Exploration of fermion zero modes and spherically symmetric gauge fields

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

MSc. Lu YANG
aus Baoji, China

Tag der mündlichen Prüfung: 03.07.2015

Referent: Prof. Dr. Frans R. Klinkhamer (Karlsruhe)
Korreferent: Prof. Dr. Janos Polonyi (Strasbourg)

DOI 10.5445/IR/1000048738

Abstract

This thesis consists of an introduction, four chapters, a discussion and an appendix.

The introduction will provide the background to the work covered in this thesis, as well as an outline of the structure of the thesis.

Chapter 1 presents a study related to fermion zero modes. The aim is to prove the existence of a zero mode that Klinkhamer and Lee observed in their study of a fermion doublet coupled to a chiral $SU(2)$ gauge field. The proof comprises analytical and numerical analysis on the stability of the solutions obtained in the study of the Dirac equation of the fermion.

Chapter 2 sketches a new mechanism for deriving a discrete and bounded fermion mass spectrum, based on the work of Klinkhamer and the present author. The model theory used consists of two fermion fields interacting with a Higgs-like scalar field. An open extra dimension is introduced to this theory so that a set of explicit classical solutions to the equations of motion is obtained. When the wave functions are required to be normalizable in the extra dimension, the masses of the four-dimensional fermions naturally become bounded and discrete.

Chapter 3 consists of an investigation of a theory on the gauged Lorentz group. In Minkowskian space-time, the pure Yang-Mills theory of this group, with spherical symmetry imposed on the gauge field, reduces to a new theory in a two-dimensional space-time. The reduced theory has a scalar field with four degrees of freedom and a quartic potential, and two abelian gauge fields. A problem that remains to be solved, however, is that the potential of the scalar field is not bounded from below.

In Chapter 4, a method for deriving a set of identities of the correlation functions in quantum field theories is presented. It can be used to obtain a variational equation (resembling a differential equation) for the generating functional of a given theory. When the generating functional is expanded into a Taylor series in terms of the source field(s), the variational equation makes it possible to relate the correlation functions of different processes to one another. The calculation here is non-perturbative. One of the identities derived for the $\lambda - \phi^4$ theory is tested and verified.

The thesis ends with a discussion in order to address the questions left unanswered in the previous chapters. Some reflections and ideas are given here, in hope that they can stimulate interest in future research.

Zusammenfassung

Diese Arbeit umfasst eine Einleitung, vier Kapitel, ein Fazit, sowie einen Anhang.

Die Einleitung beinhaltet eine Veranschaulichung der Grundlagen und eine Zusammenfassung der in dieser Arbeit präsentierten Ausarbeitungen.

Kapitel 1 stellt Untersuchungen zu fermionischen Nullstellen vor. Das Ziel ist es, die Existenz der von Klinkhamer und Lee in ihren Untersuchungen einer Theorie mit einem Fermiondublett, gekoppelt an ein chirales $SU(2)$ Eichfeld gefundenen Nullstellen zu beweisen. Der Beweis umfasst analytische und numerische Untersuchungen der Stabilität der Lösungen der Diracgleichung der Theorie.

Kapitel 2 skizziert einen neuen Mechanismus zur Herleitung eines diskreten und beschränkten Fermionmassenspektrums, basierend auf der Arbeit von Klinkhamer und meiner selbst. Das verwendete Modell besteht aus zwei fermionischen Feldern und einem mit ihnen wechselwirkenden, Higgs-ähnlichen skalaren Feld. Nach der Einführung einer zusätzlichen, offenen Dimension, können explizite Lösungen der Bewegungsgleichungen hergeleitet werden. Durch die Bedingung der Normalisierbarkeit der Wellenfunktionen in der zusätzlichen Dimension, werden die Massen der vierdimensionalen Fermionen auf ganz natürliche Weise beschränkt und diskretisiert.

Kapitel 3 beinhaltet Untersuchungen einer Theorie mit geeichter Lorentzgruppe. In der Minkowskischen Raumzeit, lässt sich die reine Yang-Mills-Theorie dieser Gruppe, unter Beschränkung auf sphärisch symmetrische Felder, auf eine Theorie in zwei Raumzeitdimensionen reduzieren. Die hieraus gewonnene Theorie ist die eines skalaren Feldes mit vier Freiheitsgraden und einem quartischen Potential, sowie zwei abelschen Eichfeldern. Das Potential ist jedoch nicht von unten beschränkt, welches ein noch zu lösendes Problem darstellt.

In Kapitel 4 wird eine Methode zur Herleitung diverser Korrelationsfunktionsidentitäten in Quantenfeldtheorien präsentiert. Diese ermöglicht es uns für die erzeugende Funktion einer beliebigen Theorie eine Variationsgleichung (in Form einer Differentialgleichung) herzuleiten. Wenn die erzeugende Funktion in den Quellenfeldern Taylorentwickelt wird, ermöglicht es die Variationsgleichung Korrelationsfunktionen unterschiedlicher Prozesse in Bezug zu setzen. Die vorliegende Rechnung ist nicht perturbativ. Eine Identität der $\lambda - \phi^4$ Theorie wurde explizit verifiziert.

Im Fazit werden die Probleme angesprochen, welche in den vorigen Kapiteln offen gelassen wurden. Es werden zudem einige Gedanken und Ideen vorgestellt, mit der Hoffnung, Interesse an zukünftiger Forschung auf diesem Gebiet anzuregen.

Contents

Introduction

The main body of this thesis is divided into four chapters, with each chapter covering a different topic. These topics describe a sequence of problems that I have worked on in the past three years. They also reflect the course I have followed exploring the field of topology of gauge fields.

First, let me introduce the background of the research on these topics.

When I started the PhD program, the problem I was given was to determine the spectral flow in the study of fermion number violation. The interest in this problem traces back to anomalies, one of the most peculiar features of quantum field theory. Consider a special quantum field theory of electrodynamics, where the electron field is massless. This theory possesses two types of $U(1)$ gauge symmetries: one vector gauge symmetry, and one chiral (also called axial) gauge symmetry, which means that the left and right handed fermions transform differently. Classically, it is straightforward to verify that the Nöther currents associated to both symmetries are conserved. However, upon quantization it turns out that the two currents cannot be simultaneously conserved, when quantum effects are included. This phenomenon was so puzzling that it deserved a striking name: anomaly. Three physicists in particular dedicated a great deal of research to this subject, so that it became known as the "Adler-Bell-Jackiw anomaly", after them. Since then, different methods have been invented to evaluate the quantum divergence of the axial current, i.e. to quantify to what degree the conservation of the axial current is violated. They all, of course, led to the same result: the divergence is a topological term totally depending on the gauge field. It is worth to note the three most prominent of these methods, as they provide different perspectives to understanding the nature of the anomaly: the perturbative calculation of the triangle diagram, the point separation method in operator calculations, and the path integral method [1].

The concept of anomalies found its first application in clarifying the confusion from the Sutherland-Veltman paradox, which arose in the context of the decay process of a pion into two photons. A clear discussion of this issue can be found in [2]. Later, physicists realized that anomalies typically imply a disturbing phenomenon: fermion number non-conservation. 't Hooft was the first to carry out the difficult calculation and showed that the amplitude of a fermion number violating process is indeed not zero [3]. In fact, the amplitude is typically very small and cannot be calculated perturbatively. 't Hooft's calculation was done in Euclidean space-time by making use of the instanton solution to the gauge field equations. The process involved is called vacuum tunneling. The mysterious expression, vacuum tunneling, indicates that there can be many vacua (although they all have zero energy, they are nevertheless different). What then tells the different vacua apart is a topological quantity of the non-abelian gauge fields associated to the vacua. A common illustration of this concept makes use of a pure Yang-Mills theory with the gauge group $SU(2)$ in the four-dimensional Minkowskian space-time $R^{(1,3)}$. The fact that the vacuum has zero energy means that the gauge field strength must vanish everywhere in the three-dimensional space. This implies that the gauge field is in pure gauge, i.e. it takes the form $ig^{-1}(x)\partial g(x)$ for some gauge function $g : R^3 \to SU(2)$. At infinity in the space R^3, the map is topologically equivalent to one from the two-sphere, S^2, to the gauge group. Topologically the gauge group $SU(2)$ is the three-dimensional unit hypersphere S^3. Therefore, the map g restricted at infinity is one from S^2 to S^3. According to a known result of topology ($\pi_2(S^3) \simeq 0$) this restricted map can be continuously deformed to the identity map, thus extended to the whole space R^3 without inducing any singularity. This implies, as far as only the topological properties of the map $g(x)$ are concerned, that we can identify all the points at infinity in R^3 as one single point. With this identification R^3 becomes S^3. Mathematically this procedure is called one-point compactification. Now the gauge function can be viewed as a map from one S^3 to another S^3. This map possesses a nontrivial topological property, called the winding number, which is an integer number. This property is shared by the gauge field and the corresponding vacuum. Vacuum tunneling, then, can be described as the transition of a vacuum with a certain winding number number into another vacuum with a different winding number, at a tiny probability called tunneling rate. For the transition from one vacuum to another the physical system has to overcome a positive energy barrier (in principle it is the energy of the sphaleron [4]) which obviously exceeds the available energy possessed by the vacuum, thus the name tunneling. In the anomaly equation of this illustration, the divergence of the axial current is proportional to the Pontrjagin density, which, upon integration over the four space-time, gives exactly the change in winding number. From the anomaly equation we conclude that a charge corresponding to the spatial integration of the axial current, the fermion number, must be violated. The violation of fermion number is given precisely by the change of winding number. In other words, when the vacuum tunneling occurs, (anti) fermions are created, equal to the change in the winding number between vacua in number.

It is natural to guess that anomalies also occur in the standard model theory. This indeed happens. The story is, however, a bit more complicated. In the standard model theory there are several types of fermionic fields, which are grouped as leptonic and baryonic fields. Both leptonic and baryonic fields have anomaly equations for their axial currents. Miraculously, when we take the difference of leptonic and baryonic axial currents, the net divergence vanishes. This phenomenon is called anomaly cancellation. It occurs in the standard model theory because of the special structure of the theory: the number of the leptonic fields, that of the baryonic fields, and the quantum charges associated with these fields come in such a way, that the cancellation is possible. The anomaly cancellation implies that the charge, which is the difference of the baryon number and the lepton number, is conserved. It is, however, still possible to get an anomaly equation with non-vanishing divergence, by taking the sum of the two equations. This manipulation has a physical implication: it suggests that the sum of the baryon and lepton numbers is not conserved. Put in a nutshell, in the standard model theory, the anomaly causes the violation of the lepton and baryon numbers, while the difference of the two is still conserved due to the anomaly cancellation.

As stated earlier, 't Hooft's calculation on the vacuum tunneling rate was done in Euclidean space-time. The method was based on an extrapolation of the WKB approximation in quantum mechanics: the tunneling rate can be approximated by using the classical solution to the equation of motion in imaginary time. This idea is physically plausible, however, not totally convincing for the following reason. Many topological results on the gauge fields are obtained in Euclidean space-time. Their counter-part in Minkowskian space-time is, however, either not known or very different. One example of this is that, the winding number of gauge fields takes integer number values in Euclidean space-time, while in Minkowskian space-time it can be an arbitrary fractional number [5]. Another example is the index theorem. In Euclidean space-time without boundaries, the degree of fermion number violation is quantified by the index of the Dirac operator in question, according to the Atiyah-Singer index theorem. In Minkowskian space-time the Dirac operator is no longer elliptic. As a result, its index cannot be defined in this case. These facts warn us not to rely on the calculation performed in Euclidean space-time. Instead, we should search diligently for a direct calculation in Minkowskian space-time.

To quantify the fermion number violation, a more general term, has been introduced:

spectral flow. Suppose, on a time interval we can track all the fermionic states in the theory under question, i.e. the eigenmodes of the Dirac operator. We know how their eigenvalues change over time. It might be that the eigenvalue of a mode turns from negative to positive, or the other way around. Now imagine that the state corresponding to this mode is occupied by an anti-fermion (or fermion). Such a scenario is then physically interpreted as an anti-fermion becoming a fermion or vice versa, accompanied by a change of 2 (or -2) in fermion number. The net change in fermion number, taking into account all the fermion modes, is called spectral flow over the given time interval. Similar to the index of the Dirac operator, the spectral flow is also a topological quantity associated to the Dirac operator and the gauge fields, in Euclidean space-time with boundaries (the spaces at the initial and the final time make up the boundaries here). A formula for computing the spectral flow in a compact Euclidean space-time with boundaries is provided in the Atiyah-Patodi-Singer index theorem [6].

Although Dirac operators in Minkowskian space-time are not elliptic, their spectral flow on a time interval can be defined unambiguously, given that the space in question is compact. Unfortunately, in Minkowskian space-time there is no elegant result for the spectral flow, comparable to the Atiyah-Patodi-Singer index theorem. Instead, a different approach can be taken. Since the fermion number violation occurs when the eigenvalue of a fermionic state crosses zero, necessarily when a fermion zero mode occurs, we can focus on studying the zero modes. We first find all the fermion zero modes of the theory in a specific time interval, then study the direction of the level crossing at the vicinity of each zero mode. Summing up all the information on the level crossings gives the spectral flow. While this approach seems less efficient than the elegant topological approach, it is practical and provides a clear physical account of the fermion number violating phenomena.

The previous text describes the background of the research I conducted over the last three years. With this background knowledge we can appreciate better the interest in the following topical discussion and agree that following these directions was indeed a natural choice.

Chapter 1 is the most directly related to the background I have discussed. Klinkhamer and Lee studied the fermion zero modes in the chiral theory containing a gauge group $SU(2)$ and a fermion doublet [7]. All the fields of their model are constrained to be spherically symmetric in three dimensional space. It is worthwhile to add some comments here on the spherical symmetry condition. Spherical symmetry was originally imposed on gauge fields for the purpose of simplifying the Yang-Mills field equations, so that some exact solutions could be obtained. However, Witten observed that a pure Yang-Mills theory for the $SU(2)$ gauge group in a flat space-time with dimension four, upon the imposition of spherical symmetry, reduces to an abelian Higgs theory in a two dimensional curved space-time [8]. The phase of the Higgs-like complex scalar field carries the topological information of the original $SU(2)$ gauge field, i.e. its winding number. Klinkhamer and Lee showed that the change of this winding number is accompanied with the occurrence of fermion zero modes. In addition, they observed from numerical analysis that, there might be a fermion zero mode which is not associated to a change of winding number. In the following I will address this special fermion zero mode as a "non-topological" one. Apparently, all these fermion zero modes are crucial for determining the spectral flow. The work presented in the first chapter is a continuation of Klinkhamer and Lee's work. It consists mainly of a proof to the existence of a "non-topological" fermion zero mode in the theory studied by Klinkhamer and Lee. The proof consists of both analytic and numeric studies. The analytic result obtained here can help to find more fermion zero modes in theories where the equations of motion assume a similar form.

The fermion zero modes themselves are of great interest. Their appearance is closely related to the topological property of the background field coupled to the fermion field. The connection between fermion zero modes and the topology of the background field is best illustrated in a simple model of a fermion coupled to a scalar field in two dimensional space-time. Jackiw and Rebbi showed that the kink solution of the scalar field in this model is necessarily accompanied with the occurrence of a fermion zero mode [10]. From the straightforward calculation we can gain some insight about the topological fermion zero mode. A peculiar property of the fermion zero mode discovered by Jackiw and Rebbi is, that in the presence of one fermion zero mode the background field will necessarily carry $1/2$ or $-1/2$ fermion number. A deeper understanding of this observation involves the discussion of the quantization of a field theory, which goes beyond the scope of this thesis.

It's natural to consider the model studied by Jackiw and Rebbi in a higher dimensional space-time and seek new solutions carrying topological properties. However, there is no intrinsically new fermion zero mode in higher dimensions, unless gauge fields are introduced to the theory. At least, no nontrivial static solution to the equations of motion for a Higgs-like scalar field in two or three dimensional space (no gauge fields) has been found. However, we noticed an interesting byproduct of this investigation. We can find a bounded and discrete mass spectrum for fermions living in Minkowskian space-time, by making use of the kink solution in an extra dimension. It has been proposed that fermions living in the ordinary 4-d space-time are zero modes trapped in a brane in a 5-d space-time [11]. This idea inspired the speculation on a possible mechanism to determine the spectrum of fermion masses in four dimensional space-time. In the second chapter I present a working theory for such a mechanism. The simple model of a fermion coupled to a scalar field with a double well potential is modified slightly. Then it is possible to find a set of exact solutions to the classical equations of motion. To obtain these solutions, two Ansätze for the fermion solutions have been made. It will be seen that these Ansätze have a fitting interpretation. Moreover, a surprising feature of the fermion solutions was noticed. They make it possible to distinguish left- and right- handed components of a Dirac spinor in the five dimensional space-time, while the counter parts of the two spinor components in the four dimensional space-time appear symmetric.

The third chapter presents a small investigation of the topic of spherically symmetric gauge fields. It was mentioned previously, that the spherically symmetric $SU(2)$ Yang-Mills theory exhibits interesting features such as the arising of a Higgs-like field. Forgacs and Manton have shown that imposing space-time symmetries on the gauge field would necessarily result in Higgs-like scalar fields [12]. This striking phenomenon naturally gives rise to the question: can the Higgs field in the electroweak theory be derived from some pure Yang-Mills theory? First of all, such a possibility is mathematically appealing. The theory would have a compact and unified form and the Higgs field would acquire a geometrical interpretation. Second, from the physics point of view, such a theory would also be more fundamental. It could also provide a possibility to unify the four fundamental forces. But life is not that easy. It is hard to tell of this question possesses a positive answer. First, the space-time symmetry of the field results in a dimensional reduction. To obtain a final theory with Higgs field in four dimensional space-time, naively we need to start with a theory in a space-time with more than four dimensions, such as in [13], where two extra dimensions were introduced. Is it necessary to introduce extra dimensions to understand the origin of the Higgs field? If yes, are the extra dimensions physical? From a certain perspective, obtaining the Higgs field from some gauge field is similar to the Kaluza-Klein mechanism for unifying gravity and electromagnetism. In both cases, the extra dimensions play a key, if not an indispensable role. In Chapter 3 I will offer a different point of view on the issue of dimensional reduction. I will argue for the possibility that there is no extra dimension. Besides this issue, we need to find a gauge group to start with, which physically makes sense and which leads to a final theory that matches the gauge sector in the electroweak theory. Manton used the group G_2 and obtained a theory with many nice aspects. However, there was no explanation why this group was chosen and in which representation of this group the fermions fields lived. The investigation presented in this chapter uses the Lorentz group. I will gauge this group and find the spherical symmetric gauge field associated to it.The investigation focuses on the mathematical structure of the theory after the spherical symmetry is imposed. In spite of its failure to reproduce the Higgs field and $SU(2) \times U(1)$ gauge group in the electroweak theory, it may serve as a starting point for further investigations.

Chapter 4 presents a set of identities of the correlations functions of a given quantum field theory. The method for deriving these identities is similar to that used for proving the Ward-Takahashi identity. These identities are interesting because they are obtained using a non-perturbative method. I have tested one of these identities in the $\lambda - \phi^4$ theory by computing the same correlation functions using the perturbative method. The result of the test is positive. However, no attempt has been made to compute any specific correlation functions using this method, since it is estimated to be very difficult, if not impossible. Interestingly, the same method for obtaining these identities can be applied to theories with gauge fields where the gauge fixing becomes unnecessary. Therefore, it might provide a new avenue to study the anomaly-related problems.

Finally, in the discussion which ends this thesis I will reflect on the results presented in each chapter and give some thoughts and ideas for further study.

Throughout this thesis, the physical constants, the speed of light c and the Plank constant $\hbar$, are set to one, unless otherwise specified.

