrow x \end{aligned} \tag{170}$$

This means that no degree of freedom in the present model, has a counterpart with the three fields A^i. However, as was pointed out in the related discussion in the previous section, this is not intended to be a complete analogy for many reasons, one of them, being the completely different interpretations of the objects (fields and single degrees of freedom, like x).

Nevertheless, the resemblance obtained between (169) $\leftrightarrow$ (168) is meaningful for two reasons: firstable, after such long and cumbersome procedure, of finding the physical Hamiltonian, carried out in completely independent formalisms, we have reached the "equivalent" answer for the finite degrees of freedom model, by starting from the field theory and after taking the corresponding approximations, *i.e.* taking only the 0-momentum mode of the mode expansion, re-interpretation of the objects: fields $\rightarrow$ demoting them to simple degrees of freedom, and recalling that we did not introduce the corresponding *dofs* to A^i in the model with finite *dofs*. This can be considered as a possible limiting case of field theory, that helps to relate a difficultier model, to a simpler one, whose properties are much more "accessible".

Second, and highly related to the last comment, we identify that also in this model, ther term $+\frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_\gamma+\eta\right)^2$ appears, which will help to understand the possibly problematic case in field theory, when $w \neq 0$ and $g \in \mathbb{C}$ (massive higher derivative scalar field).

Eliminating the Ostrogradsky's instability

As was done before in the field theory case, we are concerned with theories that do not include the Ostrogradsky's instability because among other things, in a quantum theory based on an ill classical theory, undesirable properties like ghost states in the Fock space, lost of unitarity or "unstable vacuum", appear. In fact, as was presented in section (3)(see [2], [3]), all of these undesirable properties can be tracked to the non-positive definiteness of the Hamiltonian. Therefore, we are mainly concerned with this problem.

A broader discussion regarding the necessary conditions to promote the classical theory to a quantum theory, by canonical quantization, were discussed in section 4.3.6, and here, only the main results motivated by that discussion, will be presented.

A short remark before going on, is that the non-positive definiteness of the Hamiltonian, may be present in a model for reasons completely different to that of Ostrogradsky's instability. But, for the present case, since the models were written with higher derivatives, it is sufficient to write a positive definite Hamiltonian in order to get rid of the problem, even if other dynamical instabilities are unintentionally also taken care of, by this procedure.

The possible approaches to deal with the described problem in this model, are much more, than the parameter-restricted field theory. Even though the restriction turned out to be very interesting, relating the coupling of the higher derivative term to the mass of the stabilizer vector field, it may be that this does not play a fundamental role in eliminating the Ostrogradsky's instability. In the commented field theory case, the only possibility to get rid of the Ostrogradsky's instability was to take: ($\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \neq 0$ $w \overset{!}{=} 0$, $\mathbb{C} \ni g \overset{!}{=} \pm i\sqrt{\alpha|\beta|} \to \alpha > \frac{|\beta|}{m}$), however, the condition $w \overset{!}{=} 0$ was only imposed "on the safe side", with the argument that in any case, this is a toy model and we only want to be sure that at least one field theory model can stabilized.

In what follows, the different possibilities to write a positive definite physical Hamiltonian, therefore eliminating the Ostrogradsky's instability, will be exposed. After the instability is dealt with, the theory could in principle be quantized, because a unique, "stable vacuum", could in principle be defined in Hilbert space. However this discussion can be highly non-trivial, and we only restrict to the weaker case, of assuring that at least, no Ostrogradsky's instability should cause trouble in a possible quantum theory. In that sense, for the cases below, after stating the conditions to eliminate the instability, they will be canonically quantized by the usual prescription of promoting $\{x', P_{x'}, y', P_{y'}\}$ to operators, and the Dirac brackets (Not Poisson) to commutators (171). For some interesting enough cases, the energy spectrum will be presented.

$$\{\cdot,\cdot\}_{D(\Omega)} \to \frac{1}{i}[\cdot,\cdot] \tag{171}$$

Finally, let us recall the initial Lagrangian, how the new set of variables are defined, and the physical Hamiltonian that was obtained after the Dirac's programme:

$$\begin{aligned} L(x,y) &= -\frac{1}{2}\alpha\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 - \frac{\beta}{2}\dot{x}^2 + \frac{m}{2}x^2 - g\ddot{y}\dot{x} \\ H_{phys} &= -\frac{\beta}{2}\frac{1}{(m\alpha+\beta)^2}P_{x'}^2 + \frac{w}{2}\left(\frac{g}{m\alpha+\beta}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} - \frac{m}{2\beta}x'^2\,(m\alpha+\beta) \end{aligned} \tag{172}$$

for,

$$\Phi =: \begin{pmatrix} Q \\ y \\ x \\ \lambda \\ P_Q \\ P_y \\ P_x \\ P_\lambda \end{pmatrix} \to \begin{pmatrix} (Q - P_y - \frac{\alpha}{g}mx) \\ y - P_Q \\ x \\ \lambda + P_y \\ \frac{m\alpha+\beta}{g}P_Q \\ P_y \\ P_x - \frac{g}{\alpha}P_Q \\ P_\lambda \end{pmatrix} =: \begin{pmatrix} \Xi' \\ y' \\ x' \\ \Lambda \\ P_{x'} \\ P_{y'} \\ \zeta \\ \Delta \end{pmatrix} := \Upsilon$$

- 1- Let $\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{=} 0$, $w \overset{!}{\geq} 0$, $\mathbb{C} \ni g \overset{!}{=} \pm i\sqrt{\alpha|\beta|}$. The Lagrangian is:

$$L(x,y) = -\frac{1}{2}\alpha\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 + \frac{|\beta|}{2}\dot{x}^2 \mp i\sqrt{\alpha|\beta|}\ddot{y}\dot{x} \tag{173}$$

With the procedure described above, the new set of canonical variables (215) in the Dirac brackets (211,212), and the quantization by promotion of the dynamical *dofs* (215) to operators, together with (171), the physical Hamiltonian operator is:

$$H_{phys} = \frac{1}{2|\beta|}P_{x'}^2 + \frac{w}{2}\left(\pm i\sqrt{\frac{\alpha}{|\beta|}}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} \tag{174}$$

To find the spectrum, let $|\Psi\rangle$ be a state in the domain of H_{phys}, in Hilbert space. Then, we have the time independent Schrödinger equation:

$$H_{phys}(x', y', P_{x'}, P_{y'})|\Psi\rangle = E|\Psi\rangle \tag{175}$$

Let us recall the commutators:

$$\begin{aligned} \{x', P_{x'}\}_{D(\Omega)} = 1 \to [x', P_{x'}] &= i \\ \{y', P_{y'}\}_{D(\Omega)} = 1 \to [y', P_{y'}] &= i \end{aligned} \tag{176}$$

It would be possible to take the "position basis" $\{|x', y'\rangle\}$ to project the states, but we notice that x' does not appear explicitly in H_{phys}, then, if we consider the set of states $\{|P_{x'}, y'\rangle\}$ as a basis, all computations will be highly simplified. In any case, any possibility to assign a "position operator" interpretation to $\{x', y'\}$ and "momentum operator" to $\{P_{x'}, P_{y'}\}$, has been lost in the cubersome redefinition of dynamical variables (215), that was necessary to quantize the system. Now, $\{x', y', P_{x'}, P_{y'}\}$ must be considered canonical variables by the pairs (176), without necessity of further classification.

Assuming that the basis is complete ($\int dP_{x'}dy'|P_{x'}, y'\rangle\langle P_{x'}, y'| = I$), we get the differential equation (now in eigenvalues of the $\{x', y', P_{x'}, P_{y'}\}$ operators, and we take units such that $\hbar = 1$):

$$\begin{aligned} \langle P_{x'}, y'|H_{phys}(x', y', P_{x'}, P_{y'})|\Psi\rangle &= E\langle P_{x'}, y'|I|\Psi\rangle \\ \left(-\frac{1}{2}\frac{\partial^2}{\partial y'^2} + \frac{w}{2}\left(\pm i\sqrt{\frac{\alpha}{|\beta|}}P_{x'} + y'\right)^2 - \left(E - \frac{1}{2|\beta|}P_{x'}^2\right)\right)\Psi(P_{x'}, y') &= 0 \end{aligned} \tag{177}$$

By displacing the $y' \to y''$ in the complex plane, we get the Schrödinger differential equation for a simple harmonic oscillator. We are only interested in the energy spectrum, which as always, after imposing the boundary condition that the wave function $\Psi(P_{x'}, y'')$ must be finite everywhere, we get the usual result:

$$\begin{aligned} E'_n &= E_n - \frac{1}{2|\beta|}P_{x'}^2 = \sqrt{w}\left(n + \frac{1}{2}\right) \\ &\downarrow \\ E_n(P_x) &= \sqrt{w}\left(n + \frac{1}{2}\right) + \frac{1}{2|\beta|}P_{x'}^2 \qquad (178) \\ n &\in \mathbb{N} \cup \{0\} \qquad (179) \end{aligned}$$

By the completeness of $\{|P_{x'}\rangle\}$ basis for $P_{x'}$ operator, the energy spcetrum is continuum. It can be seen from (178) that the spectrum is also real, positive and bounded from below.