1 An extra fermion zero mode in the chiral SU(2) theory

1.1 Introduction

In a quantum field theory with dynamical fermion fields subjected to some static external gauge field, fermion zero modes may occur, i.e. when we attempt to solve the Dirac equation with certain gauge background fields, we may find solutions associated with zero value of the Dirac Hamiltonian. If the external gauge field is time dependent, we can solve the Dirac equation for each time instant and find a time dependent fermion energy spectrum, i.e. the eigenvalues of the Dirac Hamiltonian as functions of time. We can speak of flows of the eigenvalues of the Dirac Hamiltonian. Curiously, the following phenomena may be observed: a particular flow goes from a negative value to a positive one or the other way around, i.e. it crosses zero as time elapses. When the corresponding state is occupied, the physical interpretation of such phenomena is: an anti-fermion turns into a fermion or vice verse. That is to say, the fermion number is violated. Suppose all the states are occupied, the net change of the fermion number in this course is called spectral flow. We may then ask ourselves if we can determine the spectral flow by tracking the evolution of the external gauge field?

To answer this question, we should constrain ourselves to a specific space-time. In fact, when the space-time is Euclidean, the question is already answered by the renowned Atiyha-Singer-Patodi index theorem [6]. In this theorem the spectral flow is given using "η" invariant, evaluated on the boundaries of the space-time manifold. This means, the net change of fermion number in question concerns only the topological properties of the initial and final background fields, provided that the background fields evolve smoothly. However, in Minkowskian space-time such a neat result has not yet been obtained. An alternative approach must be taken. We can find all the possible zero modes occurring in the time interval and study the level crossing at the vicinity of each zero mode. The spectral flow can be computed by synthesizing the information of level crossing at each zero mode. We might even be able to find some general rule in this approach so that an elegant quantity can be found to determine the spectral flow. Klinkhamer and Lee took took this approach in [7]. They investigated the Dirac equation in the background of the spherically symmetric $SU(2)$ gauge field. These gauge fields are solutions to the Yang-Mills equations discovered by Luescher and Schechter [16]. Imposing spherical symmetry makes the 1+3-d $SU(2)$ gauge field reduce to a single component complex scalar field with a Higgs-like potential and a $U(1)$ gauge field in 1+1-d curved space-time [8]. These special fields can be expressed in terms of elliptic functions [16]. If we make spherical symmetric *Ansatz* for the spinor field coupled to the gauge field as well, the corresponding Dirac equation also reduces to one in 1+1- space-time [17]. The zero mode equation will be a system of two coupled first order ordinary differential equations. These equations together are equivalent to a Schroedinger equation with a complex potential in one dimensional space. A very nice result has been obtained for this reduced theory: in the presence of a zero in the scalar field, which corresponds to a change of the winding number of the background field, a so-called inverse symmetry of the background fields (scalar and $U(1)$) emerges. One particular consequence of this symmetry (corresponding to the change of the winding number of the gauge field) is the occurrence of a fermion zero mode [7]. This result becomes valuable when we consider it is one of the very few concrete examples in Minkowskian space-time, which clearly displays the link between the topological property of the gauge field and the occurrence of fermion zero modes. Since the interplay between the fermion zero modes and the topological property of the gauge background field is often seen in Euclidean space-time (e.g. in theories with vortex or instanton solutions), we tend to assume that these are all the possible zero fermion modes.

Curiously, besides the zero modes accompanied with the change of winding number of the gauge fields, Klinkhamer and Lee also claimed that another fermion zero mode exists [7]: one that is not linked to the appearance of the zero in the complex scalar field (component of the spherically symmetric gauge field), hence not to the change of the winding number in the gauge background. It seems that this additional fermion zero mode is not associated to any topological property of the background gauge fields. More precisely, it seems that this fermion zero mode can appear or disappear under continuous deformation of the background gauge fields. If ture, this would be a peculiar phenomenon indeed. However, the evidence for the existence of this extra fermion zero mode provided in [7] was insufficient. Thus it is important to investigate if this extra fermion zero mode exists; and if it does, can it be identified as non-topological? In this chapter, only the first question will be answered firmly. I will prove that the extra fermion zero mode claimed by Klinkhamer and Lee in [7] does exist. First, some analytic properties of a general solution to the Dirac equation will be proven rigorously. Once certain numerical conditions are satisfied, these analytic properties then allow us to draw firm conclusions regarding the existence of the extra fermion zero mode. The existence of this extra fermion zero mode is the main focus of this chapter. Whether or not if this extra fermion zero mode interlocks with a topological quantity of the background gauge field is a question that, for now, remains unanswered.

In addition to the determination of the spectral flow, there are three more aspects of fermion zero modes worth noting. First, in the presence of a single fermion zero mode, the corresponding background field acquires fractional fermion number. This can be argued rigorously using the standard field quantization method [10]. This is a peculiar feature and its deep implication remains to uncover. Second, fermion zero modes are also found in the topological band theory in the field of condensed matter physics. If the system has a topologically non-trivial bulk-surface structure, a fermion zero mode in the edge states must occur. This zero mode can connect the valence and conduction bands. In other words, the gap existing in a insulator, which confines the electrons in the valence band at low energy, is closed. As a result, with a little extra energy (corresponding to low temperature) electrons can move from valence band to the conduction band. This then makes the system a conductor [14]. At last, fermion zero mode are often related to Majorana particles. A Majorana particle is its own anti-particle. A particle occupying a fermion zero mode necessarily carries zero charge. This enables us to form a Majorana state using the fermion zero mode. In the theory of topological superconductors Kitaev has made use of this idea [15].

This chapter is organized as follows. First, I will present a brief review of the origin of the extra zero mode claimed by Klinkhamer and Lee. Second, I will prove a theorem about the analytic properties of the solutions to the Dirac equation we are studying. Third, I will discuss the full condition for the existence of a fermion zero mode, and prove the existence of the extra fermion zero mode claimed by Klinkhamer and Lee. Fourth, I will briefly touch on a small case where we could apply the theorem proven in part two. I will end this chapter with a short discussion and conclusion will be made.

1.2 A brief review of Klinkhamer and Lee's result on the extra fermion zero mode

The theory studied by Klinkhamer and Lee is a chiral theory with one fermion coupling to $SU(2)$ gauge fields. With the spherically symmetric *Ansätze* for the gauge field and spinor, the fermion zero mode equation can be reduced to Eq.(3.9) in [7], i.e. :

$$\partial_r \Psi(t,r) = \begin{pmatrix} -\lambda(t,r) & R(t,r) \\ -R(t,r) & \lambda(t,r) \end{pmatrix} \Psi(t,r), \tag{1.1}$$

where Ψ is the two-component spinor field, λ and R are quantities formed of the spherically symmetric $SU(2)$ gauge fields. They are invariant under the gauge rotation of the reduced $U(1)$ gauge group (upon imposing the spherical symmetry, the gauge group $SU(2)$ is reduced to $U(1)$ [8]). These two quantities will be specified later. At the moment only two important properties are relevant here:

$$\lim_{r\to 0} \frac{R}{\lambda} = \lim_{r\to\infty} \frac{R}{\lambda} = 0, \tag{1.2}$$

and λ is always non-negative.

Taking the two-component spinor as a complex number and expressing it in its modulus-phase form as:

$$\Psi(t,r) = |\Psi(t,r)| e^{i\sigma^2 \Theta(t,r)} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.3}$$

with σ^2 the second Pauli matrix, we can bring Eq. (1.1) into the following form:

$$\partial_r \Theta = -\lambda \sin 2\Theta + R, \tag{1.4a}$$

$$\partial_r |\Psi| = \lambda |\Psi| \cos 2\Theta, \tag{1.4b}$$

with the boundary (initial) conditions:

$$\Theta(t,0) = 0 \mod \pi, \tag{1.5a}$$

$$|\Psi(t,0)| = 0. \tag{1.5b}$$

Klinkhamer and Lee studied the solutions to Eq. (1.4a) and noticed that they all fall into two classes: one with $\Theta(t,\infty) = m\pi$ called stable class; and the other with $\Theta(t,\infty) = n\pi + \frac{\pi}{2}$ called unstable class, where both m and n are integers. Furthermore, if a solution occurs in the unstable class, there has to be a zero mode solution to the original Dirac equation in question, because the solution to Eq. (1.4b) would, in this case, necessarily be normalizable, as $|\Psi|$ would vanish exponentially at $r = \infty$ and the boundary condition Eq. (1.5b) would also be satisfied.

In [7] Klinkhamer and Lee also pointed out, if the solution to Eq. (1.4a) makes a discontinuous jump at $r = \infty$ by π at some time t^*, then an unstable solution with $\Theta(t^*,\infty) = \frac{\pi}{2} \mod \pi$ must exist. This consequently follows the existence of a fermion zero mode at $t = t^*$. Klinkhamer and Lee found a numerical value for $t^* = -2.924$ and observed that $\Theta(t,\infty)$ jumps from π to 0 around $t = -2.924$. The relevant parameters to specify the gauge fields relevant for this observation are given in the caption of Fig. 8 in [7]. In this figure numerical solutions to Eq. (1.4a) at different time around $t = -2.924$ are plotted using Mathematica. From the plot we can derived that as r becomes large, different solutions approach different limits, differed by π. Following the intuition of "continuity" an unstable solution at $t = -2.924$ must exist. Furthermore, the occurrence of such a fermion zero mode cannot be due to the change in the winding number of the background field, because around $t = -2.924$ there cannot be zero in the modulus of the complex scalar field (this can be concluded from Eq.s (5.8) and (5.9) in [7]; discussion on the locations of zeros can also be found in [9]). As follows, this zero mode is not related to the change of the winding number of the background gauge field. From now on it will be referred to as the extra fermion zero mode.

However, to convince ourselves about the existence of the extra fermion zero mode, the numerical evidence and intuitive continuity argument are insufficient. At the very most, we can draw use the numerical evidence to draw the conclusion that there is a jump in the solution, $\Theta(t,\infty)$, around $t = -2.924$. The reasoning is twofold: no matter how closely the solution at a particular time approaches an unstable limit on a finite interval of r, it is not sufficient to exclude the possibility that it leaves the unstable limit and approaches a stable one when r gets bigger. In addition, numerical solutions necessarily involve errors. A small error can make an unstable solution stable in certain circumstances. The continuity argument, on the other hand, is, strictly speaking, invalid. We can construct a function $\Theta(t,r)$ that is continuous at any finite (t,r) yet $\Theta(t,\infty)$ takes only two values, π and $-\pi$. To construct such a counter example, it is better to think of the graph of $\Theta(t,r)$ as a beam of flows, parameterized by the time t. Imagine those flows in the middle of the beam (corresponding to values of t in the middle of the time interval) splitting smoothly into three groups as r increases. The upper and lower groups flow towards the limits π and $-\pi$ respectively, while the middle group of flows splits again at larger r. The upper and lower groups can be considered as the stable classes of flows. As we can safely say that all the flows eventually fall in these two classes, it is necessary to have a more rigorous argument for the existence of the unstable solution.

Besides the behavior of the solutions to Eq. (1.4a) at $r=\infty$, at $r=0$ their behavior needs to be examined more carefully as well. This is necessary to fully establish the existence of the fermion zero modes.

1.3 More on the stability

Due to the asymptotic behavior of the function $\frac{R}{\lambda}$, it is possible for us to get more information about solutions to Eq. (1.4a).

Since $\frac{R}{\lambda}(t,r) \xrightarrow{r\to\infty} 0$, there exists such a large r^* that if $r > r^*$,$|\frac{R}{\lambda}(t,r)|$ is smaller than a small positive number ϵ. In the following discussion $\frac{R}{\lambda}$ is taken to be positive. For the case where it is negative the discussion will be similar and the conclusion will not be altered.

Consider the value of $\Theta(t,r^*)$. Suppose $\sin 2\Theta(t,r^*) > \epsilon$, i.e. $2\Theta(t,r^*)$ is in the range $[2n\pi+\delta,(2n+1)\pi-\delta]$ with $\delta = \arcsin\epsilon$. From Eq. (1.4a) we can see that the derivative of $\Theta(t,r)$, $\partial_r\Theta(t,r^*)$, is negative in this situation. In a small interval $(r^*, r^*+\alpha)$ with α small, $2\Theta(t,r)$ gets smaller as r increases. In this interval as long as the condition $\sin 2\Theta(t,r^*) \geq \epsilon$ holds true, 2Θ decreases monotonously, until hitting the value $2n\pi+\delta$.

Now consider that $2\Theta(t,r^*)$ is close to but still bigger than $2n\pi+\delta$. $\partial_r\Theta(t,r^*)$ is still negative, so $2\Theta(t,r)$ continues to decrease in a short interval starting from r^*. But it is certain that $2\Theta(t,r)$ cannot get smaller than $2n\pi-\delta$, as $\partial_r\Theta(t,r)$ would have turned positive and $2\Theta(t,r)$ would have started growing back towards $2n\pi$. Likewise, once $2\Theta(t,r)$ becomes smaller than $2n\pi+\delta$ it cannot grow bigger than $2n\pi+\delta$, as the $\partial_r\Theta(t,r)$ would have already turned negative and $2\Theta(t,r)$ would move toward $2n\pi$ again. As r grows, $\frac{R}{\lambda}$ becomes ever smaller, we can expect that $2\Theta(t,r)$ stabilizes around $2n\pi$, with some tinier and tinier oscillations.

At r^* it could also occur that $\sin 2\Theta(t,r^*) < -\epsilon$, i.e. 2Θ is in the range $((2n-1)\pi+\delta, 2n\pi-\delta)$ for some integer n and the same δ as defined above. In a similar manner, we can conclude that in either case, $2\Theta(t,r)$ will first grow a bit and then stabilize around

$2n\pi$.

The complex situation occurs when $2\Theta(t, r^*)$ falls in the interval $[(2n+1)\pi - \delta, (2n+1)\pi + \delta]$. In this case an unstable solution might arise. In order to convince ourselves of the appearance of the unstable solution, we need more input. In section 1.4.1, a rigorous proof will be presented based on this analysis, provided that there are two solutions to Eq. (1.4a) at t_1 and t_2 with $2\Theta(t_1, \infty) = 0$ and $2\Theta(t_2, \infty) = 2\pi$, which are equivalent to the condition $2\Theta(t_1, r^*) \leq \pi - \delta$ and $2\Theta(t_2, r^*) \geq \pi + \delta$ for some appropriate r^* and δ.

We can sum up the above analysis as follows, at a large r with $|R/\lambda| < \epsilon$ if for the solution to Eq. (1.4a) $2\Theta(t, r)$ is in the range $[(2n-1)\pi + \delta, (2n+1)\pi - \delta]$, the solution will converge to $2n\pi$ at $r = \infty$. For $2\Theta(t, r)$ in the range $[(2n-1)\pi - \delta, (2n-1)\pi + \delta]$, there is a possibility that the solution can converge to the unstable value $(2n-1)\pi$ at $r = \infty$.

1.4 The occurrence of extra zero modes

1.4.1 Proof for the existence of unstable solutions to the phase equation

In this subsection, I will prove rigorously the existence of the unstable solution to Eq. (1.4a) under some special condition. It is necessary to clarify beforehand that, here, an unstable solution means a solution to Eq. (1.4a), with a proper initial value condition, which at $r = \infty$ falls in the unstable class, regardless of its value at $r = 0$. Before we can claim the extra fermion zero mode exists, however, we must examine the behavior of the solution at $r = 0$ carefully. This will be discussed in section section 1.4.2.

The proof will be derived through two steps. A particular property of continuous functions will be shown in the first step as the following lemma.

Lemma1 Suppose a function $f(x)$ is continuous on an interval $[a, b]$ and $f(a) < f(b)$, then for any two numbers α and β with $f(a) < \alpha < \beta < f(b)$, there are always such two numbers s and t with $a < s < t < b$ that $f(s) = \alpha$, $f(t) = \beta$ and for any $y \in [s, t]$ the inequality $\alpha \leq f(y) \leq \beta$ holds true.

Proof. Since $f(x)$ is continuous on $[a, b]$ and $f(a) < \alpha < \beta < f(b)$, the set $\{x | f(x) = \alpha \text{ and } x \in [a, b]\}$ is not empty. In particular, this set is closed and hence compact because $f(x)$ is continuous. Therefore, the greatest number in this set exists, which is denoted as s.Now consider the set $\{x | f(x) = \beta \text{ and } x \in [s, b]\}$. For a similar reason this set is also closed and hence compact. Therefore, the smallest number in this set exists, which is defined as t. The two numbers, s and t, chosen this way satisfy the requirement in the hypothesis. On the one hand, $f(x) \geq \alpha$ whenever $x \in [s, b]$. If this weren't true, we would be able to choose another s' in this interval with $f(s') = \alpha$, which contradicts that s is the largest in the previously defined set. On the other hand,following the same logic, we may also conclude that $f(x) \leq \beta$ whenever $x \in [s, t]$. □

Given that the functions $\lambda(t, r)$ and $R(t, r)$ are smooth, we can conclude that the solution to Eq. (1.4a) with initial value $\Theta(t, r_o) = \Theta_i$, denoted as $\Theta(t, r)$, has continuous first order derivative with respect to t and r when both parameters are finite. This follows from Theorem 6.1 in [18]. Therefore, lemma 1 applies to $\Theta(t, r)$ as a function of t on any interval $[t_1, t_2]$, once r is fixed to a finite value. We will use this fact frequently in the proof of the following theorem.

Theorem2 In the differential equation Eq. (1.4a) with initial value condition $\Theta(t, r_o) = \Theta_i$ (we must assume r_o is different than 0 for reasons explained in section 1.4.2), suppose the smooth functions $\lambda(t,r)$ and $R(t,r)$ have the properties as were specified in Eq. (1.2). If there are two solutions at t_1 and t_2 with $\Theta(t_1,\infty)=0$ and $\Theta(t_2,\infty)=\pi$, then there must be such a solution at some $t^*\in[t_1,t_2]$ that $\Theta(t^*,\infty)=\frac{\pi}{2}$.

Proof. We can prove the theorem using a contradiction. The idea is to show that there is a t^* such that $2\Theta(t^*,r)$ is bounded in an interval $[\pi-\delta,\pi+\delta]$ for arbitrarily small positive number δ and arbitrarily large r.

Step 1. Since $\lambda(t,r)$ and $R(t,r)$ are smooth functions of finite t and r, and $\frac{R}{\lambda}\xrightarrow{r\to\infty}0$, for any small positive number ϵ there must be such a large but finite r^* that $|\frac{R}{\lambda}|\le\epsilon$ for all $(t,r)\in[t_1,t_2]\times[r^*,\infty)$. Note that this r^* can be chosen independent of t, since $R/\lambda\xrightarrow{r\to\infty}0$ for all t in the compact interval $[t_1,t_2]$. Thus, according to the previous analysis on the stability of the solutions, we conclude that for $(t,r)\in[t_1,t_2]\times[r^*,\infty)$, if $2\Theta(t,r^*)$ is in the interval $[\pi+\arcsin\epsilon,3\pi-\arcsin\epsilon]$ or $[-\pi+\arcsin\epsilon,\pi-\arcsin\epsilon]$, $\Theta(t,r)$ will converge to π or 0 at $r=\infty$ (note that sometimes a factor 2 is included to save words). In the following text $\arcsin\epsilon$ will be denoted as δ.

Step 2. Since $2\Theta(t_1,\infty)=0$ and $2\Theta(t_2,\infty)=2\pi$, for any small positive δ it is possible to find a number r_1 such that $2\Theta(t_1,r_1)<\pi-\delta$ and $2\Theta(t_2,r_1)>\pi+\delta$. Since $\Theta(t,r_1)$ is continuous on the interval $[t_1,t_2]$, by lemma 1 there exist such two numbers a_1 and b_1 with $t_1<a_1<b_1<t_2$ that $2\Theta(a_1,r_1)=\pi-\delta$ and $2\Theta(b_1,r_1)=\pi+\delta$, and $\pi-\delta\le 2\Theta(t,r_1)\le\pi+\delta$ for $t\in[a_1,b_1]$. According to the discussion in step 1, we conclude that the solutions at a_1 and b_1 both fall into stable classes, i.e. $2\Theta(a_1,\infty)=0$ and $2\Theta(b_1,\infty)=2\pi$.

Step 3. Now consider the value of $2\Theta(\frac{a_1+b_1}{2},r)$. There are three cases.

Case 0: $2\Theta(\frac{a_1+b_1}{2},\infty)=\pi$, theorem is proven.

Case 1: $2\Theta(\frac{a_1+b_1}{2},\infty)=2\pi$,

We consider this condition together with $2\Theta(a_1,\infty)=0$. There must be such a number $r_2>r_1+1$ that $2\Theta(a_1,r_2)<\pi-\delta$ and $2\Theta(\frac{a_1+b_1}{2},r_2)>\pi+\delta$. By lemma 1 there must be such two numbers, a_2 and b_2, that $a_1<a_2<b_2<\frac{a_1+b_1}{2}$, $2\Theta(a_2,r_2)=\pi-\delta$, $2\Theta(b_2,r_2)=\pi+\delta$, and $\pi-\delta\le2\Theta(t,r_2)\le\pi+\delta$ for all $t\in[a_2,b_2]$.

Case 2: $2\Theta(\frac{a_1+b_1}{2},\infty)=0$.

We consider this condition together with $2\Theta(b_1,\infty)=2\pi$. There must be such a number $r_2'>r_1+1$ that $2\Theta(\frac{a_1+b_1}{2},r_2')<\pi-\delta$ and $2\Theta(b_1,r_2)>\pi+\delta$. By lemma 1 there must be such two numbers, a_2' and b_2' that $\frac{a_1+b_1}{2}<a_2'<b_2'<b_1$, $2\Theta(a_2',r_2')=\pi-\delta$, $2\Theta(b_2,r_2')=\pi+\delta$, and $\pi-\delta\le2\Theta(t,r_2')\le\pi+\delta$ for all $t\in[a_2',b_2']$.

Case 1 and case 2 cannot occur simultaneously. It is possible to define a unique set (a_2,b_2,r_2) when case 0 does not occur.

Step 4. Now repeat step 3 and define (a_3,b_3,r_3), (a_4,b_4,r_4)..., unless case 0 in step 3 is encountered, in presence of which the theorem is already proven and the sequence terminates. In the following discussion we will assume that an infinite sequence has been constructed. We claim that the two sequences $\{a_n\}$ and $\{b_n\}$ converge to the same limit, denoted by t^*. This is because $t_1<a_n<b_n<t_2$, $a_n\le a_{n+1}$, $b_{n+1}\le b_n$ and in particular the gap sequence $\{b_n-a_n\}$ converges to 0 faster than the geometric sequence $\{2^{-n}\}$.

It is clear from step 2 and step 3, that t^* is in the interval $[a_n,b_n]$ for arbitrary positive integer n, hence $t^*\in[t_1,t_2]$. We also conclude, $r_n>r^*+n-1$ and $\pi-\delta\le2\Theta(t^*,r_n)\le\pi+\delta$.

Evidently, if $\Theta(t^*, \infty) \neq \frac{\pi}{2}$ we would run into a contradiction because we know that $\Theta(t^*, \infty)$ can only be integer multiples of $\pi/2$. □

1.4.2 Another condition for the existence of fermion zero modes

If a fermion zero mode were to exist, it would not only put constraint on the asymptotic behavior of $\Theta(t, r)$ at large r, but also on its behavior at $r = 0$. It was noted in [7] that for a regular function for $|\psi|$ at $r = 0$, the necessary boundary condition is

$$\Theta(t, 0) \equiv 0 \mod \pi. \tag{1.6}$$

This boundary condition should be examined carefully. Since $\lambda(t, r)$ is divergent as $1/r$ at $r = 0$, Eq. (1.6) cannot be taken as an initial value condition because it cannot determine a unique solution to Eq. (1.4a). A proper way to treat this problem is to set the value for $\Theta(t, r)$ at a nonzero finite r_o (here t is taken as a constant parameter):

$$\Theta(t, r_o) = \Theta_i. \tag{1.7}$$

Eq. (1.4a) and Eq. (1.7) together form a proper initial value problem for a first order differential equation. The study on ODE (e.g. Theorem 6.1 [18]) says that there is a unique solution to this problem. However, the boundary condition specified in Eq. (1.6) is not likely to be satisfied by this solution. To see why Eq. (1.6) does not serve as a proper initial condition, we should make a change of variable: $x = \ln r$. Then Eq. (1.4a) becomes

$$\partial_x \Theta = -\rho \sin 2\Theta + e^x R. \tag{1.8}$$

Now x ranges in the interval $(-\infty, \ln r_o]$ and $x = -\infty$ corresponds to $r = 0$. The requirement that $\Theta(t, 0) \equiv 0 \mod \pi$ translates into $\Theta|_{(t,x=-\infty)} \equiv 0 \mod \pi$. An analysis similar to that in section 1.3 shows that the value of a solution to Eq. (1.8) at $x = -\infty$ (or the value of a solution to Eq. (1.4a) at $r = 0$) necessarily falls into two classes: $\Theta|_{(t,x=-\infty)} \equiv 0 \mod \pi$ and $\Theta(t, x = -\infty) \equiv \frac{\pi}{2} \mod \pi$. This means that if we attempt to determine a unique solution to Eq. (1.8), especially in a numerical way, it is insufficient to require $\Theta(t, x = -\infty)$ $(\Theta(t, r = 0))$ to take some specific value. More importantly, the analysis shows that the class with $\Theta|_{(t,x=-\infty)} \equiv 0 \mod \pi$ is unstable: any perturbation would make it fall into the class of $\Theta|_{(t,x=-\infty)} \equiv \frac{\pi}{2} \mod \pi$. Therefore, for a fermion zero mode to exist, it is necessary that the solution to Eq. (1.4a) approaches to a value in the unstable class not only at $r = \infty$, but also at $r = 0$.

Even if we find a particular value for Θ_i in Eq. (1.7) that makes the corresponding solution satisfy the condition at $r = \infty$, then in general this solution would not satisfy the condition at $r = 0$; or the other way around. Thus the numerical evidence provided in [7] bears one more defect: because of its numerical nature it does not give information about the stability of the solution at $r = 0$. In other words, the value of that solution at $r = 0$ probably falls into the stable class, thereby ruining the normalizablity of the corresponding spinor field at $r = 0$. For this reason, it is necessary to revise the condition on the existence of the extra fermion zero mode discussed in [7].

In general, it is hard to establish the occurrence of such a solution, because we simply cannot solve the corresponding Dirac equation (equivalent to a Schrödinger equation with general complex potential). One alternative is to combine analytic analysis and numerical study. In the following I will introduce a method combining the the analytic and numeric analysis.

For a given time t_o, let $\Theta(t_o, r, \Theta_i)$ denote the solution to Eq. (1.4a) with the initial value condition as in Eq. (1.7). Now $\Theta(t_o, r, \Theta_i) + \pi = \Theta(t_o, r, \Theta_i + \pi)$ is also a solution with a different initial value, $\Theta_i + \pi$. Comparing these two solutions at $r = \infty$, there is a change by π. If Θ_i changes smoothly, the solution $\Theta(t_o, r, \Theta_i)$ would change smoothly at finite r as well. Thus, according to theorem 2 (we have to assign the role played by t to Θ_i) there must be such a Θ^* in the interval $[\Theta_i, \Theta_i + \pi]$ that the corresponding solution to Eq. (1.4a) satisfies $\Theta(t_o, r = \infty, \Theta^*) = \frac{\pi}{2} \mod \pi$. There could be more than one value in this interval that satisfies the same condition. When the smallest value in this interval is chosen, a function $\Theta^*(t)$ can be defined: $\Theta^*(t)$ equals the smallest value in the interval $[\Theta_i, \Theta_i + \pi]$ that $\Theta(t, r_o)$ in Eq. (1.7) has to take at time t, so that $\Theta(t, r = \infty, \Theta^*(t)) = \frac{\pi}{2} \mod \pi$. Furthermore, it can be proven that this function is continuous, provided that $R(t, r)$ and $\lambda(t, r)$ have proper continuity properties (see lemma 3 in the appendix). Those values that differ by an integer multiple of π to the Θ^* chosen here will lead to the same spinor solution, and will therefore not be of interest.

Using the function $\Theta^*(t)$ the following corollary of theorem 2 can be obtained:

Corollary3 There exists such an interval

$$[-\arcsin|\frac{R(t_o, r_o)}{\lambda(t_o, r_o)}|, \arcsin|\frac{R(t_o, r_o)}{\lambda(t_o, r_o)}|], \tag{1.9}$$

defined for a sufficiently small r_o and some $t_o \in [t_1, t_2]$, that if $\Theta^*(t_1)$ and $\Theta^*(t_2)$ fall on different sides outside this interval, there must exist such a $t^* \in [t_1, t_2]$ that $\Theta(t^*, r = 0, \Theta^*(t^*)) = 0 \mod \pi$. Consequently, a fermion zero mode exists at t^*.

Proof. In the proof of theorem 2, we recall that at some sufficiently large r, if the value of $\Theta(t, r)$ falls outside a certain small interval $([\pi/2 - \delta/2, \pi/2 + \delta/2])$, the solution will necessarily fall into a stable class as $r \to \infty$. Moreover, those solutions whose values at r fall on different sides of the interval will end up in different stable classes. We can make a similar observation on the behavior of the solutions to Eq. (1.4a) around $r = 0$. We can first find a sufficiently small r_o and some $t_o \in [t_1, t_2]$ so that $\arcsin|R(t_o, r_o)/\lambda(t_o, r_o)|$ attains the maximal value on the domain $[t_1, t_2] \times (0, r_o]$. If the hypothesis in the corollary is satisfied, $\Theta(t_1, r = 0, \Theta^*(t_1))$ and $\Theta(t_1, r = 0, \Theta^*(t_2))$ will differ by π. Therefore, there is a $\Theta^*(t^*) \in [\Theta^*(t_1), \Theta^*(t_2)]$ with the desired property. In addition, we can argue that $t^* \in [t_1, t_2]$ using lemma 1. □

Corollary 3 provides theoretical validation and guidance for the numerical search for fermion zero modes. It is possible to obtain a numerical plot for the function $\Theta^*(t)$ with high precision. The stability of the stable solutions to Eq. (1.4a) enables us to predict their behavior as $r \to \infty$ with certainty by knowing their values at some finite (possibly large) r. We can then test a pair of initial values in the interval $[\Theta_i, \Theta_i + \pi]$ for some Θ_i, and see if they give solutions in different stable classes. If this is the case by theorem 2 we know that the initial value corresponding to an unstable solution is bounded by this pair chosen in the interval. By testing pairs that are getting closer and closer, the estimation of the desired initial value can be made more and more accurate (it is possible to write a program in mathematica to do this work). Once the numerical plot for the function $\Theta^*(t)$ is made, we can check if the hypothesis in corollary 3 is true. As soon as it is known to be true, the existence of a fermion zero mode is established.

1.5 Evidence for the existence of the extra fermion zero mode

I have clarified that the condition for a fermion zero mode to exist consists of two sides: the corresponding solution to Eq. (1.4a) falls both in the unstable class at $r = 0$ as well as in the unstable class at $r = \infty$. According to the analysis in the previous sections, the condition can be realized on each side separately. Klinkhamer and Lee discussed in [7] in case of change in the winding number of the spherically symmetric field, there is an additional symmetry, the inverse symmetry. This symmetry ensures that if the condition is fulfilled on one side, it will also be fulfilled on the other side. But can the condition hold true on both sides simultaneously without this symmetry?

The answer is yes in the case of the Luescher-Schechter solutions. It is crucial to observe the following numerical fact: there exists such a continuous interval $t^* \in [-2.9245, -2.9235]$, that for any t^* in this interval $\Theta(t^*, r = \infty, \Theta_i) = \pi/2 \mod \pi$, for some initial value around $\Theta_i = 0.138$ at $r_o = 0.8$ (in comparison, in [7] only a single time point $t^* = -2.924$ was mentioned). If we take a moment to reflect upon the problem for a moment, the existence of this continuous interval, instead of a single value, is not surprising. This is due to the continuous property of $\lambda(t, r)$ and $R(t, r)$. Precisely because of the existence of this continuous critical time interval, the probability of finding an extra fermion zero mode in this neighborhood in the parameter space is increased substantially. We can assume without loss of generality that in this time interval, for a given Θ_i there is only one t^* with $\Theta(t^*, r = \infty, \Theta_i) = \pi/2 \mod \pi$. Intuitively, when Θ_i varies continuously, the function $t^*(\Theta_i)$ also varies continuously. In the appendix, a proof for this continuity is provided (lemma 2). This continuity will turn out to be essential for the existence of the extra fermion zero mode that I am going to prove.

The full analytic and numerical evidence for the existence of this solution is presented below. I will assume that the difference between the numerical solutions to Eq. (1.4a) and the true solutions is bounded by a small error at finite r. I assume that this error is small enough that it won't affect the stability of a truely stable solution. As long as this assumption is accepted (more evidence can be obtained by checking the numerical calculation with a different parameterization of the equation), I will be able to conclude that the extra fermion zero mode must occur.

1.6 Luescher and Shechter solutions and the numerical method

The claim for the existence of the extra fermion zero mode is only for a particular type of gauge background, namely the Luescher-Shechter solutions to the Yang-Mills field equations. In this section, these special gauge backgrounds are specified first. In order to carry out the numerical calculations, the parameters in $\lambda(t, r)$ and $R(t, r)$ will also be specified. At the end of this section, numerical results will be presented that will prove the existence of the extra-fermion mode.

The spherically symmetric solutions to the $SU(2)$ Yang-Mills equations in the 4-D Minkowskian spacetime, discovered by Luescher and Schechter separately [16], are expressed as follows (the same notations are taken as in [7]):

$$
\begin{aligned}
\alpha =& \rho \cos \varphi = 1 + q(\tau) \cos^2 w, \\
\beta =& \rho \sin \varphi = \frac{1}{2} q(\tau) \sin 2w, \qquad (1.10) \\
a_r =& q(\tau) \partial_r w,
\end{aligned}
$$

with

$$\begin{aligned}\tau \equiv&\mathrm{sgn}(t)\arccos(\frac{1+r^2-t^2}{\sqrt{(1+t^2-r^2)^2+4r^2}}),\\ w\equiv&\arctan(\frac{1-r^2+t^2}{2r}).\end{aligned} \tag{1.11}$$

The function $q(\tau)$ is a solution to the following E.O.M for a Higgs-like field:

$$\frac{d^2q}{d\tau^2}+2q(q+1)(q+2)=0. \tag{1.12}$$

The last equation can be solved exactly and the solution $q(\tau)$ is essentially a Jacobian elliptic function of the type $dn(\tau)$ or $cn(\tau)$. The complex scalar field $\Phi=\alpha+i\beta\equiv\rho e^{i\varphi}$ with $\varphi(t,0)=0$ is the Higgs like field, and a_r is the spatial component of the reduced abelian gauge field. $\alpha,\ \beta,\ a_r,\ \rho$ and φ are all real. In addition, ρ is non-negative.

The parameter functions used before, $\lambda(t,r)$ and $R(t,r)$, are related to the above quantities in the following way:

$$\lambda(t,r) = \rho(t,r)/r, \tag{1.13}$$

$$R(t,r) = (a_r-\partial_r\varphi)/2. \tag{1.14}$$

Once $\lambda(t,r)$ and $R(t,r)$ are specified, Eq. (1.4a) can then be numerically solved with a choice for the initial value for the solution according to Eq. (1.7). However, it will be convenient to define a new quantity to solve Eq. (1.4a):

$$\phi(t,r)\equiv 2\Theta(t,r)+\varphi(t,r). \tag{1.15}$$

With some algebra Eq. (1.4a) can then be transformed into the following equation:

$$\partial_r\phi=\frac{2}{r}(\alpha\sin\phi-\beta\cos\phi)+a_r. \tag{1.16}$$

Before we begin to numerically solve this equation for ϕ, we need to first fix the function $q(\tau)$. The numerical study presented in this chapter is done for the same $q(\tau)$ as in [7], namely,

$$q(\tau)=-1+(1+\sqrt{2\epsilon})^{1/2}cn[(8\epsilon)^{1/4}(\tau+(8\epsilon)^{-1/4})K(m)|m], \tag{1.17}$$

where $m=\frac{1+\sqrt{2\epsilon}}{2\sqrt{2\epsilon}}$ is the modulus, $K(m)$ is the complete elliptic integral of the first kind (for more information on elliptic functions, please see [7] and references therein), and $\epsilon=20$.

At first glance, it appears that theorem 2 cannot be applied to Eq. (1.16). However, we will see this is not true. Since there is no zero in the complex scalar field in the neighborhood of $t=-2.924$, the winding number of complex scalar field, i.e. $\varphi(t,\infty)$ does not change in this neighborhood. Hence we can conclude that, a change of π in $\Theta(t,\infty)$ is equivalent to a change of 2π in $\phi(t,\infty)$, in the neighborhood of $t=-2.924$ according to Eq. (1.15). Suppose under some suitable conditions, applying theorem 2 we obtain an unstable solution to Eq. (1.4a). The substitution of this solution into Eq. (1.15) will result in an unstable solution to Eq. (1.16). Thus we can derive a similar result for Eq. (1.16) as theorem 2: if the solution to Eq. (1.16) with some fixed initial value has such a property that its value at $r=\infty$ changes discontinuously by 2π around a time point t^*, then there must exist an unstable solution at t^* with the same initial value and $\phi(t^*, \infty) = \pi$. The following analysis is based on this fact.

Numerical study prompts us to focus on these two intervals: $[\phi_1, \phi_2] = [0.13977, 0.14007]$, for the initial value ϕ_i at $r_o = 0.8$, and $[t_1, t_2] = [-2.9245, -2.9235]$, for the time. This data is found in the following way: in [7] numerical result was obtained that suggested a solution $\Theta(t = -2.924, r)$ to Eq. (1.4a) with $\Theta(-2.924, r \sim 0) = 0$ and $\Theta(-2.924, r \to \infty) \to \pi/2 \mod \pi$. Equivalently, this implies that there is a solution $\phi(t = -2.924, r)$ to Eq. (1.16) with $\phi(-2.924, r \sim 0) = 0$ and $\phi(-2.924, r \to \infty) \to \pi \mod 2\pi$. On closer inspection I found that around $t = -2.924$ for a wide range of initial values ϕ_i at $r \sim 0$, the solution quickly converges to 0.13980 at $r \sim 0.8$. Thus, it is better to zoom into the initial value at around $r = 0.8$ to have a closer look. Therefore, I came to the choice of parameter space mentioned at the beginning of the paragraph.

Numerical study shows that for whatever initial value taken at $r = 0.8$ in the interval $[0.13977, 0.14007]$, at time $t = -2.9235$, the solutions to Eq. (1.16) fall in the same stable class: $\phi(-2.9235, r = \infty, \phi_i) = -2\pi$ (see Fig. 1.1-a). On the other hand, at time $t = -2.9245$, the solutions for all initial values in the same interval fall in another stable class: $\phi(-2.9245, r = \infty, \phi_i) = 0$ (see Fig. 1.1-b). This implies that there is a continuous function, $t^*(\phi_i)$, defined on the interval $[0.13977, 0.14007]$ with t^* takes values in the interval $[-2.9245, -2.9235]$ according to lemma 2 in the appendix. Note that $t^*(\phi_i)$ is conceptually the same as the function $t^*(\Theta_i)$ defined before.