This is a simple yet meaningful result. Firstable, it can be explicitly seen in the boundedness from below of the spectrum (178), that there is a unique vacuum in Hilbert space, whose non-existance was the biggest, and origin of all problems for the Ostrogradsky's instability [2, 3]. We can finally assure that this model with a finite number of degrees of freedom and with higher derivatives, can indeed be stabilized by the method proposed by [39] (i.e. demanding the existance of a particular kind of constraint. *i.e.* ζ in (204)). And furthermore, the stabilization does not get lost when quantizing the theory, which had not been exemplified before.

On the other hand, let us point out that the term $+\frac{w}{2}\left(\pm i\sqrt{\frac{\alpha}{|\beta|}}P_{x'} + y'\right)^2$ which was also present in the corresponding case in field theory, does not affect the desired properties of the energy spectrum. There is however, a difference between these two models besides that one, very important, of the interpretation of the objects under consideration. This is, we have taken here, besides ($\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $w \neq 0$, $\mathbb{C} \ni g \overset{!}{=} \pm i\sqrt{\alpha|\beta|}$) also $m \overset{!}{=} 0$, which was not the case in field theory. In fact, it was not possible at all in field theory, to remove *all* the possible sources of instability by taking $m = 0$. But as was discussed in section (4.3.6), it was identified that the conditions by means of which the physical Hamiltonian density, was found to have the desired properties, were sufficient but not necessary. In particular, it was discussed that there there may be other possibilities to avoid the Ostrogradsky's instability in those models, that could be reached by taking theories whose parameters were restricted to other regions in parameter space. These, however, may leave the model with certain dynamical instabilities not related with that of Ostrogradsky. In short, m could be taken to be zero in the field theory, and yet the Ostrogradsky instability seems to disappear, but further analysis should be done in that case.

Furthermore, as was also commented in (4.3.6), the mass term for the vector field, $\frac{m}{2}A_\mu A^\mu$, was not even demanded to be included in the model, by the stabilization procedure proposed in [39]. Hence, this may be another indication of the fact that the Ostrogradsky's instability can be removed even taking $m = 0$, for specific conditions on the other parameters. There is however, another possibility: since the procedure given in [39], was proposed for systems with finite degrees of freedom, which obviously do not say anything related to the extension to field theories. Problems may arise, and one could, in principle, for some not yet known reason, demand the existance of the term $\frac{m}{2}A_\mu A^\mu$, in order for the Ostrogradsky's instability to disappear. This is not a very unreasonable possibility. One example was already found in this work, when demanding explicit covariance in the Lagrangian density. It was identified, that it is necessary for the stabilizer field, to transform as a vector and not as a Lorentz scalar.

Finally, there is yet another point that may incline the previous discussion to the fact, that with the particular field theory proposed in section 4.3, we just introduced more possible sources of dynamical instability, which forced upon us the conditions ($\alpha, m \in \mathbb{R}^+$, $\beta \in \mathbb{R}^-$, $\mathbb{C} \ni g^2 \overset{!}{=} \alpha\beta$, $w \overset{!}{=} 0$, and $\alpha > \frac{|\beta|}{m}$) in order to guarantee ("on the safe side") the elimination of Ostrogradsky's instability. This is, there are much more possibilities to eliminate the instability in this model with finite degrees of freedom, than in field theory[40]

The following, are other possibilities to eliminate the Ostrogradsky's instability. Since the procedure and analysis is almost the same as for the first case, only the results will be given.

[40]We are making explicit reference to the fact that these two completely different models, in the sense of interpretation, may be compared upon some considerations. This was widely discussed at the end of section (4.3.3) and at the beginning of this section. It was roughly related to the fact that there is a possible limit, reaching the current model with finite degrees of freedom, from field theory, when taking the 0-momentum mode, of the mode expansion of the physical Hamiltonian density and then re-interpreting the objects being treated.

- 2- Let $\mathbb{R} \ni \alpha \overset{!}{<} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{=} 0$, $w \overset{!}{\geq} 0$, $\mathbb{R} \ni g \overset{!}{=} \pm\sqrt{|\alpha||\beta|}$. The Lagrangian is:

$$L(x,y) = \frac{1}{2}|\alpha|\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 + \frac{|\beta|}{2}\dot{x}^2 \mp \sqrt{|\alpha||\beta|}\ddot{y}\dot{x} \tag{180}$$

With the new set of canonical variables (215) in the Dirac brackets (211), after canonical quantization, the physical Hamiltonian operator is:

$$H_{phys} = \frac{1}{2|\beta|}P_{x'}^2 + \frac{w}{2}\left(\pm\sqrt{\frac{|\alpha|}{|\beta|}}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} \tag{181}$$

This is a very similar case to the first one. This Lagrangian no longer includes the Pais-Uhlenbeck oscillator, because we have taken $\alpha \overset{!}{<} 0$. However, this is still a higher-derivative degree of freedom, coupled to the stabilizer x. It is also important to note that $g \in \mathbb{R}$. The energy spectrum can be found as above, with almost the same results:

$$\begin{aligned} E_n(P_x) &= \sqrt{w}\left(n + \frac{1}{2}\right) + \frac{1}{2|\beta|}P_{x'}^2 \\ n &\in \mathbb{N} \cup \{0\} \end{aligned} \tag{182}$$

The only difference is not noticeable in the spectrum but in the wave functions, which now depend on a variable y'', obtained after displacing the origin strictly in the real line, $y' \to y''$, and not in the complex plane, as in the first case. Nevertheless, the important result for our present purposes is that again, the spectrum is bounded from below, positive, and in this particular case, also continuous. Again, we have explicitly verified the elimination of Ostrogradsky's instability.

- Now, let us explore the more general case of $m \neq 0$:

3- Let $\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{>} 0$, $w \overset{!}{\geq} 0$, $\mathbb{C} \ni g \overset{!}{=} \pm i\sqrt{\alpha|\beta|}$. The Lagrangian is:

$$L(x,y) = -\frac{1}{2}\alpha\ddot{y}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}wy^2 + \frac{|\beta|}{2}\dot{x}^2 + \frac{m}{2}x^2 \mp i\sqrt{\alpha|\beta|}\ddot{y}\dot{x} \tag{183}$$

With the new set of canonical variables (215) in the Dirac brackets (211), after canonical quantization only when $\alpha > \frac{|\beta|}{m}$, the physical Hamiltonian operator is:

$$H_{phys} = \frac{|\beta|}{2}\frac{1}{(m\alpha - |\beta|)^2}P_{x'}^2 + \frac{w}{2}\left(\pm i\frac{\sqrt{\alpha|\beta|}}{(\alpha m - |\beta|)}P_{x'} + y'\right)^2 + \frac{P_{y'}^2}{2} + \frac{m}{2|\beta|}x'^2(\alpha m - |\beta|) \tag{184}$$

We encounter again a complex term but instead of computing the energy spectrum, which in this case requires more work (because all variables are now dynamical), we pass to a simpler case which has better properties. It is worth noting that if we make $m = 0$ in this model, we recover the first case with the nice spectrum described above. Furthermore, if we set $w \stackrel{!}{=} 0$ we recover an equivalent model to that described in the field theory (154) in section 4.3.3. In particular, let us write again, for matter of comparison, both, the model in quantum field theory, and its 0-momentum component of its mode expansion:

$$\begin{aligned}\mathcal{H}_{phys} \quad (\omega) \quad &= \frac{|\beta|}{2}\frac{1}{(m\alpha-|\beta|)^2}\left(P_\gamma^2+\alpha\nabla P_\gamma\cdot\nabla P_\gamma\right)+\frac{1}{4}(F^{ij})^2+\frac{m}{2|\beta|}(m\alpha-|\beta|)\gamma^2\\ &+ \frac{1}{2}\left(P_{\Theta_j}^2+\frac{\alpha}{m\alpha-|\beta|}\nabla P_{\Theta_j}\cdot\nabla P_{\Theta_j}\right)+\frac{m}{2}(\Theta^i)^2+\frac{P_\eta^2}{2}+\frac{1}{2}\nabla\eta\cdot\nabla\eta\end{aligned}$$

Let us recall that this function acquires the status of the physical Hamiltonian density operator only after demanding $\alpha > \frac{|\beta|}{m}$, which is the same condition for (184). And the 0-momentum function of the mode expansion, after demanding $\mathbb{R} \ni \alpha \stackrel{!}{>} 0$, $\mathbb{R} \ni \beta \stackrel{!}{<} 0$, $m \stackrel{!}{>} 0$, $w \stackrel{!}{=} 0$, $\mathbb{C} \ni g \stackrel{!}{=} \pm i\sqrt{\alpha|\beta|}$:

$$\begin{aligned}h_{phys_{\vec{p}=\vec{0}}} \quad &\approx \frac{|\beta|}{2}\frac{1}{(m\alpha-|\beta|)^2}P_\gamma^2+\frac{P_\eta^2}{2}+\frac{m}{2|\beta|}\gamma^2(m\alpha-|\beta|)+\frac{1}{2}P_{\Theta_j}^2+\frac{m}{2}(\Theta^i)^2\\ &\updownarrow\\ H_{phys} \quad &= \frac{|\beta|}{2}\frac{1}{(m\alpha-|\beta|)^2}P_{x'}^2+\frac{P_{y'}^2}{2}+\frac{m}{2|\beta|}x'^2(\alpha m-|\beta|)\end{aligned} \tag{185}$$

As commented before, there are only two terms not possible to identify (upon re-interpretation of the objects) in (185), these are the ones containing $\Theta^i =: A^i$ and $P_{\Theta_i} =: P_{A_i}$. Let us recall that the following was a possible assignment: $\varphi \leftrightarrow y$ and $A^0 \leftrightarrow x$. Leaving A^i without counterpart in the current model with only finite degrees of freedom.

At the end of section (4.3.6) a wide discussion on the interesting condition $\alpha > \frac{|\beta|}{m}$ was given. However, in the following case, additional comments will be given.

- 4- Let $\mathbb{R} \ni \alpha \stackrel{!}{<} 0$, $\mathbb{R} \ni \beta \stackrel{!}{<} 0$, $m \stackrel{!}{<} 0$, $w \stackrel{!}{\geq} 0$, $\mathbb{R} \ni g \stackrel{!}{=} \pm\sqrt{\alpha|\beta|}$. The Lagrangian is:

$$L(x,y) = \frac{1}{2}|\alpha|\ddot{y}^2+\frac{1}{2}\dot{y}^2-\frac{1}{2}wy^2+\frac{|\beta|}{2}\dot{x}^2-\frac{|m|}{2}x^2 \mp i\sqrt{\alpha|\beta|}\ddot{y}\dot{x} \tag{186}$$

With the new set of canonical variables (215) in the Dirac brackets (211), after canonical quantization only when $|\alpha| < \frac{|\beta|}{|m|}$, the physical Hamiltonian operator is:

$$H_{phys} = \frac{|\beta|}{2}\frac{1}{(|m||\alpha|-|\beta|)^2}P_{x'}^2+\frac{w}{2}\left(\pm\frac{\sqrt{|\alpha||\beta|}}{(|\alpha||m|-|\beta|)}P_{x'}+y'\right)^2+\frac{P_{y'}^2}{2}+\frac{|m|}{2|\beta|}x'^2(|\beta|-|\alpha||m|) \tag{187}$$

Notice that this is a much more interesting case, because now α has an upper limit defined by the mass of the stabilizer $|\alpha| < \frac{|\beta|}{|m|}$. If two degrees of freedom x and y, the latter possibly with higher derivatives, describe some physical system, then, one could say that the higher derivative terms may play a role in the evolution of the system, but only under certain conditions can be detected. For instance, if the mass parameter of x were too small, we would know the order of high energies, given by the mentioned upper limit, at which the higher derivatives should be found to affect the dynamics of the y degree of freedom. Such estblishment of an order of magnitude of energies, at which higher derivatives may describe real dynamics, could also be obtained in the case found for fields. Neverthelss, the relation $\alpha > \frac{1}{m}$ is not as meaningful, because if no information is found at experiments in certain range of energies, it tells nothing about which should be the next try, but rather, only says that the effects may take importance above the already tried range of energies. It is important to note that this analysis, though in the finite degree of freedom case, is mainly given thinking of possible extensions to QFT, as the one found in section (4.3.6), but with the more interesting relation $|\alpha| < \frac{1}{m}$.

However, we must be very careful with this analysis for two main reasons: firstable this is only a toy model, which only suggests that at least in an example (the one already found), such relation between mass parameters of stabilizers and coupling to the dynamics of possibly higher derivatives of *dofs*, may arise if we desire a stable model. Second, because at least in the finite degree of freedom case, other possibilities to stabilize without the mentioned relation $\alpha(m)$, were found.

Finally, a point supporting the idea that this is in fact an interesting result: this is the only toy model with the "right" signs in the Lagrangian for the stabilizer x, *i.e.* x is now a harmonic oscillator. It must be noted that y however, is no longer the Pais-Uhlenbeck oscillator, but only a particular higher derivative model, with well properties given by its interaction with the harmonic oscillator x.

- Let us note that there is another possibility for the Hamiltonian (168) to be free of Ostrgradsky's instability. This can be reached by demanding ($\mathbb{R} \ni \alpha \overset{!}{>} 0$, $\mathbb{R} \ni \beta \overset{!}{<} 0$, $m \overset{!}{<} 0$, $w \overset{!}{\geq} 0$, $\mathbb{C} \ni g \overset{!}{=} \pm i\sqrt{\alpha|\beta|}$), but it is quite similar to those described above and is not worth to write it explicitly.

5 Conclusions

The main intention with this thesis has been to contribute in the understanding of the properties that arise when new, higher-derivative structure, is conferred to the mathematical models that may be claimed to describe some dynamics of some physical phenomena. The discussion has been centered on a toy model that was specifically built, such that all the key properties were included, in order to analyse how they could be put together to give a nice mathematical model, *i.e.* with healthy properties that would in principle, allow quantization of the classical theory. By key properties we refer to those commonly assigned to higher-derivative theories, in particular, the Ostrogradsky's instability and also to the stabilization conditions suggested in [39].

More precisely, a continuum higher-derivative scalar field theory was proposed in this thesis. On its own, without any interaction, it leads to an unstable theory. This fact was already known for the Pais-Uhlenbeck model [2, 3, 6, 7, 8, 9, 10, 11]. However, by a naive application of the stabilization procedure that was proposed only for systems with finitely many DOFs [39], it was found that the scalar field theory requires of another field, such that between their momenta a very specifical kinetic constraint exists and controls the instability. In the case of field theory, additional structure of the stabilizer field was found to be required. In particular, the stabilizer needs to transform like a vector field, in order for this theory to be explicitly covariant. Something that was not anticipated in [39, 40, 41]. It was also argued that this tranformation condition on the stabilizer cannot be bypassed if the instability is to be controled already at the level of a free theory, which is mandatory, if one expects a possible extension of the stable properties to interacting higher derivative theories. Roughly, the discussion centered around the need of a healthy free theory such that the Feynman propagator does not show a ghost DOF.

Starting from such a model, and imposing the constraint that was naively expected to control the Ostrogradsky's instability, a Hamiltonization with constraints in the extended Hamiltonian formalism, was carried out. More precisely, the complete constraint content of the theory, the Dirac brackets, the new set of dynamical fields and the physical Hamiltonian were found. This was also done for some models with a finite number of degrees of freedom. This equivalent system had much nicer properties that greatly simplified the analysis of the dynamics of the system. Therefore, upon the classical Physical Hamiltonian propagating physical healthy DOFs, the conditions for which this function was positively defined and bounded from below, were found. This guaranteed that the Ostrogradskian instability had been systematically removed.

Furthermore, related to the interpretation of this new structure, it was found that the only way to stabilize the higher-derivative scalar field, forces a relation between the coupling parameter of the higher-derivative scalar field (α), and the mass parameter of the stabilizer field (m). The condition is a lower bound of the kind $\alpha > 1/m$. Such relation was completely unexpected but more meaningful for the physical interpretation of higher-derivative terms. For the cases with finite DOFs, a model with upper bound was also found.

Since the model without stabilization conditions is manifestly unstable, then, only the imposed conditions could have lead to the new nice properties of the theory, and it is concluded that after some subtleties regarding the construction of the model and the transformation properties of the stabilizer field, it is possible to to systematically build a field theory with higher derivative structure that can be, in principle, be brought to a healthy quantum theory. In particular, the Dirac brackets, which include by construction all the constraint content of the theory, were promoted to commutators for the initial set of canonical fields. Equivalently, the same canonical quantization was proposed directly for the physical DOFs propagated by the theory, which are canonical in the Dirac brackets, and therefore, for this set of variables, the promotion to commutators was simply taken from the Poisson brackets.

Now, in a quantum field theory, the restriction for stability relating the higher-derivative coupling and the mass of the stabilizer field ($\alpha > 1/m$), could be interpreted at least inside the very restricted panorama allowed by this simple toy model, as a lower bound on energies at which a higher-derivative term, may indeed, appreciably describe some dynamics in a scattering process.

Even though it is not claimed that this kind of relation should appear in every possible stable higher-derivative field theory, the sole fact that this toy model with these interesting properties exists, speaks about possibly physically interesting cases.