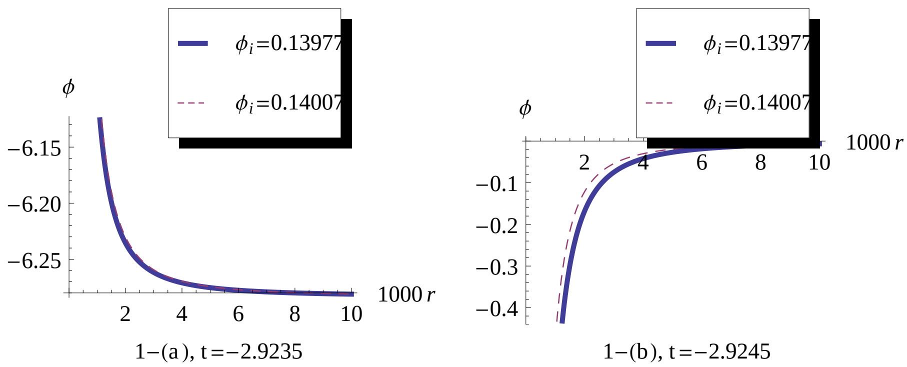

Figure 1.1: The function $t^*(\phi_i)$ of the initial value ϕ_i can be defined on the interval $[0.13977, 0.14007]$. Four solutions to Eq. (1.16) at different times with different initial values of ϕ_i are plotted. It can be predicted that the two solutions in 1-(a) will converge to the limit -2π, while those in 1-(b) will converge to the limit 0, as $r \to \infty$.

Now look at the other direction $r \to 0$. Numerical study shows that for whatever value of t in the time interval, as long as the initial value at $r = 0.8$ is taken to be 0.13977, the solutions to Eq. (1.16) fall into the same stable class, $-\pi$, at $r = 0$ (Fig. 1.2-a). On the other hand, when the initial value is taken to be 0.14007, the solutions for all t in the time interval fall into the same stable class, π, at $r = 0$ (Fig. 1.2-b). This implies that there is a continuous function, $\phi_i^*(t)$, defined on the interval $[-2.9245, -2.9235]$ with ϕ_i^* taking value in the interval $[0.13977, 0.14007]$ according to lemma 3 in the appendix. Note that the function $\phi_i^*(t)$ is of the same concept as of $\Theta_i^*(t)$ defined before, the only difference being that $\phi_i^*(t)$ is defined according to the behavior of the solutions at $r = 0$ while $\Theta_i^*(t)$ according to that at $r = \infty$.

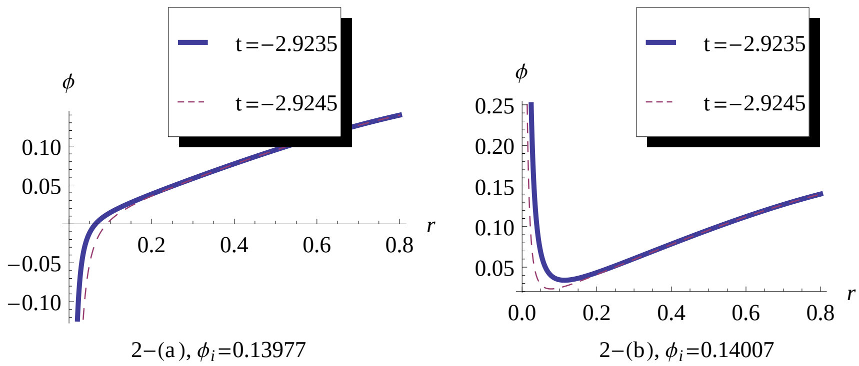

Figure 1.2: The function $\phi_i^*(t)$ can be defined on the time interval $[-2.9245, -2.9235]$. Four solutions to Eq. 1.16 at different times with different initial values of ϕ_i are plotted. It can be predicted the two solutions in 2-(a) will converge to the limit $-\pi$, while those in 2-(b) will converge to the limit π as $r \to 0$.

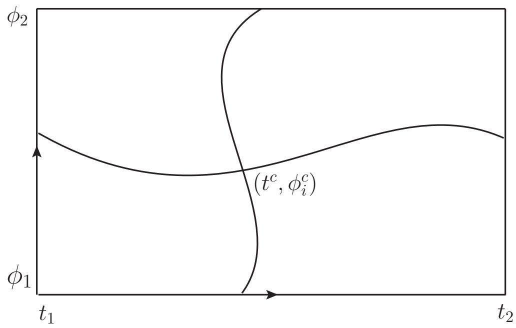

Figure 1.3: Two random curves corresponding to $\phi_i(t)$ and $t(\phi_i)$ are plotted in the same rectangle. The intersection point is denoted as (t^c, ϕ_i^c).

If graphs of the two functions, $t^*(\phi_i)$ and $\phi_i^*(t)$, are drawn on the same plane $t - \phi_i$, there must be at least one intersection point (t^c, ϕ_i^c) (here c stands for 'critical'). A typical picture of such a case is shown in Fig. 1.3.

Our intuition suggests that the two curves must intersect. Rigorous analytic proof for this fact is tedious and not necessary here. A relatively intuitive argument exists that may serve as a proof. Observe that the two graphs have to fall in the same rectangle, $[t_1, t_2] \times [\phi_1, \phi_2]$, and that each curve connects a pair of edges of the rectangle continuously. Now suppose there is no intersection point, the rectangle could be cut along one of the curves, say the graph of $t^*(\phi_i)$. The rectangle would then be separated into two disjoint parts, with each containing an edge that is connected to the other curve. Since there was no intersection point, the other curve would not be cut and should still connect the two edges from the two disjoint parts continuously. This is a contradiction.

According to the previous discussions in this section, it is clear that the solution $\phi(t^c, r, \phi_i^c)$ to Eq. (1.16) falls into an unstable class at both $r = 0$ and $r = \infty$. From Eq. (1.15) we can obtain a solution $\Theta(t^c, r)$ to Eq. (1.4a). It will have the property that $\Theta(t^c, 0) = 0 \mod \pi$ and $\Theta(t^c, \infty) = \pi/2 \mod \pi$. From the solution $\Theta(t^c, r)$, a fermion zero mode can be constructed. As was mentioned before, around t^c (close to $t = -2.924$) the winding number does not change, this fermion zero mode is not apparently related to the topological number (winding number) of the gauge fields, and is thus referred to as the extra fermion zero mode. If we slightly adjust the parameters in the numerical study, e.g. the energy of the gauge field, we could find other fermion zero modes that share the same nature as this extra fermion zero mode. But this is not going to be pursued.

1.7 A special case where fermion zero modes can be predicted

It is surprising that the chiral abelian Higgs model in $1+1$ dimension displays a large similarity to the spherically reduced $SU(2)$ theory. Therefore, it is straightforward to apply the methods used in the previous sections to the chiral abelian Higgs model. The action of this theory is [19]:

$$S = \int_{\mathbf{R}^{1,1}} d^2x(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}+|D_\mu\chi|^2-V(|\chi|^2)+f\bar{\psi}|\chi|e^{i\gamma^5\phi}\psi(x)+\bar{\psi}\gamma^\mu(\partial_\mu+iA_\mu\gamma^5)\psi, \quad (1.18)$$

where $\chi = |\chi|e^{i\phi}$ is the complex scalar field with the double well potential $V(|\chi|^2)$, γ^5 is chosen to be the second Pauli matrix, σ^2. Note that the Yukawa coupling between the fermion and the Higgs fields is chiral. The Lagrangian density of this theory is invariant under the following gauge transformation:

$$\begin{aligned} \psi &\rightarrow e^{\frac{i}{2}\alpha\gamma^5}\psi, \\ \phi &\rightarrow e^{-i\alpha}\phi, \\ A_\mu &\rightarrow A_\mu - \frac{1}{2}\partial_\mu\alpha, \\ F_{\mu\nu} &\rightarrow F_{\mu\nu}. \end{aligned}$$

The transformation rule indicates that the charge of the fermion is half of that of the Higgs particle. The other gamma matrices are chosen to be:

$$\gamma^0 = i\sigma^1, \quad \gamma^1 = -\sigma^3. \quad (1.19)$$

Now the zero mode Dirac equation reads as:

$$\gamma^1\partial_1\psi + i\gamma^\mu A_\mu\gamma^5\psi + f(\mathrm{Re}\chi + i\mathrm{Im}\chi\gamma^5)\psi = 0. \quad (1.20)$$

By using the explicit expressions for the gamma matrices here, this equation is transformed into

$$\partial_1\psi = -[iA_0 + iA_1\gamma^5) + f(\mathrm{Re}\chi\gamma^1 + i\mathrm{Im}\chi\gamma^0)]\psi. \quad (1.21)$$

The time component of the gauge field contributes only a phase factor to the fermion field, and can therefore be left out. Now this equation looks almost identical to Eq. (3.5) in [7].

Following the same procedure carried out in [7] we multiply to both sides of Eq. (1.21) the unitary matrix

$$-i\gamma^0 e^{i\phi\frac{\gamma^5}{2}}. \quad (1.22)$$

Then the equation is transformed into

$$\partial_1\psi = [-|\chi|\sigma^3 + i(A_1 - \partial_1\phi)\sigma^2]\psi. \quad (1.23)$$

By using Eq. (1.3) this equation is equivalent to:

$$\begin{aligned} \partial_1\Theta &= -|\chi|\sin 2\Theta + A_1 - \partial_1\phi, \\ \partial_1|\psi| &= |\chi|\cos 2\Theta. \end{aligned} \quad (1.24)$$

These equations are exactly the same as Eq. (1.4) by expressing them in terms of $x = \ln r$, even the boundary conditions become correspondent. Therefore, the results obtained in the previous sections are applicable to Eq. (1.24) as well. In case χ is real and has opposite signs at $x = -\infty$ and $x = \infty$, there is necessarily a fermion zero mode. This special case is essentially the same as fermions interacting with a kink background. The difference is that here the presence of gauge field makes the problem more complicated. Thus it is necessary to invoke theorem 2 in order to prove the existence of the fermion zero mode.

1.8 Conclusion and discussion

The main result presented in this chapter functions as a rigorous proof for the existence of the extra fermion zero mode discovered by Klinkhamer and Lee [7]. The stability of solutions to the nonlinear ordinary differential equation, Eq. (1.4a), is studied carefully. The results of this study may be useful for searching for fermion zero modes in other theories.

What makes the extra fermion zero mode so important is that it may change the spectral flow, which is directly linked to fermion number violation. How exactly this extra fermion zero mode influences the spectral flow has yet to be investigated. Such investigation requires us to keep track of the zero mode and to examine how its energy changes smoothly with the variation of the background gauge field. This is generally a difficult problem. However, in the case of 1+1-d theory, Kaufhold found a very elegant method to determine the global spectral flow without having to track each individual fermion zero modes [19]. Kaufhold's method inspires us with a new approach to the difficult problem of fermion number violation. It deserves attention, and effort for generalization to the higher dimensional spaces.

Whether or not there exists a topological quantity of the background field responsible for the appearance of this extra fermion zero mode is another hard question to answer. On the one hand, results in [7] show that most of the fermion zero modes in the theory studied there are related to the change of the topological quantity, the winding number of the gauge field, via the appearance of zeros in the complex scalar field Φ (page 16). On the other hand we can assert that at the time when the extra fermion zero mode appears, there is no zero in the complex scalar field. From this we can only conclude that, the extra fermion zero mode is not related to the winding number. But in principle, there could be some other topological number of the background gauge field associated to the extra fermion zero mode.

2 A bounded and discrete fermion mass spectrum

In this chapter I will present a mechanism to determine a fermion mass spectrum, based on my work with F. Klinkhamer [20]. This presentation will compact. For more details please refer to [20].

As far as we know, matter (dark matter is not considered) in our world consists of fundamental fermions such as electrons, neutrinos as well as up and down quarks. The rest mass, as a property of a specie of fundamental fermions, is as important as the electric charge and the spin. After quantum mechanics was discovered, we have made great progress in understanding these properties of fermions. Dirac's research on magnetic monopoles provides a probable explanation of the quantization of electric charges, carried by the fermions. We also know that the spins of fermions must take half odd integer values from the theory of group representation. The special values of the mass parameters of fermions, however, remain a puzzle to solve. Although people believe that the Higgs mechanism of the standard model theory explains how the fermions acquire masses, it is clear that the mechanism does not determine the mass value of a given fermion.

In spite of their shortcomings, some mechanisms help to derive fermion mass spectra. One such mechanism makes use of Kaluza and Klein's idea of extra dimensions. In this mechanism one introduces a compact extra dimension and writes down a Klein-Gordon equation for a "zero-mass" particle of the five dimensional space-time. The fifth component of the momentum is interpreted as the mass and its value is quantized due to the boundary condition in the fifth dimension. The prediction is, that there are an infinite tower of particles with evenly gaped masses. The other prediction was made by Dirac. In his study on field theory in conformal space, an equation of motion with conformal symmetry was obtained. From this equation he was able to obtain a similar fermion mass spectrum as given by the first mechanism.

If we believe that the masses of fermions can be predicted by some mechanism at all, we would probably hope that the mechanism provides a finite spectrum. More pleasing would be one where the values in the spectrum have some interesting structure. Ideally, the values should match with measurement. The new mechanism to be introduced here does provide a finite spectrum and the values in the spectrum do have an interesting structure. Unfortunately the values do not match with the experiment.

2.1 Theoretical derivation

The theory we start with describes a pair of fermionic fields interacting with a Higgs-like scalar field in a five dimensional spacetime. The two fermionic fields are assumed to have opposite coupling constants and the fifth dimension is open. The Lagrangian density for this theory reads as follows:

$$\mathcal{L}_5 = \bar{\Psi}\, i\partial\!\!\!/\, \Psi + \bar{\Omega}\, i\partial\!\!\!/\, \Omega - f\, \bar{\Psi}\Psi\, \phi + f\, \bar{\Omega}\Omega\, \phi + \frac{1}{2}\, \partial_a\phi\, \partial^a\phi - \frac{\lambda^2}{2}\, \left(\phi^2 - M^2\right)^2 . \tag{2.1}$$

We can readily write down the equations of motion for the two fermionic fields Ψ and Ω, and for the scalar field ϕ.

The equation of motion for the scalar field contains a source contribution from the fermionic fields. In the absence of this contribution, the equation admits the well-known kink solution, which in this case is chosen to depend solely on the coordinate of the fifth dimension, w:

$$\phi(w) = M \tanh(\lambda M w). \tag{2.2}$$

In the following approach we assume that the fermionic source terms cancel consistently. We then find solutions to the equations of motion for the fermionic fields, and verify that the fermionic source terms indeed cancel for the solutions we find.

Substituting the kink solution into the equation of motion for the fermionic field, say, Ψ, we obtain:

$$\left(i\gamma^\mu\partial_\mu + i\Gamma^5\partial_w\right)\,\Psi = +f\,M\tanh(c\,w)\,\Psi, \tag{2.3}$$

where we introduced a new gamma matrix: $\Gamma^5 = -i\gamma^5$. Upon writing Ψ as

$$\Psi = \begin{pmatrix} \psi_l \\ \psi_r \end{pmatrix} \tag{2.4}$$

Eq. (2.3) becomes

$$\begin{aligned} i\sigma^\mu\partial_\mu\psi_r - \partial_w\psi_l &= f\,M\tanh(c\,w)\,\psi_l\,, \\ i\overline{\sigma}^\mu\partial_\mu\psi_l + \partial_w\psi_r &= f\,M\tanh(c\,w)\,\psi_r\,. \end{aligned} \tag{2.5}$$

It is necessary for us to make two Ansätze to proceed further:

$$\psi_l = v_l(w)\chi(x), \quad \psi_r = v_r\xi(x) \tag{2.6}$$

and

$$i\sigma^\mu\partial_\mu\xi(x) = m_4\chi(x), \quad i\bar{\sigma}^\mu\partial_\mu\chi(x) = m_4\xi(x), \tag{2.7}$$

where x stands for the coordinates in the four dimensional space-time, and σ^μ and $\bar{\sigma}^\mu$ are four vectors made of the identity and the three Pauli matrices. It is natural to interpret the spinors $\chi(x)$ and $\xi(x)$ as the left and right handed components of the Dirac spinor in the four dimensional space-time. The parameter m_4 then acquires the natural meaning of the mass of the Dirac spinor. With these Ansätze the $v_{l(r)}$ factors are decoupled from the $\chi(x)$ and $\xi(x)$ functions. We obtain equations for v_l and v_r with straightforward calculations:

$$\begin{aligned} (1-s^2)\,\frac{d^2v_l(s)}{ds^2} - 2s\,\frac{dv_l(s)}{ds} + \left[F(F+1) - \frac{F^2-m_f^2}{1-s^2}\right]\,v_l(s) = 0\,, \\ (1-s^2)\,\frac{d^2v_r(s)}{ds^2} - 2s\,\frac{dv_r(s)}{ds} + \left[F(F-1) - \frac{F^2-m_f^2}{1-s^2}\right]\,v_r(s) = 0\,, \end{aligned} \tag{2.8}$$

where $s = \tanh w$, $m_f = m_4/(\lambda M)$ and $F = \frac{f}{\lambda}$. From these equations we can conclude that, in the case $F \geq 2$ being an integer (for non-integer $F > 2$ similar result can be obtained, see the appendix of [20]), m_f can only take values in the following set:

$$m_f \in \left\{\pm\sqrt{F^2-1^2}, \pm\sqrt{F^2-2^2}, \pm\sqrt{F^2-3^2}, \ldots, \pm\sqrt{F^2-(F-1)^2}\right\}\,, \tag{2.9}$$

which gives us a fermion mass spectrum upon multiplication of λM. The zero mass should also be included in the mass spectrum. This can be seen by setting $m_4 = 0$ in Eqs. (2.6) and (2.7) .

Let's also write down the solutions for v_l and v_r explicitly:

$$v_l(s) \;=\; P_F^{\pm\sqrt{F^2-m_f^2}}(s)\,, \tag{2.10a}$$

$$v_r(s) \;=\; P_{F-1}^{\pm\sqrt{F^2-m_f^2}}(s)\,. \tag{2.10b}$$

Eq. (2.10) exhibits a curious fact: the left and right handed-fermions are associated with different factors living in the fifth dimension. An explanation of this is: the matrix Γ^5 involved in the previous calculation has two roles: it is both the gamma matrix associated with the fifth dimension and the gamma matrix used to define the left- and right- handedness of a spinor. Via Γ^5, the chirality and the scalar field living in the fifth dimension are connected. Now the scalar field is specified to be the kink solution which distinguishes the positive and negative direction in the fifth dimension. This discrimination translates into the discrimination between the handedness of fermions in the four-dimensional space-time via Γ^5.

We can do the same calculations for the field Ω and obtain solutions that are pairwise matched to those for Ψ. This matching guarantees the fermion-source cancellation in the equation of motion for the scalar field.

2.2 Discussion

Having sketched the derivation of a discrete and bounded fermion mass spectrum, let me identify the key ingredients responsible for such a spectrum. Obviously, the discreteness comes from our requirement that the v_{lr} factors of the fermion wave functions are square-integrable. This means, the "scattering states" in the fifth dimensions are excluded. Why could this be a plausible assumption? To answer this question, it is necessary to have some insight into the fifth dimension. In spite of its similarity to the other three spatial dimensions, the fifth dimension is different — it's an effective dimension which merely represents a new degree of freedom. We may speculate that the fifth dimension is related to some energy and the integration of the square of the $v_{l(r)}$ factor is some sort of physical energy which must be finite. As for the boundedness of the spectrum, further analysis shows that the "vacuum expectation value" of the scalar field ϕ provides an upper bound for the fermion mass, in the same way as the potential barrier does for the eigenvalues of the Hamiltonian of the bounded states in quantum mechanics.

Let me also comment briefly on the fermion-source cancellation in the equation of motion of the scalar field:

$$\partial_a \partial^a \phi = \bar{\Psi}\Psi - \bar{\Omega}\Omega - 2\lambda^2(\phi^2 - M^2)\phi. \tag{2.11}$$

The two fermionic fields were assigned opposite coupling constants to facilitate the fermion-source cancellation. This made it possible for us to obtain a set of classically exact solutions to the equations of motion of the theory. Eq. (2.11) as it stands is a classical equation. Once quantum effects are taken into account, it should be modified slightly. We can modify it in the following way:

$$\partial_a \partial^a \phi + 2\lambda^2(\phi^2 - M^2)\phi = < \bar{\Psi}\Psi - \bar{\Omega}\Omega > . \tag{2.12}$$

The modified equation can be interpreted as follows: the scalar field is a classical background field (i.e. it is an effective field), only the fermionic fields are treated as quantum fields. The modified equation is a semi-classical equation that we hope can capture the essential dynamics of the scalar field. Even if the quantum effects are taken into account, the fermion source cancellation still occurs due to the pairwise matching between the Ψ and Ω modes.

3 The spherically symmetric gauge field of the Lorentz group

In July 2012, scientists at CERN announced the discovery of a new particle. It wasn't immediately said to be the long awaited Higgs particle. But most people assumed it was (as time is passing by, more and more evidence has been collected in favor of a Higgs particle). There was undoubtedly a huge excitement in the community of particle physics. I believe physicists in the other fields also shared this joy. Discussions on the discovery have been going on since then and will continue for a long time. Reactions to this discovery are twofold. On the one hand, many physicists have been expecting the particle for a long time, seeing the discovery as a natural outcome. Their argument is: without the field of Higgs field the standard model theory does not work; the theory is beautiful, so the particle must be there. On the other hand, some physicists do not feel comfortable with the introduction of a unique scalar field by hand. Therefore they are surprised that a bold prediction from such an ugly theory in the end turned out to be correct. Consequently, they believe that there must be something more fundamental behind the Higgs mechanism! Of course, these two opposite opinions are both subjective and none is wiser than the other. Today those holding the first opinion focus on detecting phenomenological discrepancy between experiment and theory, those insisting on the second opinion delicately look for a more natural and "beautiful" explanation for the structure of the standard model theory. I find myself in the group having the second opinion, to me there is at least one hint that implies there might be something more fundamental. In [8] Witten presented a calculation which bears a surprisingly beautiful feature: starting with a pure Yang-Mills theory of $SU(2)$ in a four dimensional space-time, one imposes spherical symmetry on the gauge field and sees that the original theory reduces to one that describes a Higgs-like complex scalar field coupled to the gauge field of $U(1)$, a subgroup of the original $SU(2)$. The most prominent two features are, the derived scalar field receives a double-well potential and the covariant derivative associated to the $U(1)$ gauge field comes about naturally. Later, Bergmann, Forgacs and Manton proved that the occurrence of this surprising phenomenon can be generalized: A theory with a Higgs-like scalar field subjected to some gauge group can be obtained from a pure Yang-Mills theory of a larger gauge group, in higher dimensional spacetime, by imposing some spacetime symmetry [12] [22].

In this chapter, following Forgacs and Manton, the special case of the Lorentz group is investigated. The interest in this group originates from the following consideration: the Higgs field in the electro-weak theory is a complex doublet and has four degrees of freedom. Thus the group to start with must be big enough. Also, the Higgs doublet only couples to the left handed fermion doublets. Since the Lie algebra of the Lorentz group has a natural left-right decomposition. Thus the Lorentz group is a good candidate as the gauge group of a "fundamemtal" theory (of course, one has to be aware that the spin representation of the Lorentz group here corresponds to the physical concept of "isospin"). Due to the same considerations the group $SO(4)$, which is locally isomorphic to $SU(2) \times SU(2)$), is equivalently a good candidate group to start with as the Lorentz group. One might wish to identify the origin of the group as space-time symmetry transformations. Due to this heuristic argument, the Lorentz group has more priority. By gauging the Lorentz group we are led to a theory of gravity. This is beyond the scope of this thesis and we are not going to make speculative comments at this point. In fact, the resulted theory from the pure Yang-Mills theory of the Lorentz group will differ from the Electroweak theory in several ways, it makes little sense to speculate further. I present the calculation, note some theoretically interesting points and hope that it can inspire some further ideas.

3.1 The basic formalism

In this section I will sketch Forgacs and Manton's study very briefly. Also, I will stick to the notations used by them as much as possible while substituting the Lorentz group in the place a generic group is referred to.

The four dimensional space-time has a Minkowsian metric:

$$h_{\mu\nu} = \mathrm{diag}(1, -1, -r^2, -r^2 \sin^2\theta), \tag{3.1}$$

where the spherical coordinate system(r, θ, ϕ) for the spatial section has been chosen. The space-time symmetry considered here is simply the spherical symmetry of the space, corresponding to the $SO(3)$ rotation group. Now let's define what is actually meant when speaking of gauge fields possessing spherical symmetry.

Suppose a symmetry transformation of the space is generated by the vector field ξ^μ. It is an active transformation: a point x is transformed to another point $\bar{x} = x + \epsilon\xi$, where ϵ is an infinitesimal number. This is equivalent to a passive transformation, a coordinate transformation T. The gauge field, A_μ, transforms as a vector field under T and the transformed gauge field is denoted as $A^{'}_\mu(x'^\nu)$ [here, x'^ν denotes the new coordinates of the point x]. Now suppose that due to the transformation T the point x acquires the coordinates $\bar{x}$ had in the old coordinate system, i.e. $x'_\mu = \bar{x}_\mu$. The symmetric gauge field will satisfy the following equation:

$$A^{'}_\mu(x'^\nu) = A_\mu(\bar{x}^\nu). \tag{3.2}$$

This condition translates to the following equation in terms of the symmetry generator ξ^μ

$$L_\xi A_\mu := (\partial_\mu \xi^\rho A_\rho + \xi^\rho \partial_\rho) A_\mu = 0. \tag{3.3}$$

The condition can be relaxed to that the gauge field is symmetric up to an infinitesimal gauge transformation generated by $g = 1 + i\epsilon W^a T^a$, i.e.

$$A_\mu(\bar{x}^\nu) = A^{g'}_\mu(x'^\nu) = [g A_\mu g^{-1} + i(\partial_\mu g) g^{-1}]'. \tag{3.4}$$

Note that on the right hand side of the equation, gauge transformation is done before coordinate transformation. The order can certainly be reversed, there would be no real difference caused. Then the following symmetry equation is obtained:

$$L_\xi A_\mu = \partial_\mu W - [A_\mu, W] := D_\mu W. \tag{3.5}$$

At this point Forgacs and Manton made a further observation: if a set of symmetry generators, ξ^μ_m (for the case of $SO(3)$, m runs through 1, 2 and 3) are considered simultaneously, the corresponding W_m must satisfy the following constraint:

$$L_{\xi_m} W_n - L_{\xi_n} W_m - [W_m, W_n] - f_{mnp} W_p = 0, \tag{3.6}$$

where the Lie derivative of W_n is defined as for a coordinate scalar and f_{mnp} are the structure constants of $SO(3)$.

The whole point of Forgacs and Manton's article is then to solve Eqs. (3.5) and (3.6) using a smart trick which makes use of invariant vector fields on the Lie group manifold.

Since the spacetime symmetry of the gauge field in our context is the rotation group in three space, the generators of the group can be represented by tangent vectors fields on a sphere centered at the origin, with the common coordinate system (θ, ϕ). The other two coordinates of spacetime, namely time and the radius of the sphere, can be, at the moment, ignored. This is because in Eqs. (3.5) and (3.6) the Lie derivatives concern only variation of the quantities on the same sphere.

Now notice that there is a subgroup $SO(2)$ of $SO(3)$ that fixes a point (actually two antipodal points). Via this subgroup, the sphere introduced in the previous paragraph can be identified as the coset space $SO(3)/SO(2)$. This identification reminds of the principle bundle map $SO(3) \hookrightarrow S^2$ with the fiber $SO(2) \cong U(1)$. However, for the purpose of solving the equations, only the local property of this map is concerned. Further details of this identification can be found in Forgacs and Manton's article. They then embedded the coset space into the full space of the group, i.e. they considered Eq. (3.6) on the space of $SO(3)$ with coordinates (χ, θ, ϕ). After this embedding, it can be seen from the definition of the Lie algebra that the Lie derivatives in Eq. (3.6) become normal derivatives with respect to the coordinates of on the $SO(3)$ space. At this point, defining a new quantity:

$$W_{\hat{\alpha}} = \xi^m_{\hat{\alpha}} W_m \tag{3.7}$$

will be of help, where $\hat{\alpha}$ ranges through (χ, θ, ϕ), and $\xi^m_{\hat{\alpha}}$ is the matrix-inverse of the vector fields, generators $\xi^{\hat{\alpha}}_m$, defined on the space of $SO(3)$. It turns out that $W_{\hat{\alpha}}$ is a pure gauge field on the full $SO(3)$ space, which can be brought to zero by a gauge transformation.

The gauge field A_μ is dealt with in a similar manner. For the moment we consider only the components A_θ and A_ϕ, which are also viewed as quantities defined on the coset space. Then adding one trivial component: $A_\chi = 0$, it is seen that Eq. (3.5) can be brought to

$$L_{\xi_m} A_{\hat{\alpha}} = 0, \tag{3.8}$$

by a gauge transformation on the space of the symmetry group, $SO(3)$. Eq. (3.8) looks much simpler and can be solved exactly. However, to obtain a solution to the original symmetry equation, $A_{\hat{\alpha}}$ has to bear such properties that it can be gauge transformed to a case where $A_\chi = 0$ and where neither A_θ nor A_ϕ depends on χ.

The solutions to Eq. (3.8) are the matrix inverse of vector fields $\tilde{\xi}^{\hat{\alpha}}_n$ which generate left translations on $SO(3)$(in comparison the vector fields $\xi^{\hat{\alpha}}_n$ generate right translations) defined via:

$$\tilde{\xi}_{m\hat{\alpha}} \tilde{\xi}^{\hat{\alpha}}_n = \delta_{mn}, \tag{3.9}$$

with

$$\begin{aligned} \tilde{\xi}_{1\hat{\alpha}} =& (0, -\cos\xi, -\sin\theta\sin\xi), \\ \tilde{\xi}_{2\hat{\alpha}} =& (0, -\sin\xi, \sin\theta\cos\xi), \\ \tilde{\xi}_{3\hat{\alpha}} =& (-1, 0, -\cos\theta). \end{aligned} \tag{3.10}$$

Thus the most general solution for $A_{\hat{\alpha}}$ is:

$$A_{\hat{\alpha}} = \Phi_m \tilde{\xi}_{m\hat{\alpha}}, \tag{3.11}$$

where $\Phi_x = \Phi^a_m T^a$ and Φ^a_m are constants on the space $SO(3)$, thus they can only be functions of space-time coordinates t and r. The gauge field A also has components $A_t(t, r)$ and $A_r(t, r)$. Then we consider the condition that there exists a gauge transformation on $SO(3)$ which brings $A_{\hat{\alpha}}$ into a form with $A_\chi = 0$ and neither A_θ nor A_ϕ depending on χ. This condition is equivalent to requiring some field strength components, namely

$F_{t\chi}, F_{r\chi}, F_{\theta\chi}$ and $F_{\phi\chi}$ to vanish. Then Forgacs and Manton were able to simplify the condition to the following:

$$\begin{aligned} \partial_i\Phi_3 - [A_i, \Phi_3] = 0,\ \text{for } i = t \text{ or } r, \\ f_{3mn}\Phi_n + [\Phi_3, \Phi_m] = 0,\ \text{for } n = 1, 2, \text{ or } 3, \end{aligned} \tag{3.12}$$

where the number 3 comes from the third generator of $SO(3)$, which served as the generator of the subgroup $SO(2)$ identified previously. Solving these contraints gives the desired gauge field. It is possible to make a gauge rotation which depends on t and r only, such that $\Phi_3(t, r)$ becomes constant. Then, A_i commutes with Φ_3, which simplifies the constraints further.

3.2 The reduced Yang-Mills theory of the Lorentz group

Forgacs and Manton have found the general solutions for gauge fields with space-time symmetries. In this section I make use of their result to investigate the special case of the Lorentz group. There are 6 group generators: 3 rotations J^1, J^2, J^3 and 3 boosts $\Sigma^1, \Sigma^2, \Sigma^3$. The Lie algebra is defined by the following commutation relations:

$$[J^i, J^j] = i\epsilon^{ijk}J^k,\ \ [\Sigma^i, \Sigma^j] = -i\epsilon^{ijk}J^k \text{ and } [J^i, \Sigma^j] = i\epsilon^{ijk}\Sigma^k, \tag{3.13}$$

where ϵ^{ijk} is the Levi-Civita tensor with $\epsilon^{123} = 1$ and the index i and imaginary unit i should not be mixed, despite the abuse of notations. In the later calculation I will also use the following Weyl representation of the Lie algebra:

$$J^i = \frac{1}{2}\begin{pmatrix} \sigma^i & 0 \\ 0 & \sigma^i \end{pmatrix} \text{ and } \Sigma^i = \frac{-i}{2}\begin{pmatrix} \sigma^i & 0 \\ 0 & -\sigma^i \end{pmatrix} \tag{3.14}$$

and the γ matrices:

$$\gamma^5 = \begin{pmatrix} -\mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix},\ \gamma^0 = \begin{pmatrix} 0 & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix}, \tag{3.15}$$

and

$$\gamma^r = \begin{pmatrix} 0 & \sigma^1 \\ -\sigma^1 & 0 \end{pmatrix},\ \gamma^\theta = \frac{1}{r}\begin{pmatrix} 0 & \sigma^2 \\ -\sigma^2 & 0 \end{pmatrix} \text{ and } \gamma^\phi = \frac{1}{r\sin\theta}\begin{pmatrix} 0 & \sigma^3 \\ -\sigma^3 & 0 \end{pmatrix}, \tag{3.16}$$

where the $\sigma^i s$ are Pauli matrices, $\mathbb{I}$ is the rank two identity matrix and the γ matrices are already scaled according to the metric in polar coordinates.

To solve Eq. (3.12), it is necessary to identify Φ_3 first. In the following calculation the choice is made as $\Phi_3 = J^3$. It is then found that:

$$\begin{aligned} \Phi_1 =&\ \phi_1 J^1 + \phi_2 J^2 + \eta_1\Sigma^1 + \eta_2\Sigma^2 \\ \Phi_2 =&\ \phi_2 J^1 - \phi_1 J^2 + \eta_2\Sigma^1 - \eta_1\Sigma^2 \end{aligned} \tag{3.17}$$

From this and Eq. (3.11) the following can be obtained:

$$\begin{aligned} A_\theta =& (-\phi_1\cos\chi - \phi_2\sin\chi)J^1 + (-\phi_2\cos\chi + \phi_1\sin\chi)J^2 + \\ &+ (-\eta_1\cos\chi - \eta_2\sin\chi)\Sigma^1 + (-\eta_2\cos\chi + \eta_1\sin\chi)\Sigma^2, \\ A_\phi =& \sin\theta[(\phi_2\cos\chi - \phi_1\sin\chi)J^1 + (-\phi_1\cos\chi - \phi_2\sin\chi)J^2 + \\ &+ (\eta_2\cos\chi - \eta_1\sin\chi)\Sigma^1 + (-\eta_1\cos\chi - \eta_1\sin\chi)\Sigma^2], \\ A_\chi =& - J^3. \end{aligned} \tag{3.18}$$

It can be verified that by a gauge rotation $U(\chi,\theta,\phi)=e^{-i\chi J^3}$, according to

$$A'_\mu = UA_\mu U^{-1} + i(\partial_\mu U)U^{-1}, \tag{3.19}$$

the above gauge field components are transformed to:

$$\begin{aligned} A'_\theta =& -\phi_1 J^1 - \phi_2 J^2 - \eta_1 \Sigma^1 - \eta_2 \Sigma^2, \\ A'_\phi =& \sin\theta\left[\phi_2 J^1 - \phi_1 J^2 + \eta_2\Sigma^1 - \eta_1 \Sigma^2\right] - \cos\theta J^3, \\ A'_\chi =& 0. \end{aligned} \tag{3.20}$$

With the choice made for Φ_3, four real scalar fields and two abelian gauge fields are obtained. Only the number of degrees of freedom of the scalar fields is acceptable when these fields are compared to the field content of the Eletroweak theory. A second choice of Φ_3 is Σ^3, but this would result in a similar gauge field A with the same number of degrees of freedom. Thus this possibility will not be investigated further.

Now the Lagrangian density for the spherically symmetric Lorentz gauge field can be computed. However, it must be noted that Weinberg gives a proof in [24] that for non-compact gauge group, there is no gauge invariant Lagrangian density with a positive definite metric on the associated Lie algebra, while such a Lagrangian density always exists for a compact gauge group. For the latter the Lagrangian density typically takes the following form:

$$L = -\frac{1}{4}\mathrm{tr} F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2) \text{ or } L = \epsilon^{\mu\nu\alpha\beta}\mathrm{tr} F_{\mu\nu}F_{\alpha\beta}, \tag{3.21}$$

where the last expression is a topological term of the gauge field configuration, which after integration over four space-time becomes a constant. The first expression, upon Legendre transformation gives the correct Hamiltonian density:

$$H = \frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2). \tag{3.22}$$

What Weinberg's proof then suggests is that for non-compact gauge groups the gauge invariant Lagrangian density will lead to a problematic Hamiltonian (most likely not bounded from below). But let's ignore this difficulty and move on.