6 Appendix

6.1 A1. Some results for section 4.3

- The Dirac brackets for the theory (123), with (129) and (130), are:

$$
\begin{aligned}
\{\mathcal{F}(x),\mathcal{G}(y)\}_{D(\Pi)} = \{\mathcal{F}(x),\mathcal{G}(y)\} \; - & \int d^3z d^3z' \Big(\{\mathcal{F}(x),\zeta_0(z)\}\{\zeta_0(z),\zeta_j(z')\}^{-1}\{\zeta_j(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\zeta_0(z)\}\{\zeta_0(z),\Xi_1(z')\}^{-1}\{\Xi_1(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\zeta_i(z)\}\{\zeta_i(z),\zeta_0(z')\}^{-1}\{\zeta_0(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\zeta_i(z)\}\{\zeta_i(z),\Delta_0(z')\}^{-1}\{\Delta_0(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\zeta_i(z)\}\{\zeta_i(z),\Xi_{3_j}(z')\}^{-1}\{\Xi_{3_j}(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Delta_0(z)\}\{\Delta_0(z),\zeta_j(z')\}^{-1}\{\zeta_j(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Delta_0(z)\}\{\Delta_0(z),\Delta_j(z')\}^{-1}\{\Delta_j(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Delta_0(z)\}\{\Delta_0(z),\Lambda^0(z')\}^{-1}\{\Lambda^0(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Delta_i(z)\}\{\Delta_i(z),\Delta_0(z')\}^{-1}\{\Delta_0(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Delta_i(z)\}\{\Delta_i(z),\Xi_{2_j}(z')\}^{-1}\{\Xi_{2_j}(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Delta_i(z)\}\{\Delta_i(z),\Xi_{3_j}(z')\}^{-1}\{\Xi_{3_j}(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Lambda^0(z)\}\{\Lambda^0(z),\Delta_0(z')\}^{-1}\{\Delta_0(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Xi_1(z)\}\{\Xi_1(z),\zeta_0(z')\}^{-1}\{\zeta_0(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Xi_{2_i}(z)\}\{\Xi_{2_i}(z),\Delta_j(z')\}^{-1}\{\Delta_j(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Xi_{3_i}(z)\}\{\Xi_{3_i}(z),\zeta_j(z')\}^{-1}\{\zeta_j(z'),\mathcal{G}(y)\} \\
+ & \; \{\mathcal{F}(x),\Xi_{3_i}(z)\}\{\Xi_{3_i}(z),\Delta_j(z')\}^{-1}\{\Delta_j(z'),\mathcal{G}(y)\}\Big)
\end{aligned}
\tag{188}
$$

where,

$$
\begin{aligned}
\zeta_0(x) &:= P_{A_0}(x) - \frac{g}{\alpha}P_{\psi_0}(x) \approx 0 \\
\zeta_i(x) &:= P_{\psi_i}(x) \approx 0 \\
\Delta_\mu(x) &:= P_{\lambda_\mu}(x) \approx 0 \\
\Lambda^0(x) &:= P_\varphi(x) + \lambda^0(x) \approx 0 \\
\Xi_1(x) &:= mA_0(x) + \frac{g}{\alpha}(P_\varphi(x) - \psi_0(x)) - \partial_i P_{A_i}(x) \approx 0 \\
\Xi_{2_i}(x) &:= -(\psi^i(x) + \lambda^i(x) + \partial_i P_{\psi_0}(x)) \approx 0 \\
\Xi_{3_i}(x) &:= -\psi^i(x) - \partial_i\varphi(x) \approx 0
\end{aligned}
$$

For the computation of the new set of canonical fields, an alternative form of the Dirac brackets, already developed in terms of the canonical fields (after using the fundamental Poisson brackets (101)), is:

$$
\begin{aligned}
\{\mathcal{F}(x),\mathcal{G}(y)\}_{D(\Pi)} &= \{\mathcal{F}(x),\mathcal{G}(y)\}\\
&- \int d^3z d^3z' \Big(- \frac{\frac{g}{\alpha}\delta_{ij}}{m+\frac{g^2}{\alpha^2}} \{\mathcal{F}(x),(P_{A_0}-\frac{g}{\alpha}P_{\psi_0})(z)\}\{\partial_i^{(z')}P_{\psi_j}(z'),\mathcal{G}(y)\}\\
&+ \frac{1}{m+\frac{g^2}{\alpha^2}}\{\mathcal{F}(x),(P_{A_0}-\frac{g}{\alpha}P_{\psi_0})(z)\}\{(mA_0+\frac{g}{\alpha}(P_\varphi-\psi_0)-\partial_i^{(z')}P_{A_i})(z'),\mathcal{G}(y)\}\\
&+ \frac{1}{m+\frac{g^2}{\alpha^2}}\frac{g}{\alpha}\delta_{ij}\{\mathcal{F}(x),\partial_j^{(z)}P_{\psi_i}(z)\}\{(P_{A_0}-\frac{g}{\alpha}P_{\psi_0})(z'),\mathcal{G}(y)\}\\
&+ \delta_{ij}\{\mathcal{F}(x),\partial_j^{(z)}P_{\psi_i}(z)\}\{P_{\lambda_0}(z'),\mathcal{G}(y)\}\\
&- \delta_{ij}\{\mathcal{F}(x),P_{\psi_i}(z)\}\{(-\psi^j-\partial_j^{(z')}\varphi)(z'),\mathcal{G}(y)\}\\
&- \delta_{ij}\{\mathcal{F}(x),P_{\lambda_0}(z)\}\{\partial_i^{(z')}P_{\psi_j}(z'),\mathcal{G}(y)\}\\
&+ \delta_{ij}\{\mathcal{F}(x),P_{\lambda_0}(z)\}\{\partial_i^{(z')}P_{\lambda_j}(z'),\mathcal{G}(y)\}\\
&+ \{\mathcal{F}(x),P_{\lambda_0}(z)\}\{(P_\varphi+\lambda^0)(z'),\mathcal{G}(y)\}\\
&- \delta_{ij}\{\mathcal{F}(x),\partial_j^{(z)}P_{\lambda_i}(z)\}\{P_{\lambda_0}(z'),\mathcal{G}(y)\}\\
&- \delta_{ij}\{\mathcal{F}(x),P_{\lambda_i}(z)\}\{-(\psi^j+\lambda^j+\partial_j^(z')P_{\psi_0})(z'),\mathcal{G}(y)\}\\
&+ \delta_{ij}\{\mathcal{F}(x),P_{\lambda_i}(z)\}\{(-\psi^j-\partial_j^{(z')}\varphi)(z'),\mathcal{G}(y)\}\\
&- \{\mathcal{F}(x),(P_\varphi+\lambda^0)(z)\}\{P_{\lambda_0}(z'),\mathcal{G}(y)\}\\
&- \frac{1}{m+\frac{g^2}{\alpha^2}}\{\mathcal{F}(x),(mA_0+\frac{g}{\alpha}(P_\varphi-\psi_0)-\partial_i^{(z)}P_{A_i})(z)\}\{(P_{A_0}-\frac{g}{\alpha}P_{\psi_0})(z'),\mathcal{G}(y)\}\\
&+ \delta_{ij}\{\mathcal{F}(x),-(\psi^i+\lambda^i+\partial_i^{(z)}P_{\psi_0})(z)\}\{P_{\lambda_j}(z'),\mathcal{G}(y)\}\\
&+ \delta_{ij}\{\mathcal{F}(x),(-\psi^i-\partial_i^{(z)}\varphi)(z)\}\{P_{\psi_j}(z'),\mathcal{G}(y)\}\\
&- \delta_{ij}\{\mathcal{F}(x),(-\psi^i-\partial_i^{(z)}\varphi)(z)\}\{P_{\lambda_j}(z'),\mathcal{G}(y)\}\Big)\delta^{(3)}(\vec{z'}-\vec{z})
\end{aligned}
\tag{189}
$$

This is the final form used for all the computations.

In the search for the new set of canonical variables, the square extended matrix of Dirac brackets between the initial canonical fields, is formed by the following results:

$$
\begin{aligned}
\{\psi^\mu(x),\psi^\nu(y)\}_{D(\Pi)} &= \frac{g^2}{\alpha^2}\frac{1}{m+\dfrac{g^2}{\alpha^2}}\left(\delta_{\mu0}\delta_{\nu j}\partial_j^{(y)}\delta^{(3)}(\vec{y}-\vec{x})-\delta_{\mu i}\delta_{\nu 0}\partial_i^{(x)}\delta^{(3)}(\vec{x}-\vec{y})\right)\\
\{\psi^\mu(x),\varphi(y)\}_{D(\Pi)} &= -\frac{g^2}{\alpha^2}\frac{1}{m+\dfrac{g^2}{\alpha^2}}\delta_{\mu0}\delta^{(3)}(\vec{y}-\vec{x}))\\
\{\psi^\mu(x),A^\nu(y)\}_{D(\Pi)} &= \frac{g}{\alpha}\frac{1}{m+\dfrac{g^2}{\alpha^2}}\left(\delta_{\mu0}\delta_{\nu i}+\delta_{\mu i}\delta_{\nu 0}\right)\\
\{\psi^\mu(x),\lambda^\nu(y)\}_{D(\Pi)} &= \delta_{\mu i}\delta_{\nu 0}\partial_i^{(x)}\delta^{(3)}(\vec{x}-\vec{y})-\delta_{ij}\delta_{\mu0}\delta_{\mu0}\partial_j^{(y)}\delta^{(3)}(\vec{y}-\vec{x})\\
\{\psi^\mu(x),P_{\psi_\nu}(y)\}_{D(\Pi)} &= \left(\delta_{\mu\nu}-\frac{g^2}{\alpha^2}\frac{1}{m+\dfrac{g^2}{\alpha^2}}\delta_{\mu0}\delta_{\nu0}-\delta_{ij}\delta_{\mu i}\delta_{\nu j}\right)\delta^{(3)}(\vec{x}-\vec{y})\\
\{\psi^\mu(x),P_{\varphi}(y)\}_{D(\Pi)} &= -\delta_{ij}\delta_{\mu i}\partial_j^{(x)}\delta^{(3)}(\vec{x}-\vec{y})\\
\{\psi^\mu(x),P_{A_\nu}(y)\}_{D(\Pi)} &= \frac{g}{\alpha}\frac{1}{m+\dfrac{g^2}{\alpha^2}}m\delta_{\mu0}\delta_{\nu0}\delta^{(3)}(\vec{x}-\vec{y})\\
\{\psi^\mu(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= 0\\
\{\varphi(x),\varphi(y)\}_{D(\Pi)} &= 0\\
\{\varphi(x),A^\mu(y)\}_{D(\Pi)} &= -\frac{g}{\alpha}\frac{1}{m+\dfrac{g^2}{\alpha^2}}\delta_{\mu0}\delta^{(3)}(\vec{x}-\vec{y})\\
\{\varphi(x),\lambda^\mu(y)\}_{D(\Pi)} &= -\delta_{\mu0}\delta^{(3)}(\vec{x}-\vec{y})\\
\{\varphi(x),P_{\psi_\mu}(y)\}_{D(\Pi)} &= 0\\
\{\varphi(x),P_{\varphi}(y)\}_{D(\Pi)} &= \delta^{(3)}(\vec{x}-\vec{y})\\
\{\varphi(x),P_{A_\mu}(y)\}_{D(\Pi)} &= 0\\
\{\varphi(x),P_{\lambda_\mu}(y)\}_{D(\Pi)} &= 0\\
\{A^\mu(x),A^\nu(y)\}_{D(\Pi)} &= \frac{1}{m+\dfrac{g^2}{\alpha^2}}\left(\delta_{\mu i}\delta_{\nu0}\partial_i^{(y)}\delta^{(3)}(\vec{y}-\vec{x})-\delta_{\mu0}\delta_{\nu i}\partial_i^{(x)}\delta^{(3)}(\vec{x}-\vec{y})\right)\\
\{A^\mu(x),\lambda^\nu(y)\}_{D(\Pi)} &= 0\\
\{A^\mu(x),P_{\psi_\nu}(y)\}_{D(\Pi)} &= \frac{g}{\alpha}\frac{1}{m+\dfrac{g^2}{\alpha^2}}\delta_{\mu0}\delta_{\nu0}\delta^{(3)}(\vec{x}-\vec{y})\\
\{A^\mu(x),P_{\varphi}(y)\}_{D(\Pi)} &= 0\\
\{A^\mu(x),P_{A_\nu}(y)\}_{D(\Pi)} &= \left(\delta_{\mu\nu}-\frac{m}{m+\dfrac{g^2}{\alpha^2}}\delta_{\mu0}\delta_{\nu0}\right)\delta^{(3)}(\vec{x}-\vec{y})\\
\{A^\mu(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= 0
\end{aligned}
$$

$$
\begin{aligned}
\{\lambda^{\mu}(x),\lambda^{\nu}(y)\}_{D(\Pi)} &= \left(\delta_{\mu 0}\delta_{\nu j}\partial_j^{(y)}\delta^{(3)}(\vec{y}-\vec{x}) - \delta_{\mu i}\delta_{\nu 0}\partial_i^{(x)}\delta^{(3)}(\vec{x}-\vec{y})\right)\\
\{\lambda^{\mu}(x),P_{\psi_\nu}(y)\}_{D(\Pi)} &= 0\\
\{\lambda^{\mu}(x),P_{\varphi}(y)\}_{D(\Pi)} &= \delta_{ij}\delta_{\mu i}\partial_j^{(x)}\delta^{(3)}(\vec{x}-\vec{y})\\
\{\lambda^{\mu}(x),P_{A_\nu}(y)\}_{D(\Pi)} &= 0\\
\{\lambda^{\mu}(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= \left(\delta_{\mu\nu} - \delta_{\mu 0}\delta_{\nu 0} - \delta_{ij}\delta_{\mu i}\delta_{\nu j}\right)\delta^{(3)}(\vec{x}-\vec{y})\\
\{P_{\psi_\mu}(x),P_{\psi_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{\psi_\mu}(x),P_{\varphi}(y)\}_{D(\Pi)} &= 0\\
\{P_{\psi_\mu}(x),P_{A_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{\psi_\mu}(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{\varphi}(x),P_{\varphi}(y)\}_{D(\Pi)} &= 0\\
\{P_{\varphi}(x),P_{A_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{\varphi}(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{A_\mu}(x),P_{A_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{A_\mu}(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= 0\\
\{P_{\lambda_\mu}(x),P_{\lambda_\nu}(y)\}_{D(\Pi)} &= 0
\end{aligned}
\tag{190}
$$

Based on this results, we define the extended matrix of Dirac brackets by adding $\{\partial_i P_{\psi_0}, \partial_i P_{A_i}, \partial_i \varphi\}$ to the set, above. This is, for $\tilde{\Phi} \in \{\psi^{\mu}, P_{\psi_\mu}, A^{\mu}, P_{A_\mu}, \lambda^{\mu}, P_{\lambda_\mu}, \varphi, P_{\varphi}, \partial_i P_{\psi_0}, \partial_i P_{A_i}, \partial_i \varphi\}$. The entries of the extended matrix of Dirac brackets are:

$$
\tilde{J}_{ab} = \{\tilde{\Phi}_a, \tilde{\Phi}_b\}_{D(\Pi)} \tag{191}
$$

this matrix (33×33), is schematically presented. It is built using the results (189), and only the first row, specifically the last three columns emphasizes specific form of this matrix.

$$
\tilde{J}_{\{\cdot,\cdot\}} = \begin{pmatrix}
0 & \frac{g^2}{\alpha^2}\frac{1}{m+\frac{g^2}{\alpha^2}}\partial_j^{(y)}\delta^{(3)}(\vec{y}-\vec{x}) & \cdots & \cdots & \cdots & -\left(1-\frac{g^2}{\alpha 2}\frac{1}{m+\frac{g^2}{\alpha^2}}\right)\partial_j^{(y)}\delta^{(3)}(\vec{y}-\vec{x}) & 0 & -\frac{g^2}{\alpha^2}\frac{1}{m+\frac{g^2}{\alpha^2}}\partial_j^{(y)}\delta^{(3)}(\vec{y}-\vec{x}) \\
 & \ddots & & & & & & \\
\cdots & & & & & & & \ddots
\end{pmatrix}
\tag{192}
$$

6.2 A2. Some results for section 4.4

Finding the new set of canonical variables

As was discussed in section (4.4), the procedure to be presented is very similar to the one followed in section (4.3), therefore the search for the new set of variables, although with less comments, and much less broader discussion, is given here. This will be put in constant reference to the development of section (4.3). It must be noted that even though this computations are similar in form, the objects being tretated are completely different. Therefore, the computations were completely independent.

Let us define an equivalent lagrangian to (160) by introducing $Q =: \dot{y}$ and one Lagrange multiplier λ. (See (99)):

$$L(x, y, Q, \lambda)_{eq} = -\frac{1}{2}\alpha\dot{Q}^2 + \frac{1}{2}Q^2 - \frac{1}{2}wy^2 - \frac{\beta}{2}\dot{x}^2 + \frac{m}{2}x^2 - g\dot{Q}\dot{x} + \lambda(Q - \dot{y}) \tag{193}$$

the canonical momenta are:

$$\begin{aligned} P_Q &=: \frac{\partial L_{eq}}{\partial \dot{Q}} = -\alpha\dot{Q} - g\dot{x} \\ P_y &=: \frac{\partial L_{eq}}{\partial \dot{y}} = -\lambda \\ P_x &=: \frac{\partial L_{eq}}{\partial \dot{x}} = -\beta\dot{x} - g\dot{Q} \\ P_\lambda &=: 0 \end{aligned} \tag{194}$$

The basic Poisson brackets are:

$$\{x, P_x\} = 1 \quad \{y, P_y\} = 1 \quad \{Q, P_Q\} = 1 \quad \{\lambda, P_\lambda\} = 1 \tag{195}$$

Following the discussion in (160), we can compute the Hessian matrix $M_{ab} =: \frac{\partial^2 L_{eq}}{\partial \dot{q}_i \partial \dot{q}_j}$, with $q_i \in \{x, y, Q, \lambda\}$ for $i = 1, 2, 3, 4$, which in this case has nullity 2. The degeneracy is introduced by the construction of L_{eq} and has nothing to do with the original model (160). Therefore, we can focus on a submatrix of the Hessian matrix, given by

$$\begin{pmatrix} \delta P_Q \\ \delta P_x \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 L_{eq}}{\partial \dot{Q}^2} & \frac{\partial^2 L_{eq}}{\partial \dot{Q}\partial \dot{x}} \\ \frac{\partial^2 L_{eq}}{\partial \dot{x}\partial \dot{Q}} & \frac{\partial^2 L_{eq}}{\partial \dot{x}^2} \end{pmatrix} \begin{pmatrix} \delta \dot{Q} \\ \delta \dot{x} \end{pmatrix} \tag{196}$$

$$\begin{pmatrix} \delta P_Q \\ \delta P_x \end{pmatrix} = \begin{pmatrix} -\alpha & -g \\ -g & -\beta \end{pmatrix} \begin{pmatrix} \delta \dot{Q} \\ \delta \dot{x} \end{pmatrix} \tag{197}$$

therefore, according to the results of stabilization in [39], that were originally derived for these kind of systems with a finite number of degrees of freedom, we must impose a degeneracy on this submatrix of the Hessian. This implies the existance of a constraint in the theory, which at the end should control the Ostrogradsky's instability. And, as derived for fields, the condition is $g^2 \stackrel{!}{=} \alpha\beta$.

There are three primarily unexpressible velocities (or non-invertible in terms of momenta) $\dot{\lambda}, \dot{y}$ and either $\dot{x}$ or $\dot{Q}$, because their momenta are linearly dependent. As before, we pick $\dot{x}$ to be unexpressible, and for the reasons given for fields, the final result is independent of this choice. We can therefore, form the following function:

$$f = P_Q\dot{Q} + P_x\dot{x} + P_y\dot{y} + P_\lambda\dot{\lambda} - L_{eq} \quad (198)$$

Which after the introduction of the yet undetermined functions $V_y = \dot{y}$, $V_\lambda = \dot{\lambda}$ and $V_x = \dot{x}$, and after using the definition of momenta (194) we find the following Hamiltonian in the extended formalism, where we also identify the functions $V_y = \dot{y}$, $V_\lambda = \dot{\lambda}$, $V_x = \dot{x}$ as Lagrange multipliers:

$$\begin{aligned} H^{(1)} &= H_0 + V_x\zeta + V_y\Lambda + V_\lambda\Delta \\ H_0 &= -\frac{1}{2\alpha}P_Q^2 - \frac{1}{2}Q^2 + \frac{w}{2}y^2 - \frac{m}{2}x^2 + P_yQ \\ \zeta &=: P_x - \frac{g}{\alpha}P_Q \approx 0 \\ \Lambda &=: P_y + \lambda \approx 0 \\ \Delta &=: P_\lambda \approx 0 \end{aligned} \quad (199)$$

Furthermore, we also defined the primary constraints, which vanish weakly, after computing the evolution of the dynamical variables of the system, and setting them to follow the equations of motion (For a much broader discussion see section (4.3)). Their Poisson brackets are:

$$\{\zeta, \lambda\} = 0 \quad \{\zeta, \Delta\} = 0 \quad \{\Lambda, \Delta\} = 1 \quad (200)$$

Now, as described above, let us impose the conservation of constraints in time. this is:

$$\begin{aligned} \dot{\zeta}(t) &= \{\zeta(t), \mathcal{H}^{(1)}(t)\} \stackrel{!}{=} 0 \\ \dot{\zeta}(t) &= mx(t) + \frac{g}{\alpha}(P_y(t) - Q(t)) \stackrel{!}{=} 0 \end{aligned}$$

$\Rightarrow$ Define it as a secondary constraint:

$$\Xi =: mx(t) + \frac{g}{\alpha}(P_y(t) - Q(t)) \approx 0 \quad (201)$$

$$\begin{aligned}
\dot{\Delta}(t) &= \{\Delta(t), \mathcal{H}^{(1)}(t)\} \stackrel{!}{=} 0 \\
\dot{\Delta}(t) &= -V_y \stackrel{!}{=} 0
\end{aligned}$$

$\Rightarrow$ Fix the Lagrange multiplier:

$$V_y =: 0 \tag{202}$$

$$\begin{aligned}
\dot{\Lambda}(t) &= \{\Lambda(t), \mathcal{H}^{(1)}(t)\} \stackrel{!}{=} 0 \\
\dot{\Delta}(t) &= -wy(t) - V_\lambda \stackrel{!}{=} 0
\end{aligned}$$

$\Rightarrow$ Fix the Lagrange multiplier:

$$V_\lambda =: wy(t) \tag{203}$$

Now, let us define the Hamiltonian in the second stage, introducing a new Langrange multiplier χ with the secondary constraint Ξ:

$$\begin{aligned}
H^{(2)} &= H_0 + V_x\zeta + V_y\Lambda + V_\lambda\Delta + \chi\Xi \\
H_0 &= -\frac{1}{2\alpha}P_Q^2 - \frac{1}{2}Q^2 + \frac{w}{2}y^2 - \frac{m}{2}x^2 + P_yQ \\
\zeta &=: P_x - \frac{g}{\alpha}P_Q \approx 0 \\
\Lambda &=: P_y + \lambda \approx 0 \\
\Delta &=: P_\lambda \approx 0 \\
\Xi &=: mx(t) + \frac{g}{\alpha}(P_y(t) - Q(t)) \approx 0
\end{aligned} \tag{204}$$

The Poisson brackets between constraints, are:

$$\begin{array}{lll}
\{\zeta, \lambda\} = 0 & \{\zeta, \Delta\} = 0 & \{\Lambda, \Delta\} = 1 \\
\{\Lambda, \Xi\} = 0 & \{\Xi, \Delta\} = 0 & \{\Xi, \zeta\} = m + \dfrac{g^2}{\alpha^2}
\end{array} \tag{205}$$

And now, impose the same conservation for the Ξ constraint in time:

$$\begin{aligned}
\dot{\Xi}(t) &= \{\Xi(t), \mathcal{H}^{(2)}(t)\} \stackrel{!}{=} 0 \\
\dot{\Xi}(t) &= -\frac{g}{\alpha} w y(t) + \frac{g}{\alpha^2} P_Q(t) + V_x \left(m + \frac{g^2}{\alpha^2} - \frac{g}{\alpha} w P_\lambda(t) \right) \stackrel{!}{=} 0 \\
&\Rightarrow \text{ Fix the Lagrange multiplier:}
\end{aligned}$$

$$V_x =: \frac{g}{\alpha} \frac{1}{m + \frac{g^2}{\alpha^2}} \left(w(y(t) + P_\lambda(t)) - \frac{P_Q}{\alpha} \right) \tag{206}$$

Because no other constraint arises, we can assure that by fixing the Lagrange multipliers as in (202, 203, 206), all the constraints are preserved in time. In fact, all the constraint content of the theory has already been given in (204). Furthermore, from (205) we can assure that the system of constraints is of second class and consequently, the inverse of the Dirac matrix, as well as the Dirac brackets for the complete theory, exist. Finally, we can identify the Hamiltonian (204), together with the evolution of the degrees of freedom given by the Poisson brackets with the second stage Hamiltonian, with the complete Hamiltonian formulation for fields (123).

Furthermore, we can identify from the complete Hamiltonian formulation (204), that there are 8 canonical variables and 4 constraints among them. Therefore, we now desire to build the physical Hamiltonian, depending on two disjoint sets of new canonical variables $\Upsilon_a \in \{\omega_g, \Omega_f\}$ with $g = 1 \ldots 4$ $f = 1 \ldots 4$. The set $\{\omega\}$ being the dynamical variables and $\{\Omega\}$ the set of non-dynamical variables, that do not enter the physical Hamiltonian and vanish identically when the equations of motion are satisfied (For a much broader discussion, see the related section on fields (4.3.3)).

Based on the previous discussion, with the constraints in (204) and their Poisson brackets (205), we write the Dirac matrix as:

$$\mathcal{D} = \begin{pmatrix} \{\Xi,\Xi\} & \{\Xi,\zeta\} & \{\Xi,\Lambda\} & \{\Xi,\Delta\} \\ \{\zeta,\Xi\} & \{\zeta,\zeta\} & \{\zeta,\Lambda\} & \{\zeta,\Delta\} \\ \{\Lambda,\Xi\} & \{\Lambda,\zeta\} & \{\Lambda,\Lambda\} & \{\Lambda,\Delta\} \\ \{\Delta,\Xi\} & \{\Delta,\zeta\} & \{\Delta,\Lambda\} & \{\Delta,\Delta\} \end{pmatrix} \tag{207}$$

Using the condition $g^2 \stackrel{!}{=} \alpha\beta$,

$$\mathcal{D} = \begin{pmatrix} 0 & m + \frac{\beta}{\alpha} & 0 & 0 \\ -(m + \frac{\beta}{\alpha}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{208}$$

And the inverse of the Dirac matrix:

$$\mathcal{D}^{-1} = \begin{pmatrix} 0 & -\dfrac{1}{m+\dfrac{\beta}{\alpha}} & 0 & 0 \\ \dfrac{1}{m+\dfrac{\beta}{\alpha}} & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{209}$$