The time and radial components of the gauge field are also obtained from Eq. (3.12):

$$A_t = a_t(t,r)J^3 + b_t(t,r)\Sigma^3 \quad \text{and} \quad A_r = a_r(t,r)J^3 + b_r(t,r)\Sigma^3. \tag{3.23}$$

From the components of the gauge field we can compute the field strength. The Lagrangian density is:

$$\begin{aligned} -\frac{1}{4}\mathrm{tr} F_{\mu\nu}F^{\mu\nu} =& -\frac{1}{2}\mathrm{tr}\left[-F_{tr}F_{tr} + \frac{1}{r^2}(-F_{t\theta}F_{t\theta} + F_{r\theta}F_{r\theta}) + \frac{1}{r^2\sin^2\theta}(-F_{t\phi}F_{t\phi} + F_{r\phi}F_{r\phi}) + \right. \\ & \left. + \frac{1}{r^4\sin^2\theta}F_{\theta\phi}F_{\theta\phi}\right]. \end{aligned} \tag{3.24}$$

The terms relevant for calculating the Lagrangian density are listed below:

$$
\begin{aligned}
\mathrm{tr}F_{tr}F_{tr} =&(\partial_t a_r - \partial_r a_t)^2 - (\partial_t b_r - \partial_r b_t)^2, \\
\mathrm{tr}F_{t\theta}F_{t\theta} =&(-\partial_t\phi_1 - a_t\phi_2 + b_t\eta_2)^2 + (-\partial_t\phi_2 + a_t\phi_1 - b_t\eta_1)^2 + \\
&- (\partial_t\eta_1 + a_t\eta_2 + b_t\phi_2)^2 - (-\partial_t\eta_2 + a_t\eta_1 + b_t\phi_1)^2, \\
\mathrm{tr}F_{r\theta}F_{r\theta} =&(-\partial_r\phi_1 - a_r\phi_2 + b_r\eta_2)^2 + (-\partial_r\phi_2 + a_1\phi_1 - b_r\eta_1)^2 + \\
&- (\partial_r\eta_1 + a_r\eta_2 + b_r\phi_2)^2 - (-\partial_r\eta_2 + a_r\eta_1 + b_r\phi_1)^2, \\
\mathrm{tr}F_{\theta\phi}F_{\theta\phi} =&\sin^2\theta\left[(1+\eta_1^2+\eta_2^2-\phi_1^2-\phi_2^2)^2 - 4(\phi_1\eta_1+\phi_2\eta_2)^2\right].
\end{aligned}
\tag{3.25}
$$

as well as

$$
\mathrm{tr}F_{t\phi}F_{t\phi} = \sin^2\theta\ \mathrm{tr}F_{t\theta}F_{t\theta} \quad \text{and} \quad \mathrm{tr}F_{r\phi}F_{r\phi} = \sin^2\theta\ \mathrm{tr}F_{r\theta}F_{r\theta}. \tag{3.26}
$$

Clearly the above expressions exhibit a pattern which indicates that there are two abelian gauge transformations associated with the gauge fields a_μ and b_μ. In the case treated by Witten [8][12] where the Lorentz group would have to be replaced by $SU(2)$, there would be only one gauge field (in the above expressions we would have to set $b_\mu = 0$ and $\eta_i = 0$). In Witten's case, the expression can be put in a more elegant form by defining the following:

$$
\phi = \phi_1 + i\phi_2 \quad \text{and} \quad D_i\phi = \partial_i\phi - ia_i\phi. \tag{3.27}
$$

Then the relevant terms can be rewritten as, for instance:

$$
\begin{aligned}
\mathrm{tr}F_{tr}F_{tr} =&F_{tr}F_{tr}, \\
\mathrm{tr}F_{t\theta}F_{t\theta} =&D_t\phi(D_t\phi)^* = |D_t\phi|^2, \\
\mathrm{tr}F_{\theta\phi}F_{\theta\phi} =&\sin^2\theta(1-|\phi|^2)^2,
\end{aligned}
\tag{3.28}
$$

and the reduced action, after integration over θ and ϕ is performed, corresponds to an abelian Higgs theory in a two dimensional spac-time with constant negative curvature (the original space-time was Euclidean in that case)[8].

The fact that the term $\mathrm{tr}F_{t\theta}F_{t\theta}$ could be written as a covariant derivative term reflects the following coincidence. The adjoint representation of J^3 in the subspace of the $su(2)$ Lie algebra spanned by J^1 and J^2 coincides with the complex number multiplication structure. More explicitly, the coefficient of J^1 can be identified as the real component while that of J^2 as the imaginary one; and the action of J^3 on the subspace has the effect equivalent to multiplying the coefficients by the imaginary unit i. Now to generalize the above definition for covariant derivative of the scalar fields a new vector can be defined:

$$
\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \eta_1 \\ \eta_2 \end{pmatrix} \text{ with the norm defined by } g_{ij} = \mathrm{diag}(1,1,-1,-1). \tag{3.29}
$$

Here a special quadratic form is introduced for computing the norm of the vector formed by the four scalar fields. One could also make η_1 and η_2 pure imaginary so that the special quadratic form becomes unnecessary. However, this alternative notation is not adopted because later later a second quadratic form on the same vector space will be defined and it cannot be replaced by making use of the alternative trick.

These scalar fields denote coefficients of a vector in the subspace of Lorentz algebra spanned by J^1, J^2, Σ^1 and Σ^2. The representation of J^3 and Σ^3 on this subspace are respectively:

$$R(J^3) = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad R(\Sigma^3) = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.30}$$

Corresponding to the gauge transformation $U = e^{i\alpha J^3 + \beta\Sigma^3}$ the covariant derivative of Φ should then be defined as:

$$D_t\Phi = \partial_t - a_t R(J^3) + b_t R(\Sigma^3), \tag{3.31}$$

while the covariant derivative of the scalar fields with respect to r can be defined simply by replacing t by r in the above equation. The gauge transformation is done according to the following rules:

$$\begin{aligned} \Phi \to & e^{\alpha R(J^3) + \beta R(\Sigma^3)}, \\ a_i \to & a_i + \partial_i \alpha, \\ b_i \to & b_i + \partial_i \beta. \end{aligned} \tag{3.32}$$

The component $\mathrm{tr} F_{t\theta} F_{t\theta}$, for example, can be written as: $g_{ij}(D_t\Phi)_i(D_t\Phi)_j$. Now let me introduce a second quadratic form on the vector space with coordinates $(\phi_1, \phi_2, \eta_1, \eta_2)^T$:

$$\hat{g}_{ij} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{3.33}$$

using which the last term in $\mathrm{tr} F_{\theta\phi} F_{\theta\phi}$ can be written as:

$$\hat{g}_{ij}\Phi_i\Phi_j. \tag{3.34}$$

Then the Lagrangian density can be put in the compact form:

$$\begin{aligned} L = & -\frac{1}{4}\mathrm{tr} F_{\mu\nu}F^{\mu\nu} \\ = & -\frac{1}{2}\{-(f^a_{tr})^2 + (f^b_{tr})^2 + \frac{2}{r^2} g_{ij}\left[-D_t\Phi_i D_t\Phi_j + D_r\Phi_i D_r\Phi_j\right] + \\ & + \frac{1}{r^4}\left[(1 - g_{ij}\Phi_i\Phi_j)^2 - 4(\hat{g}_{ij}\Phi_i\Phi_j)^2\right]\}, \end{aligned} \tag{3.35}$$

where $f^a_{tr} = \partial_t a_r - \partial_r a_t$ and $f^b_{tr} = \partial_t b_r - \partial_r b_t$. A salient feature of this Lagrangian density, as is expected, is that it does not depend on the coordinates θ and ϕ. Thus upon integration in the action, the theory becomes a two-dimensional one.

Looking at the quartic term of the Lagrangian density in Eq. (3.35), one is tempted to compare it to the mexican hat potential of Higgs field in the unified Electro-weak theory. However, there is a serious problem with the potential obtained here: it is not bounded from below. Actually this was already predicted from Weinberg's proof mentioned previously. The problem comes from the fact that the generator Σ^3 is anti-Hermitian. Curing this problem for the Lorentz group would be difficult, if not impossible. However, a way around is obvious: instead of the Lorentz group one uses the the group $SO(4)$, which is compact. With this group, a possible reduced theory may have the gauge group $SU(2) \times U(1)$! Moreover, the pseudo-scalar Lagrangian density also exists for this choice of group and it can be used to tell apart the gauge fields belonging to the left and right $SU(2)$ groups.

If one insists on using the Lorentz group, there could be a possibility to cure the above mentioned problem. If the Lorentz group is interpreted as the symmetry group of the space-time, gauging it means gravity is involved. Then the already existing Lagrangian density for gravity, for example the Einstein-Hilbert action, can be made use of. The Hilbert-Einstein action can also be formulated in terms of Yang-Mills gauge fields [25]. However, there is an essential difference between the Yang-Mills action and the Einstein-Hilbert action: the former is quadratic in terms of curvature (field strength), while the latter is of first order in curvature. Thus, a Yang-Mills action can provide a quartic potential of the derived scalar fields, but the Einstein-Hilbert action cannot. What one can do then is to assume that there is a quadratic term of curvature in Einstein-Hilbert action with a small pre-factor, or simply to start with the square of the Ricci curvature $R_{\mu\nu}R^{\mu\nu}$. Investigations in this direction will however not be pursued in this thesis.

Now let me add a bit of discussion on the understanding of the space-time symmetry of the gauge field. For simplicity let me focus on the $U(1)$ gauge theory of electro-magnetism. It is easy to tell that the space-time symmetry of the gauge field implies the same symmetry on the electromagnetic field, hence the same symmetry of the sources of these fields, i.e. the charges and currents. The classical result on solving the Maxwell equations in the presence of sources is that, the electromagnetic potentials receive contributions from electric monopole, electric and magnetic multi-pole structures in the sources [23]. Clearly, the symmetry is getting weaker from monopole to dipole and so on and the contribution is also getting smaller in general as the symmetry weakens. For example, considering the spherical symmetry leaves only the electric monopole contribution in the electromagnetic potentials. It is conceivable that the monopole contribution is mostly of leading order at low energy (of course there are exceptional cases where the monopole contribution to the source vanishes and the dominant contribution comes from e.g. electric dipole). How is this observation in classical physics connected to quantum physics? It is perhaps useful to think in term of the path integrals. Suppose there is quantum state describing a collection of electrons and photons interacting with each other at time t, denoted as $|\psi(t), A(t) >$. The probability to find the system in the state $|\psi(T), A(T) >$ at some later time T is the square of the modulus of $< \psi(T), A(T)|e^{i\int_t^T H(\psi,A)}|\psi(t), A(t) >$. Switching to path integral, this quantity equals the supposition of the amplitudes of all the paths that connect the two states. It is a known fact that the contribution from those solutions to the Maxwell equations (classical paths) dominates the superposition. According to the previous arguments, the configurations with certain symmetry can be used to approximate the classical solutions to the Maxwell equations at low energy. Therefore, in path integral the integration on the $U(1)$ gauge configurations with certain space-time symmetry gives the major contribution to the quantum amplitude at low energy. Of course, the above discussion is only an intuitive understanding. It is far from a proof. It is tempting to hypothesize that a similar conclusion holds true for the non-abelian gauge fields as well. If this hypothesis is true, proving it rigorously could result in useful simplifications in the computation of S-matrix elements.

3.3 A pseudo-scalar term in the Lagrangian density

With the Weyl representation of the Lie algebra of the Lorentz group, Eq. (3.14), one can construct another gauge invariant term which may also appear in the Lagrangian density:

$$L_{pseudo} = \kappa \mathrm{tr} F_{\mu\nu}\gamma^5 F^{\mu\nu}, \tag{3.36}$$

where the subscript, *pseudo*, in the Lagrangian density indicates that it is a pseudo-scalar under Lorentz transformations, because of the appearance of γ^5, κ is a free parameter. In the expression above, γ^5 multiplies with the group generators in the normal manner of matrix multiplication. The gauge invariance comes from the fact that γ_5 commutes with all the generators J^i and Σ^i. As a quick exercise, this Lagrangian density in the case of spherically symmetric gauge field is also calculated.

The relevant terms for calculating this Lagrangian density are:

$$\begin{aligned}
\mathrm{tr} F_{tr}\gamma^5 F_{tr} =& - i\cdot 2(\partial_t a_r - \partial_r a_t)(\partial_t b_r - \partial_r b_t),\\
\mathrm{tr} F_{t\theta}\gamma^5 F_{t\theta} =& - i\cdot 2\left[(\partial_t\phi_1 + a_t\phi_2 - b_t\eta_2)(\partial_t\eta_1 + a_t\eta_2 + b_t\phi_2) + \right.\\
&\left.(-\partial_t\phi_2 + a_t\phi_1 - b_t\eta_1)(-\partial_t\eta_2 + a_t\eta_1 + b_t\phi_1)\right],\\
\mathrm{tr} F_{r\theta}\gamma^5 F_{r\theta} =& - i\cdot 2\left[(\partial_r\phi_1 + a_r\phi_2 - b_r\eta_2)(\partial_r\eta_1 + a_r\eta_2 + b_r\phi_2) + \right.\\
&\left.(-\partial_r\phi_2 + a_r\phi_1 - b_r\eta_1)(-\partial_r\eta_2 + a_r\eta_1 + b_r\phi_1)\right],\\
\mathrm{tr} F_{\theta\phi}\gamma^5 F_{\theta\phi} =& i\cdot 4\sin^2\theta(1 + \eta_1^2 + \eta_2^2 - \phi_1^2 - \phi_2^2)(\eta_1\phi_1 + \eta_2\phi_2),
\end{aligned} \tag{3.37}$$

as well as

$$\mathrm{tr} F_{t\phi}\gamma^5 F_{t\phi} = \sin^2\theta\ \mathrm{tr} F_{t\theta}\gamma^5 F_{t\theta} \quad \text{and} \quad \mathrm{tr} F_{r\phi}\gamma^5 F_{r\phi} = \sin^2\theta\ \mathrm{tr} F_{r\theta}\gamma^5 F_{r\theta}. \tag{3.38}$$

With the previously defined notations the pseudo-scalar Lagrangian density can be rewritten as:

$$\begin{aligned}
L_{pseudo} =& \kappa \mathrm{tr} F_{\mu\nu}\gamma^5 F^{\mu}\\
=& 4i\kappa\left[-f^a_{tr} f^b_{tr} + \frac{2}{r^2}\hat{g}_{ij}(-D_t\Phi_i D_t\Phi_j + D_r\Phi_i D_r\Phi_j)\right] +\\
&-\frac{2}{r^4}(1 - g_{ij}\Phi_i\Phi_j)\hat{g}_{lm}\Phi_l\Phi_m.
\end{aligned} \tag{3.39}$$

It is clear that the free parameter κ needs to be purely imaginary so that the Lagrangian density is real. The first term in the square brackets in the equation above deserves some attention. This term says that the the gauge fields from the two gauge groups are interacting. If the equation of motion for either gauge field is derived, it will be seen that its dynamics is influence by the other. In standard model theory, gauge fields interact with each in two ways: first, they can interact indirectly via fermions; second the $SU(2)$ gauge fields interact with electro-magnetic field via the Higgs mechanism, i.e. the gauge bosons $W^{\pm}$ acquire electric charges while becoming massive via the Higgs mechanism. This new term describing the interaction between two gauge fields arising from the pseudo-scalar Lagrangian density of spherically symmetric gauge field is interesting and its further suggestion is perhaps worth to investigate.

This pseudo-scalar Lagrangian density displays another peculiarity in the Yang-Mills theory with gauge group $SU(2)\times SU(2)$. First, the Yang-Mills action of this theory can be written down as the sum of that for the two $SU(2)s$. One notices the fact that $so(4) \cong su(2)_L \oplus su(2)_R$ and that $SO(4)$ admits a spin representation which allows a similar term as in Eq. (3.36). Then this pseudo-scalar term is added to the Yang-Mills action obtained previously and a new action is obtained. This procedure is explicitly carried out in the following text.

The Lie algebra of $SO(4)$ and its spin representation can be obtained with a slight modification from Eqs. (3.13) and (3.14). In the following expressions of $so(4)$ elements a prime is added to indicate the difference from those of $so(1,3)$.

The Lie algebra $so(4)$ is defined by the following commutation relationships:

$$[J'^i, J'^j] = i\epsilon^{ijk}J'^k,\ \ [\Sigma'^i, \Sigma'^j] = i\epsilon^{ijk}J'^k \text{ and } [J'^i, \Sigma'^j] = i\epsilon^{ijk}\Sigma'^k. \tag{3.40}$$

The Weyl representation of this algebra is:

$$J'^i = \frac{1}{2}\begin{pmatrix} \sigma^i & 0 \\ 0 & \sigma^i \end{pmatrix} \text{ and } \Sigma'^i = \frac{1}{2}\begin{pmatrix} \sigma^i & 0 \\ 0 & -\sigma^i \end{pmatrix}. \tag{3.41}$$

The decomposition of this algebra into $su(2) \oplus su(2)$ is obtained by defining the following new basis:

$$T^i = \frac{1}{2}\left(J'^i + \Sigma'^i\right), \quad S^i = \frac{1}{2}\left(J'^i - \Sigma'^i\right). \tag{3.42}$$

Then it is straightforward to calculate the following commutators:

$$[T^i, T^j] = i\epsilon^{ijk}T^k,\ \ [S^i, S^j] = i\epsilon^{ijk}S^k \text{ and } [T^i, S^j] = 0. \tag{3.43}$$

These commutators suggest that T^i can be identified as generators of $SU(2)_L$ and S^i of $SU(2)_R$. Now the gauge fields belonging to $SU(2)_L$ and $SU(2)_R$ can be written as:

$$A_{\mu L} = A^i_{\mu L}T^i \text{ and } A_{\mu R} = A^i_{\mu R}S^i. \tag{3.44}$$

The gauge field of $SO(4)$ can be simply identified as the sum of the above two fields:

$$A'_\mu = A_{\mu L} + A_{\mu R}. \tag{3.45}$$

For the field strength a similar relation holds:

$$\begin{aligned}
F'_{\mu\nu} =& \partial_\mu A'_\nu - \partial_\nu A'_\mu + \frac{i}{g}[A'_\mu, A'_\nu] \\
=& \partial_\mu\left(A_{\nu L} + A_{\nu R}\right) - \partial_\nu\left(A_{\mu L} + A_{\mu R}\right) + \frac{i}{g}[A_{\mu L} + A_{\mu R}, A_{\nu L} + A_{\nu R}] \\
=& \partial_\mu A_{\nu L} - \partial_\nu A_{\mu L} + \frac{i}{g}[A_{\mu L}, A_{\nu L}] + L \leftrightarrow R \\
=& F_{\mu\nu L} + F_{\mu\nu R},
\end{aligned} \tag{3.46}$$

where the fact $[T^i, S^j] = 0$ has been used in deriving the third equation, the coupling constants for left and right gauge fields have been assumed to be both g.

The pseudo-scalar term can be computed explicitly now:

$$\begin{aligned}
\mathrm{tr}F'_{\mu\nu}\gamma^5 F'^{\mu\nu} =& \mathrm{tr}(F_{\mu\nu L} + F_{\mu\nu R})\gamma^5(F^{\mu\nu}_L + F^{\mu\nu}_R) \\
=& \mathrm{tr}F_{\mu\nu L}F^{\mu\nu}_L - \mathrm{tr}F_{\mu\nu R}F^{\mu\nu}_R.
\end{aligned} \tag{3.47}$$

Eq. (3.47) shows clearly the effect of the pseudo-scalar Lagrangian density: it discriminates the gauge fields belonging to the left and right $SU(2)$ sub-groups of $SO(4)$ by altering the weights of their contribution to the total action. In the extreme case, the contribution from either the left or the right subgroup gauge field can be totally removed, leaving a Lagrangian density for a chiral theory.

At the end let me add a general remark. For any spin group $Spin(2n)$, a representation can be constructed via the Clifford algebra [26]. In this construction, there is a special matrix γ^{2n+1} playing the same role as γ^5 in the four dimensional case. Therefore, a pseudo-scalar gauge invariant Lagrangian density term can be written down.

4 Some identities of the correlation functions

In this chapter I will present a method to derive a set of identities among the correlation functions. This method makes use of the generating functional in path integral formulation. Its prominent feature is non-perturbative. However, it does not help to determine any specific correlation function.

The method exploits the polynomial structures of the Lagrangian densities. It is seen that, in almost all well-studied theories, the Lagrangian density is a polynomial function of the fields and their first order derivatives, where the field can be a scalar, a spinor, or a vector. Despite that the discussion will focus on a scalar field theory, the result can be easily extended to theories consisting of spinor and vector fields. One of the important equations obtained is Eq. (4.11), a variation equation satisfied by the generating functional of $\lambda\phi^4$- theory. It should be noted that a similar equation in Euclidean space has been derived by Witten using a mathematically more rigorous method [27]. Witten called this equation another type of Ward-Takahashi identity. Indeed, the technical details in the derivation of the Ward-Takahashi identies and the one presented in this chapter are quite similar.

This chapter is outlined in the following way: first, a simple illustration of the idea is given in term of one variable integral; then the method will be demonstrated through the application to the $\lambda\phi^4$ scalar theory and the result will be tested using perturbative calculations; after that theories with fermion fields and gauge fields will be considered. When the number of fields in the theory is increased, the complexity in the expression of the identities of correlation functions will also increase drastically. Checking the identities of the theories with spinor and vector fields will therefore be spared in this chapter.