As introduced for fields, the Dirac brackets in a classical model with finite degrees of freedom, are, for any two functions F, G depending on time:

$$\{F(t), G(t)\}_{D_\Pi} =: \{F(t), G(t)\} - \sum_{a,a'} \{F(t), \Pi_a(t)\}\{\Pi(t), \Pi(t)\}^{-1}_{a,a'}\{\Pi_{a'}(t), G(t)\} \tag{210}$$

where, following the notation introduced in section (3.4) and also used in (4.3), we have $\Pi_a \in \{\Xi, \zeta, \Lambda, \Delta\}$. Explicitly, using the constraints given in (204), the Dirac brackets are:

$$\begin{aligned} \{F(t), G(t)\}_{D_\Pi} =: \{F(t), G(t)\} \quad &+ \quad \frac{\alpha}{m\alpha+\beta}\{F(t), mx(t) + \frac{g}{\alpha}(P_y(t) - Q(t))\}\{P_x(t) - \frac{g}{\alpha}P_Q(t), G(t)\} \\ &- \quad \frac{\alpha}{m\alpha+\beta}\{F(t), P_x(t) - \frac{g}{\alpha}P_Q(t)\}\{mx(t) + \frac{g}{\alpha}(P_y(t) - Q(t)), G(t)\} \\ &+ \quad \{F(t), P_y(t) + \lambda(t)\}\{P_\lambda(t), G(t)\} \\ &+ \quad \{F(t), P_\lambda(t)\}\{P_y(t) + \lambda(t), G(t)\} \end{aligned} \tag{211}$$

With these Dirac brackets, we compute the Dirac brackets between all the degrees of freedom

$\Phi \in \{x, y, Q, \lambda, P_x, P_y, P_Q, P_\lambda\}$:

$$
\begin{array}{ccc}
\{Q(t),Q(t)\}_{D(\Pi)} = 0 & \{y(t),P_y(t)\}_{D(\Pi)} = 1 & \{\lambda(t),P_\lambda(t)\}_{D(\Pi)} = 0 \\
\{Q(t),y(t)\}_{D(\Pi)} = -\dfrac{\beta}{\alpha}\dfrac{1}{m+\dfrac{\beta}{\alpha}} & \{y(t),P_x(t)\}_{D(\Pi)} = 0 & \{\lambda(t),P_x(t)\}_{D(\Pi)} = 0 \\
\{Q(t),x(t)\}_{D(\Pi)} = 0 & \{y(t),P_\lambda(t)\}_{D(\Pi)} = 0 & \{P_Q(t),P_Q(t)\}_{D(\Pi)} = 0 \\
\{Q(t),\lambda(t)\}_{D(\Pi)} = 0 & \{x(t),x(t)\}_{D(\Pi)} = 0 & \{P_Q(t),P_x(t)\}_{D(\Pi)} = 0 \\
\{Q(t),P_Q(t)\}_{D(\Pi)} = \dfrac{m}{m+\dfrac{\beta}{\alpha}} & \{x(t),\lambda(t)\}_{D(\Pi)} = 0 & \{P_Q(t),P_y(t)\}_{D(\Pi)} = 0 \\
\{Q(t),P_y(t)\}_{D(\Pi)} = 0 & \{x(t),P_Q(t)\}_{D(\Pi)} = \dfrac{g}{\alpha}\dfrac{1}{m+\dfrac{\beta}{\alpha}} & \{P_Q(t),P_\lambda(t)\}_{D(\Pi)} = 0 \\
\{Q(t),P_x(t)\}_{D(\Pi)} = \dfrac{g}{\alpha}\dfrac{m}{m+\dfrac{\beta}{\alpha}} & \{x(t),P_y(t)\}_{D(\Pi)} = 0 & \{P_y(t),P_y(t)\}_{D(\Pi)} = 0 \\
\{Q(t),P_\lambda(t)\}_{D(\Pi)} = 0 & \{x(t),P_x(t)\}_{D(\Pi)} = 1 - \dfrac{m}{m+\dfrac{\beta}{\alpha}} & \{P_y(t),P_x(t)\}_{D(\Pi)} = 0 \\
\{y(t),y(t)\}_{D(\Pi)} = 0 & \{x(t),P_\lambda(t)\}_{D(\Pi)} = 0 & \{P_y(t),P_\lambda(t)\}_{D(\Pi)} = 0 \\
\{y(t),x(t)\}_{D(\Pi)} = -\dfrac{g}{\alpha}\dfrac{1}{m+\dfrac{\beta}{\alpha}} & \{\lambda(t),\lambda(t)\}_{D(\Pi)} = 0 & \{P_x(t),P_x(t)\}_{D(\Pi)} = 0 \\
\{y(t),\lambda(t)\}_{D(\Pi)} = -1 & \{\lambda(t),P_Q(t)\}_{D(\Pi)} = 0 & \{P_x(t),P_\lambda(t)\}_{D(\Pi)} = 0 \\
\{y(t),P_Q(t)\}_{D(\Pi)} = 0 & \{\lambda(t),P_y(t)\}_{D(\Pi)} = 0 & \{P_\lambda(t),P_\lambda(t)\}_{D(\Pi)} = 0
\end{array}
\tag{212}
$$

With these brackets, the "square matrix of Dirac brackets" can be computed as with the field theory in (4.3), (4.3.3), and it can be brought, my means of the proposed procedure in the respective section on Dirac brackets, to the already discussed form (214). Nevertheless, note that here, the "extended" matrix of Dirac brackets do not need to be used, because the constraints do not involve derivatives. This is just a consequence of the fact that this is a classical model without spatial derivatives, which were forced on us in the field theory, because of covariance.

The "square matrix of Dirac brackets" between variables is straightforward to write down from the equations (212), because there are no indices to care about. Furthermore, it is similar to the one written down for fields (192)and after finding the new set of variables Υ, takes the form:

$$J = \begin{pmatrix} 0_n & \tilde{I} \\ -\tilde{I} & 0_n \end{pmatrix} \tag{213}$$

where this matrix makes sense in the symplectic form of the Hamilton equations of motion:

$$\dot{\boldsymbol{\Upsilon}} = J\frac{\partial \tilde{H}}{\partial \boldsymbol{\Upsilon}} \tag{214}$$

where $\boldsymbol{\Upsilon}$ is a vector of all the new variables in the sets $\{\omega_g\}_{g=1,\ldots,4}$, $\{\Omega_f\}_{f=1,\ldots,4}$, 0_n is a 4×4 zero matrix and $\tilde{I}$ is a diagonal 4×4 matrix, with entries 0 or 1 in the diagonal. The reason for such a form of the $\tilde{I}$ matrix is, as explained for fields, that in the new variables, the set $\{\Omega_f\}_f$ is demanded to be identically 0. Thus, only the equations corresponding to dynamical degrees of freedom ($\{\omega_g\}_g$) have a "1" entry in the diagonal of $\tilde{I}$.

Then, the change between the set $\{\Phi\}$ to $\{\Upsilon\}$ is:

$$\Phi =: \begin{pmatrix} Q \\ y \\ x \\ \lambda \\ P_Q \\ P_y \\ P_x \\ P_\lambda \end{pmatrix} \rightarrow \begin{pmatrix} (Q - P_y - \frac{\alpha}{g} m x) \\ y - P_Q \\ x \\ \lambda + P_y \\ \frac{m\alpha + \beta}{g} P_Q \\ P_y \\ P_x - \frac{g}{\alpha} P_Q \\ P_\lambda \end{pmatrix} =: \begin{pmatrix} \Xi' \\ y' \\ x' \\ \Lambda \\ P_{x'} \\ P_{y'} \\ \zeta \\ \Delta \end{pmatrix} := \Upsilon \tag{215}$$

References

[1] F. Sbisà, Classical and Quantum ghosts. Eur. J. Phys. 36 015009 2015. DOI: 10.1088/0143-0807/36/1/015009. arXiv:1406.4550v4 [hep-th]

[2] R. P. Woodard, The theorem of Ostrogradsky, 2015. arXiv:1506.02210 [hep-th]

[3] R. P.Woodard, Avoiding Dark Energy with 1/R Modifications of Gravity, 2006. arXiv:astro-ph/0601672 , 30.

[4] E. Babichev, C. Charmousis, G. Esposito-Farése and A. Lehébel, Hamiltonian vs stability and application to Horndeski theory. JCAP, 2018. arXiv:1803.11444v1 [gr-qc]

[5] Y. Abe, T. Inami, K. Izumi, T. Kitamura and T. Noumi, S-matrix Unitarity and Renormalizability in Higher Derivative Theories. arXiv:1805.00262v1 [hep-th]

[6] C. de Rham and A. Matas. Ostrogradsky in theories with multiple fields. JCAP, 2016. DOI: 10.1088/1475-7516/2016/06/041. arXiv:1604.08638v2 [hep-th]

[7] J. López Sarrión and C. Reyes. Higher time derivatives, stability and Fermi statistics, 2011. arXiv:1105.3243v1 [hep-th]

[8] H. Motohashi and T. Suyama, Third order equations of motion and the Ostrogradsky instability. Phys. Rev. D91, 2015. DOI: 10.1103/PhysRevD.91.085009. arXiv: 1411.3721

[9] C. Bender and P. Mannheim, No Ghost theorem for the fourth order derivative Pais-Uhlenbeck oscillator model, 2007. DOI: 10.1103/PhysRevLett.100.110402. arXiv:0706.0207v1 [hep-th]

[10] M. Pavšič, Stable self-interacting Pais-Uhlenbeck Oscillator. Modern Phys. Lett. A 28, 36, 2013. DOI: 10.1142/S0217732313501654

[11] H. Dimov, S. Mladenov, R. Rashkov and T. Vetsov, Entanglement of higher-derivative oscillators in holographic systems, 2016. DOI: 10.1016/j.nuclphysb.2017.03.005. arXiv:1607.07807v3 [hep-th]

[12] G. W. Horndeski, Second-order scalar-tensor field equations in a four-dimensional space. Int.J.Theor.Phys. 10 (1974) 363-384. DOI: 10.1007/BF01807638

[13] M .S. Plyushchay, Massless point particle with rigidity. Modern Phys. Lett. A, 4 (9), 1989.