After completing this chapter, I have been informed that these relations of correlation functions are Dyson-Schwinger identities.

4.1 One variable integral case

The method used for the deriving the main results in this chapter can be best illustrated in term of a single variable integral:

$$\int_{-\infty}^{\infty} dx e^{-i(ax^2+bx^3+cx^4)}, \tag{4.1}$$

where a, b, and c are constant real numbers while $c \neq 0$. To make this integral well-defined, we can introduce a regulator, $-\epsilon x^2$, in the the exponential. $-\epsilon$ is assumed to be a small positive real number, which eventually will be set to zero. This regulator is sufficient to make all the integrals in the following text well-defined. Now define

$$F(\alpha) = \int_{-\infty}^{\infty} dx e^{-i(ax^2+bx^3+cx^4+\alpha x)-\epsilon x^2} \tag{4.2}$$

for a real number α. Noting that

$$i\frac{d}{d\alpha}F(\alpha) = \int_{-\infty}^{\infty} dx x e^{-i(ax^2+bx^3+cx^4+\alpha x)}. \tag{4.3}$$

It is straightforward to derive the following equations:

$$[4c(i\frac{d}{d\alpha})^3 + 3b(i\frac{d}{d\alpha})^2 + (2a - i\epsilon)i\frac{d}{d\alpha} + \alpha]F(\alpha)$$

$$= \int_{x=-\infty}^{x=\infty} d(ax^2 + bx^3 + cx^4 + \alpha x - i\epsilon x^2)e^{-i(ax^2+bx^3+cx^4+\alpha x)-\epsilon x^2}$$

$$\equiv \int d[y(x)]e^{-iy(x)}, \tag{4.4}$$

where the domain of the new variable, $y(x)$, is $(\infty - i\infty, \infty - i\infty)$ $[(-\infty - i\infty, -\infty - i\infty)]$ when c is positive [negative]. Suppose $c = 0$, $y(x)$ will vary in $(-\infty - i\infty, \infty - i\infty)$ $[(-\infty - i\infty, \infty - i\infty)]$ if b is positive [negative]. In general, the situation will be the same as in the previous two cases depending on the order of the highest power of x in y is even or odd. Note that the new variable y has a non-vanishing imaginary part in the domain of the integration. This imaginary part makes the last integral well-defined. From now on, we will suppress the regulator in all expressions.

The last integral in the above equation vanishes. In case of even-order highest power as discussed above, the integration domain starts and ends with the same value ($+\infty$ or $-\infty$), thus the integral is zero. In case the order of the highest power is odd, the integral is equal to the Dirac delta function evaluated at 1. Its value is also zero. This integral is the same as a Fourier transformation of a constant function. The mathematical rigor should follow from that of Fourier transformation but will not be pursued here. The following equation is obtained:

$$[4c(i\frac{d}{d\alpha})^3 + 3b(i\frac{d}{d\alpha})^2 + 2ai\frac{d}{d\alpha} + \alpha]F(\alpha) = 0. \tag{4.5}$$

Of course, this equation cannot be used to determine $F(\alpha)$. However, the specific value of $F(\alpha)$ is of no interest here. The interesting quantities are the coefficients of the powers of α in the series expansion of $F(\alpha)$, assuming that F is a smooth function of α in a small interval around zero. Expressing $F(\alpha)$ in terms of Taylor expansion, the differential equation gives the iterative relations of the coefficients. It is not hard to imagine, when $F(\alpha)$ is replaced by the generating functional in the path integral formalism, the coefficients will be replaced by the correlation functions. Therefore, the counterpart of the iterative relations will be some identities among the correlation functions. Next, this method will be applied to the $\lambda\phi^4$ theory.

4.2 Application to scalar theory

The Lagrangian density of the $\lambda\phi^4$ theory is

$$L = -\frac{1}{2}\phi\partial_\mu\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4. \tag{4.6}$$

The generating functional of this theory is:

$$Z(J) = \int \mathcal{D}\phi e^{i\int d^4x[L+J(x)\phi(x)]}, \tag{4.7}$$

where $J(x)$ is the source field. The variation of $Z(J)$ with respect to $J(x)$ is given by:

$$\frac{\delta Z(J)}{\delta J(x)} = \int \mathcal{D}\phi \;\; i\phi(x)e^{i\int d^4x[L+J(x)\phi(x)]}. \tag{4.8}$$

Similarly, the n-th order variation gives:

$$\frac{\delta^n Z(J)}{\delta J^n(x)} = \int \mathcal{D}\phi[i\phi(x)]^n e^{i\int d^4x[L+J(x)\phi(x)]}. \tag{4.9}$$

One property of these expressions that will be used later is that, any local operator applied to them can be moved into the path integral:

$$\mathcal{O}_x \frac{\delta^n Z(J)}{\delta J^n(x)} = \int \mathcal{D}\phi[\mathcal{O}_x(i\phi(x))^n]e^{i\int d^4x[L+J(x)\phi(x)]}. \tag{4.10}$$

With this preparation it is straightforward to perform an integration by part to obtain the following equation:

$$+\frac{\lambda}{3!}\frac{\delta^3 Z(J)}{\delta J^3(x)} - [\partial_\mu\partial^\mu + m^2]\frac{\delta Z(J)}{\delta J(x)} + iJ(x)Z(J) = 0. \tag{4.11}$$

The above equation could be understood as a partial differential equation for $Z(J)$ as a function of an infinite number of variables with each space time point x corresponding to one variable. This interpretation is consistent to that the integration over the field $\phi(x)$ is understood as over an infinite number of variables, $\phi(x)s$, for each x. As a result, the operator $\partial_\mu\partial^\mu + m^2$ can be interpreted as a matrix with an infinite rank.

In the last equation above if λ is taken to be zero, which corresponds to the free scalar theory, the equation can be solved in a straightforward way:

$$\frac{\delta Z(J)}{Z(J)} = i[\partial_\mu\partial^\mu + m^2]^{-1}J(x)\delta J(x) \tag{4.12}$$

and therefore

$$Z(J) = Z(0)e^{i\int d^4x \frac{1}{2}J(x)[\partial_\mu\partial^\mu + m^2]^{-1}J(x)}, \tag{4.13}$$

which agrees with the result obtained by completing the square for $\phi(x)$ in the Lagrangian density and then evaluating an infinite number of Gaussian integrals.

Eq. (4.11) is not solvable explicitly when $\lambda \neq 0$. Nevertheless, it contains rich information about the $n-$point correlation functions. Previously it was mentioned that $Z(J)$ should be interpreted as a function of an infinite number of variables, in the sense explained there. Assume that $Z(J)$ is a smooth function of these variables in the neighborhood of zero, up to a numerical factor which is infinite. This factor has been encountered in perturbation calculations. It corresponds to the sum of all the disconnected graphs. Of course this argument does not serve as an adequate justification. In fact, Eq. (4.13) provides a verification for this point.

Now a smooth function can be written in Taylor expansion form:

$$Z(J) = Z_0(0) + \sum_i Z_1^i(0)J_i + \frac{1}{2!}\sum_{i,j} Z_2^{ij}(0)J_iJ_j + \frac{1}{3!}Z_3^{ijk}(0)J_iJ_jJ_k + ..., \tag{4.14}$$

where, in the expression $Z_l^{ijk...}(0)$, the lower index l indicates the order of the term in the expansion, which is the number of external points in a correlation function, the upper indices, $i, j, k...$, are short notations of the points in the space-time (suppose the space-time has been discretized and each discrete cell is assigned a unique integer number), and the 0 in the parentheses means the expansion is done around the source configuration $J(x) = 0$.

In a more usual way, with continuous variables, the functional $Z(J)$ can be written as follows:

$$Z(J) = Z_0(0) + \frac{1}{1!}\int dx_1 Z_1(0)(x_1)J(x_1) + \frac{1}{2!}\int dx_1dx_2 Z_2(0)(x_1,x_2)J(x_1)J(x_2) + ... \tag{4.15}$$

where the zero in the parentheses again means the expansion is around $J(x) = 0$ (in the later text the zero will be omitted), and the $x_i s$ are the space-time coordinates variables (note that each x_i denotes a space-time point, instead of a coordinate component). The term, Z_0, is the value of the original path integral. The ratios between the $Z_n(x_1, ..., x_n)$ and Z_0 have physical meaning. Now let's connect them to the n-point correlation functions, which is defined as:

$$G_n(x_1, x_2, ..., x_n) = \frac{1}{Z_0} \frac{\delta^n Z(J)}{\delta J(x_1)...\delta J(x_n)}. \tag{4.16}$$

It is easy to see that in Eq. (4.15) one can make the coefficients $Z_n(x_1, x_2, ..., x_n)$ totally symmetric under the permutation of the variables $x_1, x_2, ..., x_n$. This option makes use of the bosonic nature of the source field $J(x)$ (in case of spinor field theory, the source field would be fermionic and the coefficient Z_n could be made totally anti-symmetric). Then it is straightforward to check that the n-point Green functions coincide with the coefficients divided by Z_0, i.e.

$$G_n(x_1, x_2, ..., x_n) = Z_n(x_1, x_2, ..., x_n)/Z_0. \tag{4.17}$$

Careful readers might have noticed that in the above expression uncountably many points in a continuous space-time are represented by the countable many integer numbers. This is indeed ad-hoc. One explanation is, $Z(J)$ is in general an integral of J over the domain of the space-time. Now taking $J(x)$ to be a smooth function, the integral can be approximated by a sum with $J(x)$ represented by values at a discrete set of points. This is then in the same spirit as of field theory on a lattice. An alternative approach could be taken to avoid this problem. One can constrain the space-time to be compact with periodic boundary condition imposed on the fields. Now going to the Fourier modes of the fields, the spectrum is discrete, one then deals with countable infinite variables.

Substituting Eq. (4.15) into Eq. (4.11) and comparing the powers, one then obtains a set of identities relating correlation functions with different numbers of external points. To get rid of the integrals in the identities it is necessary to make the integrand totally symmetric under permutation of variables and make use of the following fact:

$$\int dx_1 dx_2 ... dx_n f(x_1, x_2, ..., x_n) J(x_1) J(x_2) ... J(x_n) = 0 \tag{4.18}$$

holds true for arbitrary function $J(x)$ implies

$$f(x_1, x_2, ..., x_n) = 0. \tag{4.19}$$

This can be shown by making a series of special choices for the function $J(x)$ in terms of simple combinations of Dirac delta functions. The strategy can be illustrated with the case of three variables. In the first step, one takes $J(x) = \delta(x - x_o)$ to conclude that $f(x, x, x) = 0$; in the second step one sets $J(x) = \delta(x - x_o) + \delta(x - y_o)$ and then $J(x) = \delta(x - x_o) - \delta(x - y_o)$ to obtain $f(x, x, y) = 0$; lastly, one uses $J(x) = \delta(x - x_o) + \delta(x - y_o) + \delta(x - z_o)$ to obtain $f(x, y, z) = 0$.

Since Eq. (4.11) is linear, one can divide both sides by $Z_0(0)$. Then the identities will involve only the normalized correlation functions. In the following $Z_0(0)$ will be set to one and all the correlation functions will be understood as normalized ones. The results are listed below:

0-th order in J :

$$\frac{\lambda}{3!} G_3(x, x, x) - (\partial_\mu \partial^\mu + m^2) G_1(x) = 0, \tag{4.20a}$$

1-st order :

$$\frac{\lambda}{3!}G_4(x,y,y,y) - (\partial_{y^\mu}\partial^{y^\mu} + m^2)G_2(x,y) + i\delta(x-y) = 0, \quad (4.20b)$$

n-th order :

$$\frac{\lambda}{3!}G_{n+3}(x_1,x_2,...x_n,y,y,y) - (\partial_{y^\mu}\partial^{y^\mu} + m^2)G_{n+1}(x_1,x_2...x_n,y)$$

$$+i\sum_{k=1}^{n} G_{n-1}(x_1,x_2...\hat{x}_k,x_n)\delta(x_k-y) = 0, \quad (4.20c)$$

where in the last equation the hat above x indicates that the corresponding variable x_k is skipped. Due to the symmetry $\phi \leftrightarrow -\phi$ of the Lagrangian density, the odd order correlation functions are all vanishing, i.e. $G_{2n+1} = 0$ for $n = 0, 1, 2...$ Only the identities of even order correlation functions are non-trivial. At this point some comments are worth noting. The identities are obtained using a non-perturbative method. Neither the concept of regularization nor of renormalization has been encountered. Also, we do not have to assume that the original Lagrangian density is "bare". Instead, we can take everything in the identities to be physical. This is to say, the mass and coupling parameters, and the correlation functions are all physical quantities, hence finite. If we wish to apply these identities as constraints to the correlation functions obtained using the perturbative method, we would have to substitute some finite quantities obtained from the perturbative calculation into the identities. Therefore, only the renormalized correlation functions could be used. In the following text, a test of the identity involving the two-point and 4-point correlation functions, Eq. (4.20b), is presented.

The first thing to do is to express the correlation functions in momentum space, using the following identity:

$$G_n(x_1,x_2,...,x_n) = \int \frac{dp_1...dp_n}{(2\pi)^{4n}} e^{i(p_1\cdot x_1+...+p_n\cdot x_n)}\hat{G}_n(p_1,...,p_n)(2\pi)^4\delta(p_1+...p_n). \quad (4.21)$$

where $\hat{G}$ denotes a Green function in momentum space. With the aid of this identity, after getting rid of integration over one momentum variable by Fourier transformation, Eq. (4.20b) can be brought into the following form:

$$\frac{\lambda}{3!}\int \frac{dp_1dp_2}{(2\pi)^8}\hat{G}_4(-p,p_1,p_2,p-p_1-p_2) - (-p^2+m^2)\hat{G}_2(p,-p) + i = 0. \quad (4.22)$$

As stated in the discussion following Eq. (4.20b), the four- and two-point correlation functions here should be understood as renormalized. These two quantities have been computed explicitly in many text books. We take the results from Cheng and Li [28]. To one-loop level, the renormalized 2-point Green function is:

$$\hat{G}_{2R}(p,-p) = \frac{-i}{p^2 - m^2 - \tilde{\Sigma}(p^2) + i\epsilon} := i\Delta_R(p), \quad (4.23)$$

where $\tilde{\Sigma}(p^2)$ is the finite part of the self-energy term in the Talyor expansion of the self-energy in powers of p^2. To one loop level, order-λ, $\tilde{\Sigma}(p^2)$ is zero. Note there is an extra minus sign in the propagator stated above compared to that in Cheng and Li's book. This is correct for the signs adopted in the Lagrangian density and path integral used in the present text. The renormalized 4-point correlation function to one-loop level, order-λ^2, in momentum space is:

$$\hat{G}_{4R}(p_1,p_2,p_3,p_4) = \Pi_{j=1}^4(i\Delta_R(p_j))[-i\lambda + \hat{\Gamma}(s) + \hat{\Gamma}(t) + \hat{\Gamma}(u)], \quad (4.24)$$

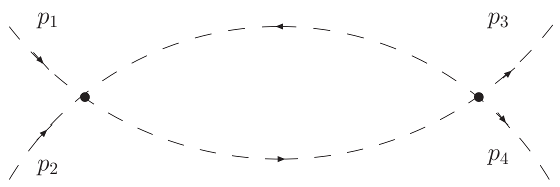

Figure 4.1: The one loop four-point diagram. When the external legs are amputated, this diagram corresponds to the Feynman integral $\hat{\Gamma}(s)$, where $s = (p_1 + p_2)^2$.

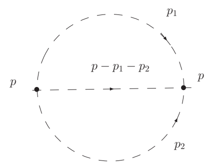

Figure 4.2: The two-loop one particle irreducible diagram. The external legs are amputated. Its contribution to the correction of the propagator is $i\Sigma(p^2)$.

where the letter R in the subscript stands for "Renormalized", and the function $\hat{\Gamma}(s)$ stands for the amplitude of the Feynman diagram in Fig. 4.1 with the Mandelstam variable s.

To the level of one-loop, $\hat{G}_{2R}$ and $\hat{G}_{4R}$ are approximated to different orders of λ. This means that Eq. (4.22) should be checked order by order. To the first order of λ, the term containing $\hat{G}_{4R}$ can be thrown away since it is at least of order two in λ. The rest terms are easily seen to cancel each other. Therefore, to the first order of λ, the identity, Eq. (4.22), is verified.

When we go to the second order of λ in the calculation, the situation becomes complicated. The one-loop computation for $\hat{G}_{4R}$ is still sufficient. Actually, only the first term in the square brackets on the right hand side of Eq. (4.24) is involved. However, for $\hat{G}_{2R}$ 2-loop corrections must be included. The only contributing one particle irreducible diagram (amputated) in this case (Prob. 10.3, page 345 of [1]) is shown in Fig. (4.2). The associated integral is

$$i\Sigma(p^2) = \frac{-\lambda^2}{3!} \int \frac{dp_1 dp_2}{(2\pi)^8} \frac{-i}{p_1^2 - m^2} \cdot \frac{-i}{p_2^2 - m^2} \cdot \frac{-i}{(p - p_1 - p_2)^2 - m^2}, \tag{4.25}$$

where the factor 3! is the symmetry factor associated to this diagram. Again, in the definition of $\Sigma(p^2)$ there is an extra minus sign compared to that in Cheng and Li. Denote the corresponding renormalized quantity as $\tilde{\Sigma}(p^2)$. Substituting it into Eq. (4.23) one obtains a correction to the two point correlation function:

$$\hat{G}_{2R}(p, -p) = \frac{-i}{p^2 - m^2 - \tilde{\Sigma}(p^2) + i\epsilon} = \frac{-i}{p^2 - m^2 + i\epsilon} + \frac{-i}{(p^2 - m^2 + i\epsilon)^2} \tilde{\Sigma}(p^2) + O(\lambda^3) \tag{4.26}$$

The λ^2-order term coming from the integral of $\hat{G}_{4R}$ in Eq. (4.22) is:

$$-i\frac{\lambda^2}{3!} \int \frac{dp_1 dp_2}{(2\pi)^8} \Pi_{j=1}^4 (i\Delta_R(p_j)) \tag{4.27}$$

which is precisely the opposite of the term arising from the non-renormalized 2-loop correction to $\hat{G}_{2R}$:

$$-(-p^2 + m^2) \frac{-i}{(p^2 - m^2 + i\epsilon)^2} \Sigma(p^2) = \frac{i}{-p^2 + m^2} \Sigma(p^2). \tag{4.28}$$

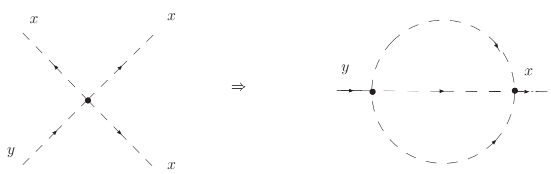

Figure 4.3: The four-point tree diagram becomes an one particle irreducible two-loop diagram when three of the four external points are identical.

Therefore, the corresponding renormalized quantities should cancel each other. Thus to the second order of λ, the identity, Eq.(4.22), holds true as well. The verification involving the higher order terms of λ is getting more and more complicated. Nevertheless, some intuition can be gained via Feynman diagrams. The four-point correlation function in the identity is evaluated at three identical points and one distinct point. In any one particle irreducible Feynman diagram with four external points, if three of them are identified, the diagram becomes one with two external points and two extra loops. The previous calculation for the λ^2 order corresponds to the case where the 4-point tree diagram becomes an one particle irreducible diagram with two external points and two loops, as is illustrated in Fig. (4.3). To be precise, all the external legs in both diagrams should be amputated so that the corresponding Feynman integrals will match with the previous calculations.