[14] M. Ostrogradsky, Mem. Ac. St. Petersbourg VI 4, 1850, 385.

[15] A. Pais and G. E. Uhlenbeck, On Field Theories with Non-Localized Action. Phys. Rev. **79**, (1) (1950). DOI: https://doi.org/10.1103/PhysRev.79.145

[16] S. W. Hawking and T. Hertog, Living with ghosts. Phys. Rev. D, vol. 65, 2002. DOI: 10.1103/PhysRevD.65.103515. arXiv:hep-th/0107088v2

[17] D. Langlois and K. Noui, Hamiltonian analysis of higher derivative scalar-tensor theories. JCAP 07 (2016) 016, 2016. DOI: 10.1088/1475-7516/2016/07/016. arXiv:1512.06820v1 [gr-qc]

[18] D. Langlois, M. Mancarella, K. Noui, F. Vernizzi. Effective Description of Higher-Order Scalar-Tensor Theories. JCAP 05 (2017) 033, 2017. DOI: 10.1088/1475-7516/2017/05/033. arXiv:1703.03797v2 [hep-th]

[19] M. Crisostomi, R. Klein, D. Roest. Higher Derivative Field Theories: Degeneracy Conditions and Classes. J. High Energ. Phys. (2017) 2017: 124. DOI: 10.1007/JHEP06(2017)124. arXiv:1703.01623v2 [hep-th]

[20] R. Kimura, A. Naruko and D. Yoshida, Extended vector-tensor theories. JCAP 01 (2017) 002, 2017. DOI: 10.1088/1475-7516/2017/01/002. arXiv:1608.07066v4 [gr-qc].

[21] D. Becker, C. Ripken, F. Saueressig, On avoiding Ostrogradski instabilities within Asymptotic Safety. J. High Energ. Phys. (2017) 2017: 121. https://doi.org/10.1007/JHEP12(2017)121. arXiv:1709.09098v2 [hep-th]

[22] K. Andrzejewski, J. Gonera, P. Maslanka. Euclidean Path Integral and Higher-Derivative Theories. Progress of Theoretical Physics, 125, 2, 2011. https://doi.org/10.1143/PTP.125.247. arXiv:0904.3055v3 [hep-th]

[23] L. Casarin, On higher-derivative gauge theories. (2017) arXiv:1710.08021v2 [hep-th]

[24] B. Grinstein, D. O'Connell and M. B. Wise. The Lee-Wick Standard Model. Phys.Rev. D77, 025012, 2008. DOI: 10.1103/PhysRevD.77.025012. arXiv:0704.1845v3 [hep-ph].

[25] C. Carone and R. Lebed. A Higher derivative Lee-Wick Standard Model. JHEP, 2009. DOI: 10.1088/1126-6708/2009/01/043. arXiv:0811.4150v2 [hep-ph].

[26] K. S. Stelle, Renormalization of Higher Derivative Quantum Gravity. Phys. Rev. D. 1977. DOI: 10.1103/PhysRevD.16.953

[27] T. Ortín, Higher order gravities and the Strong Equivalence Principle, J. High Energ. Phys. (2017) 2017: 152. DOI: 10.1007/JHEP09(2017)152. arXiv:1705.03495v2 [gr-qc]

[28] K. Muneyuki and N. Ohta, Renormalization of Higher Derivative Quantum Gravity Coupled to a Scalar with Shift Symmetry. (2013) DOI: 10.1016/j.physletb.2013.07.054. arXiv:1306.6701v3 [hep-th]

[29] A. Nicolis, R. Ratazzii and E. Trincherini, The galileon as a local modification of gravity. Phys. Rev. D, 79 2009. DOI: 10.1103/PhysRevD.79.064036. arXiv:0811.2197v2 [hep-th]

[30] J. M. Ezquiaga, J. García-Bellido and M. Zumalacárregui, Field redefinitions in theories beyond Einstein gravity using the language of differential forms. Phys. Rev. D95, 084029 (2017), 1608.01982. arXiv:1701.05476v1 [hep-th].

[31] S. A. Larin, Higher-derivative relativistic quantum gravity. Mod. Phys. Lett. A, 33, 1850028 (2018). DOI: 10.1142/S0217732318500281. arXiv:1711.02975v1 [physics.gen-ph]

[32] E. Álvarez and S. González-Martín, Ghost-free higher derivative unimodular gravity. Phys. Rev. D 96, 124031 (2017) DOI:10.1103/PhysRevD.96.124031. arXiv:1708.03133v3 [hep-th]

[33] L. Modesto, Super-renormalizable Higher-Derivative Quantum Gravity, (2012). arXiv:1202.0008v1 [hep-th]

[34] R. Kase, R. Kimura, A. Naruko and D. Yoshida, Stable cosmological solutions in degenerate theory of gravity. 2018. arXiv:1803.10876v1 [gr-qc]

[35] M. Crisostomi and K. Koyama, Self-accelerating universe in scalar-tensor theories after GW170817. Phys. Rev. D 97, 084004, 2018. DOI: 10.1103/PhysRevD.97.084004. arXiv:1712.06556v2 [astro-ph.CO]

[36] M. Crisostomi and K. Koyama, Vainshtein mechanism after GW170817. Phys. Rev. D 97, 021301(R), 2017. DOI: 10.1103/PhysRevD.97.021301. arXiv:1711.06661v2 [astro-ph.CO]

[37] G. Pulgar, J. Saavedra, G. Leon and Y. Leyva, Higher Order Lagrangians inspired by the Pais-Uhlenbeck Oscillator and their cosmological applications. JCAP, 2015 DOI: 10.1088/1475-7516/2015/05/046. arXiv:1408.5885v3 [gr-qc]

[38] D. Yoshida, J. Quintin, M. Yamaguchi and R. Brandenberger, Cosmological perturbations and stability of nonsingular cosmologies with limiting curvature. Phys.Rev. D96 (2017) no.4, 043502, 2017. DOI: 10.1103/PhysRevD.96.043502. arXiv:1704.04184v2 [hep-th]

[39] H. Motohashi, K. Noui, T. Suyama, M. Yamaguchi and D. Langlois, Healthy degenerate theories with higher derivatives. JCAP 1607 (2016) 07, 033, 2016. DOI: 10.1088/1475-7516/2016/07/033. arXiv:1603.09355v2 [hep-th].

[40] R. Klein and D. Roest, Exorcising the Ostrogradsky ghost in coupled systems. J. High Energ. Phys. (2016) 2016: 130. DOI: 10.1007/JHEP07(2016)130. arXiv:1604.01719v3 [hep-th].

[41] Tai-Jun Chen, M. Fasiello, E. Lim and A. Tolley, Higher derivative theories with constraints: exorcising Ostrogradsky's ghost. JCAP 1302 (2013) 042, 2013. DOI: 10.1088/1475-7516/2013/02/042. arXiv:1209.0583v4 [hep-th].

[42] F.J. de Urries and J. Julve. Ostrogradski formalism for Higher-derivative scalar field theories, 1998. DOI: 10.1088/0305-4470/31/33/006. arXiv:hep-th/9802115v2.

[43] H. Goldstein, C. Poole and J. Safko, Classical Mechanics, 3rd edition. Pearson, 2001.

[44] P. A. M. Dirac, Lectures on Quantum Mechanics. Dover Publications, 1964.

[45] D.M. Gitman and I.V. Tyutin, Quantization of Fields with Constraints. Springer - verlag, 1990.

[46] M. Henneaux and C. Teitelboim, Quantization of Gauge Systems. Princeton University Press, 1992.

[47] A. Hanson, T. Regge and C. Teitelboim, Constrained Hamiltonian Systems. Accademia Nazionale dei Lincei, 1976.

[48] W. Greiner and J. Reinhardt, Field Quantization. J. Springer, Berlín, 1996.

[49] V. P. Nair, Quantum Field Theory. A modern perspective. Springer, New York. 2004.

[50] M. E. Peskin and D. Schroeder, An Introduction to Quantum Field Theory. Perseus Books, Massachusetts, 1995.

[51] M. Maggiore, A Modern Introduction to Quantum Field Theory. Oxford University Press, 2005.

[52] M. Srednicki, Quantum Field Theory. Cambridge University Press, 2007.

[53] K. Schleich, Is reduced phase space quantisation equivalent to Dirac quantisation?. Class. Quatum Grav. 7, 1990. DOI: 10.1088/0264-9381/7/8/028.