4.3 Theories with spinor and vector fields

4.3.1 Application to spinor field

In this section some identities of the correlation functions for theories with spinor fields will be derived. While scalar fields and vector fields satisfy certain commutation relations, the spinor fields satisfy anti-commutation relations. The consequence is that the rules for integration in path integral over these fields are different. First, let us recall a few properties of integration over Grassmanian variables (applicable for spinor fields). The most basic rules are:

$$\int d\eta\ 1 = 0, \quad \int d\eta\ a\eta = a \quad \text{and} \quad \int d\eta \int d\bar{\eta}\ \bar{\eta}\eta = 1, \tag{4.29}$$

where η and $\bar{\eta}$ are two independent real Grassmanian variables, and a is an ordinary real number. For more properties of integral involving Grassmanian variable, the reader is referred to [1]. With these properties the following can be derived:

$$\int d\eta\ \bar{\eta} e^{i\eta\bar{\eta}} = 0. \tag{4.30}$$

In the following derivation let's work with QED. The gauge field (electric-magnetic potential) will be taken as external field. The Lagrangian density can be written as:

$$L_{QED} = \bar{\psi}(x)[i\not{D} + m]\psi(x), \tag{4.31}$$

where ψ is the 4-component Dirac spinor representing the electron-positron field and the co-variant derivative is $D_\mu = \partial_\mu - iA_\mu(x)$. The generating functional of this theory is:

$$Z(\eta, \bar{\eta}) = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi\ e^{i\int d^4x[L_{QED} + \bar{\eta}(x)\psi(x) + \bar{\psi}(x)\eta(x)]}. \tag{4.32}$$

Following the same spirit one can derive:

$$(\not{D} + m)\frac{\delta Z}{\delta \bar{\eta}} + i\eta Z = 0. \tag{4.33}$$

A second equation can also be derived, but it is equivalent to the previous one:

$$(\not{D} + m)\frac{\delta Z}{\delta \eta} - i\bar{\eta} Z = 0. \tag{4.34}$$

where the minus sign is due to switching two Grassmanian variables.

Eq. (4.33) can be solved in a similar manner as was done for Eq. (4.12). First, one obtains from Eq. (4.33) the following equation:

$$\frac{\delta Z}{Z} = -i\delta\bar{\eta}(\not{D} + m)^{-1}\eta. \tag{4.35}$$

Since $\bar{\eta}$ and η are independent, η in the above equation can be viewed as a fixed function. Then the integration with respect to $\bar{\eta}$ can be performed and the result is:

$$Z = const.\ e^{-i\int d^4x\bar{\eta}(\not{D}+m)^{-1}\eta}. \tag{4.36}$$

This result can also be obtained directly by integrating out the $\bar{\psi}$ and ψ fields. In addition, the constant factor in the above equation is known to be $det(\not{D} + m)$. However, in the derivation of the identities among the correlation functions, this factor is again playing no role.

In the standard model theory, terms in the Lagrangian density contain either two spinor fields or no spinor field. This implies that the spinor fields can be easily integrated out. Therefore, the method discussed here won't give better results when it is applied only to the spinor fields.

4.3.2 Application to gauge field

As far as the derivation of the variation equation is concerned, the vector fields (gauge fields) differ from the scalar field only in the sense there are more fields and more indices. This difference makes the derivation more complicated and errors can easily occur. Here, a simpler and more general trick is introduced:

$$\begin{aligned}
&\int \mathcal{D}\phi \left[\frac{\delta L(\phi)}{\delta \phi}\bigg|_{x=x_o} + J(x_o) \right] e^{-i\int dx[L(\phi)+J\phi]} \\
=&\Pi_{x\neq x_o} \int d\phi(x) \int d\phi(x_o) \left[\frac{\delta L(\phi)}{\delta \phi}|_{x=x_o} + J(x_o) \right] e^{-i\int dx[L(\phi)+J\phi]} \\
=&\Pi_{x\neq x_o} \int d\phi(x) \int d\left[\hat{L}(\phi(x_o)) + J(x_o)\phi(x_o) \right] e^{-i[\hat{L}(\phi(x_o))+J(x_o)\phi(x_o)]+L_{rest}} \\
=&0,
\end{aligned} \tag{4.37}$$

where x_o indicates a fixed space-time point, $\hat{L}(\phi(x_o))$ is the part of the Lagrangian density that depends on $\phi(x_o)$ and L_{rest} is independent of $\phi(x_o)$. This trick can be understood in the following way: the field $\phi(x)$ can be represented by an in finite number of variables associated to each space-time point; the differential operators ∂_x can then be represented by a matrix of infinite rank. As a result, the Lagrangian density becomes a polynomial of an infinite number of variables. The terms in the Lagrangian density can be grouped in to two, one with each term depending on $\phi(x_o)$, denoted as $\hat{L}(\phi(x_o))$, and the other totally independent of $\phi(x_o)$, denoted as L_{rest}.

In theories with gauge fields, discretizing the space-time is accompanied with the issue of gauge invariance, just as discussed in lattice gauge theory. However, the matrix representation introduced in the intermediate steps will be replaced again by the differential operators later in the derivation, i.e. the continuum limit will be taken. Therefore, the gauge invariance will be restored. This argument serves as a justification for the trick introduced in the previous paragraph.

Next, in the expression $\frac{\delta L(\phi(x_o))}{\delta\phi}|_{x=x_o} + J(x_o)$ we can replace $\phi(x_o)$ by $\delta/\delta J(x_o)$ and the differential operators (or their matrix representation) can be pulled out of the path integral. As a result, the desired variation equation is obtained. In the following an example will be given for the pure $SU(2)$ gauge theory. The Lagrangian density for this theory is:

$$L_{SU(2)} = -\frac{1}{4}\left(\partial_\mu A^a_v - \partial_\nu A^a_\mu + igf^{abc}A^b_\mu A^c_\nu\right)\left(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu} + igf^{abc}A^{b\mu}A^{c\nu}\right) \quad (4.38)$$

It is then straightforward to obtain

$$\frac{\delta L_{SU(2)}}{\delta A^a_\mu(x_o)} = -\partial_\nu F^{a\mu\nu} - igf^{abc}F^{c\mu\nu}A^b_\nu = \left(-\delta^a_c\partial_\nu - igf^{abc}A^b_\nu\right)F^{c\mu\nu}. \quad (4.39)$$

Replacing the fields by the corresponding variation operators will give:

$$\left(-\delta^a_c\partial_\nu - igf^{abc}A^b_\nu\right)F^{c\mu\nu} \to \left(-\delta^a_c\partial_\nu + gf^{abc}\frac{\delta}{\delta J^{b\nu}}\right)\left(i\partial^\mu\frac{\delta Z(J)}{\delta J^c_\nu} - i\partial^\nu\frac{\delta Z(J)}{\delta J^c_\mu} - igf^{cbe}\frac{\delta^2 Z}{\delta J^b_\mu \delta J^e_\nu}\right). \quad (4.40)$$

Thus the desired differential equation will be

$$i\left(-\delta^a_c\partial_\nu + gf^{abc}\frac{\delta}{\delta J^{b\nu}}\right)\left(\partial^\mu\frac{\delta Z(J)}{\delta J^c_\nu} - \partial^\nu\frac{\delta Z(J)}{\delta J^c_\mu} - gf^{cbe}\frac{\delta^2 Z(J)}{\delta J^b_\mu \delta J^e_\nu}\right) + J^{a\mu}Z(J) = 0. \quad (4.41)$$

Similar to the case of the scalar field, the generating functional can be expanded into the Taylor series of the sources $J^a_\mu(x)$ and the identities among the correlation functions in principle can be obtained. However, this will not be done in this thesis since it involves only tedious but straightforward calculations.

In the above derivations, theories involving only scalar field, spinor fields, or vector fields are considered. The method can of course be applied to cases where all types of fields are present. This direction will not be pursued in this thesis.

Discussion

The work presented in this thesis comprises the main results obtained during my PhD study in the past three years. Four topics were covered, all of them closely related, yet each individual topic could thrive from further extensive research.

In Chapter 1, I proved the existence of a special fermion zero mode, insofar as it is not related to the winding number of the background gauge field. Despite this unique feature, it is only one among many fermion zero modes that are essential for determining the spectral flow. The next step in determining the spectral flow is to study the level crossing behavior at the vicinity of each zero mode. Although this strategy is a practical one, it does not come without technical difficulties. Klinkhamer and Lee observed that in the case of one-dimensional space the phase of the fermion wave function at spatial infinity has a discontinuous change in the presence of a zero mode. They introduced the concept "twist number" as a tool to study this phenomenon[7]. For a given wave function, the twist number is defined as the difference between the phase at infinity and that at the origin. A discontinuous change in the twist number signifies the presence of a fermion zero mode. When the space is open, this phenomenon can be understood in the following way: in the eigenvalue equation of the Hamiltonian for the Dirac fermion interacting with gauge field, we can always set the eigenvalue to zero. There should always be at least one solution to the resulting linear differential equation. Yet such a solution is not necessarily a valid fermion zero mode, in spite of its zero eigenvalue. In order to be a valid zero mode, the solution has to meet a stringent condition: it must be square-integrable in the space. Such solution typically assumes the following exponential form:

$$\psi(x) \sim e^{\int_0^x ds f[A(s),m]} \tag{4.42}$$

where the function f depends on the background gauge field, as well as on other parameters such as mass and charge. The exponential factor will grow or damp as the gauge background field changes. For a solution to be square-integral, the exponential factor must damp, which will only happen when the gauge background fields possess certain properties. In this sense, we can perceive the underlying principle of the "twist number" to be identical to that of Levinson's theorem, in which the number of the square-integrable solutions (i.e. the bounded states) of the Schrödinger equation is related to the asymptotic phase of the solutions [29]. Levinson's theorem also exists for the Dirac equations [30][31]. In both the Schrödinger and the Dirac equations, the phases of the solutions are functionals of the potentials in the equations (be it a generic potential or a specific gauge potential). Aware of this principle, Kaufhold studied the phase change of solutions to the Dirac equation on a compact space and gained deep insight and proved that there is a monotonous dependence of the phase on the energy of the eigen-modes [19]. In one-dimensional space when the periodic (or phase periodic) boundary condition applies, he was able to obtain a concise formula for the spectral flow by tracking the phase change of the eigen-modes.

Generalizing Kaufhold's result to higher dimensional space might seem straightforward. But when we attempt to define a unique phase change for each solution to the eigenvalue equation, a problem arises. In one-dimensional space this can be done by tracking the phase varying along the only non-self-intersecting curve. In two- and three- dimensional space, however, a unique curve does not exist. Alternatively, we might examine the phase flux on the boundary of the space, which is the integral of the divergence of the phase on the bulk space. When the space has no boundaries, the phase flux will be zero and the problem will have to be reconsidered.

Two new problems arise when we attempt to generalize Kaufhold's result to open spaces. First, we must take into account the normalizability of the solutions in correspondence to the boundary condition in compact spaces. In order to determine if the solutions are normalizable, we have to examine their asymptotic behavior, a technically demanding task. This problem is echoed in the index theorem. The Atiyah-Singer and Atiyah-Singer-Patodi index theorems are both obtained for compact space-times. Although an index theorem for open space-times exists,in the form of the Callias index theorem [32], the derivation is made complex by the analysis on the normalizablity of the wave-functions. Second, in the case of open spaces the spectrum of the Dirac-Hamiltonian equation will no longer be discrete. As an alternative, we might be able to define the level crossing unambiguously by making use of the mass gap (the zero eigenvalue is isolated when it occurs). This makes determining the spectral flow more complicated.

In Chapter 2, we looked at a new mechanism for determining the fermion mass spectrum. I wish to address a number of issues related to this mechanism here. The first issue concerns Rubakov's observation of the universality problem that occurs when extra dimensions are introduced into theories where fermions interact with non-abelian gauge fields [11]. The dependence of the fields on the extra dimensions will be integrated out, so that the action of the original theory will reduce to an action of a new theory in the ordinary four-dimensional space-time. The coupling constants in the four-dimensional theory are determined by a few parameters in the original five-dimensional theory. The fermionic fields in the four-dimensional theory are living in certain representations of the gauge group. Therefore, we expect to see that their charges in this representation attain integer values up to some common factor, after integrating out the extra dimensions. It is not yet clear how these integers can be simultaneously obtained from a higher dimensional theory.

We recall that the new mechanism for determining the fermion mass spectrum has a byproduct: the left- and right-handed fermions can be distinguished via a function living in the fifth dimension. This leads us to wonder if the asymmetry between the left and right $SU(2)$ gauge fields can be explained a similar way. We can assume that there is a fundamental theory with two $SU(2)$ acting in a symmetric manner on the left- and right-handed fermion doublets. We can further assume that this theory is defined in a five-dimensional space-time. We might then ask ourselves if we can use a scalar field living in the fifth dimension to break this symmetry, so that only the left $SU(2)$ remains in the reduced four-dimensional theory.

The third issue relates to the mass spectrum. Beyond making the spectrum bounded and discrete, we wish to have some freedom to adjust the values in it. The values we obtained in Chapter 2 are essentially determined by the tangent hyperbolic function. An essentially different function would lead to a numerically different mass spectrum. Unfortunately, we have not found any new functions apart from those derived from trivial variation of the tangent hyperbolic function. According to a general argument there are plenty of them, but it would be nice to have a specific one.

In Chapter 3, we investigated a theory on the gauged Lorentz group. This investigation may serve as a starting point from which future research can be conducted and new ideas can be explored. For one, the Lorentz group could be replaced by the Poincare group, gauging the Poincare group makes more sense than gauging the Lorentz group when we hope to formulate a theory that relates to gravity. Starting with a different space-time could also be an idea worth exploring. Manton started with the space-time $R^{1,3} \times S^2$ and obtained a theory comparable to the Winberg-Salam model [13]. The constraints on the gauge field in Manton's theory come from the spherical symmetry of the two-sphere. Another possible starting point could be the anti-de Sitter space-time.

In Chapter 4, I presented a method for deriving a variational equation for a given quantum field theory. Such an equation typically results in a set of non-perturbative identities of correlation functions in theory. As it is unlikely to be solved, such an equation, by itself, cannot determine the individual correlation functions. We can, however, explore the symmetries of the Lagrangian density and exploit the constraints these symmetries imposed on the correlation functions. These constraints can help simplify the identities of the correlation functions. For example, the reflection symmetry $\phi \to -\phi$ discussed in Chapter 4 following Eq. (4.20c) tells us that all the correlation functions with odd numbers of external points vanish.

By briefly touching upon four different topics, I hope this thesis has given way for new ideas for further research. While we wait for the next exciting experimental discovery to be made, there are quite a few interesting theoretical problems to keep ourselves occupied with.

References

[1] M. E. Peskin, V. Schroeder *An Introduction to Quantum Field Theory* Westview Press (1995)

[2] R. A. Bertlmann *Anomalies in Quantum Field Theory* Clarendon Press, Oxford (1996)

[3] G. 't Hooft *Symmetry Breaking through Bell-Jackiw Anomalies* Phys. Rev. Lett. 37, 8 (1976)

G. 't Hooft *Computation of the quantum effects due to a four-dimensional pseudoparticle* Phys. Rev. D 14, 3432 (1976)

[4] F. Klinkhamer, N. Manton *A saddle-point solution in the Weinberg-Salam theory* Phys. Rev. D 30, 2212 (1984)

[5] E. Farhi, V. V. Khoze, R. Singleton,Jr. *Minkowski space non-abelian classical solutions with non-integer winding number change* Phys. Rev. D 47, 12 (1993)

arXiv:hep-ph/9212239

[6] M. F. Atiyah, V. K. Patodi, I. M. Singer *Spectral asymmetry and Riemannian Geometry I, II, III* Math. Proc. Camb. Phil. Soc. (1975),77,43; (1975), 78, 405; (1976), 79, 71

[7] F. Klinkhamer, Y. J. Lee *Spectral flow of chiral fermions in nondissipative Yang-Mills gauge field backgrounds* Phys. Rev. D 64, 065024 (2001)

arXiv:hep-th/0104096

[8] E. Witten *Some exact solutions multipseudoparticle solutions of classical Yang-Mills theory* Phys. Rev. Lett. 38, 121 (1977)

[9] V. V. Khoze *Fermion number violation in the background of a gauge field in Minkowski space* Nucl. Phys. B 445, 270 (1995)

arXiv:hep-ph/9502342

[10] R. Jackiw, C. Rebbi *Solitions with fermion number 1/2* Phys. Rev. D 13, 3398 (1976)

[11] V. A. Rubakov *Large and infinite extra dimensions* Physics-Uspekhi 44(9) 871-893 (2001)

arXiv:hep-ph/0104152

[12] P. Forgacs, N.S. Manton *Space-Time symmetries in Gauge Theories* Commun. Math. Phys. 72, 15-35 (1980)

[13] N.S. Manton *A new six-dimensional approach to the Weinberg-Salam model* Nucl. Phys. B 158, 141 (1979)

[14] M. Z. Hasan, C. L. Kane *Colloquium: Topological insulators* Rev. Mod. Phys. 82, 3045 (2010)

arXiv:1002.3895

[15] A. Y. Kitaev *Unpaired Majorana fermions in quantum wires* Phys.-Usp. 44 131 (2001)

arXiv:cond-mat/0010440

[16] M. Luescher *SO(4)-Symmetric Solutions of Minkowskian Yang-Mills Field Equations* Phys. Lett. B 70, 3 (1977)

B. Schechter *Yang-Mills theory on the hypertorus* Phy. Rev. D 16, 3015 (1977)

[17] L. Yaffe *Static solutions of SU(2)-Higgs theory* Phys. Rev. D 40, 3463 (1989)

[18] P. Hartman *Ordinary differential equations* Birkhäser, Boston, Basel, Stuttgart 2nd edition (2002)

[19] C. Kaufhold *Untersuchungen zu baryonenzahl- und lorenzsymmetrieverletzenden Prozessen* Dissertation, Univer. Karlsruhe (2007)

[20] F.R. Klinkhamer, L. Yang *Fermions with a bounded and discrete mass spectrum* Phys. Rev. D 91, 045028 (2015)

arXiv:1412.1008

[21] P. A. Dirac *Wave equations in conformal space* Annals of Mathematics Vol. 37, No. 2 (1936)

[22] P. G. Bergmann, E. J. Flaherty, Jr. *Symmetries in gauge theories* J. Math. Phys.19(1), 212 (1978)

[23] J. D. Jackson *Classical Electrodynamics* Wiley, 3rd edition (1998)

[24] S. Weinberg *The quantum theory of fields (Vol.II)* (chapter 15.2) Cambridge University Press (1996)

[25] R. Utiyama *Invariant Theoretical Interpretation of Interaction* Phys. Rev. 101, 1597 (1956)

[26] H. Murayama *Notes on Clifford Algebra and Spin(N) Representations* Lecture notes of Quantum Field Theory, Berkeley (2007)

[27] E. Witten *A New Look At The Path Integral Of Quantum Mechanics*

arXiv.10096032

[28] T.P. Cheng and L.F. Li, *Gauge theory of elementary particle physics* Clarendon Press, Oxford (1988)

[29] N. Levinson *On the uniqueness of the potential in a Schrödinger equation for a given asymptotic phase* København : det Kongelige danske videnskabernes selskab (1949)

[30] A. Calogeracos, N. Dombey *The strong Levinson theorem for the Dirac equation* Phys. Rev. Lett. 93 180405 (2004)

arXiv:quant-ph/0411026

[31] Z. Ma *The Levinson theorem* J. Phys. A: Math. Gen. 39 (2006) R625-R659

[32] C. Callias, *Axial Anomalies and Index Theorems on Open Spaces* Commun. Math. Phys. 62, 213-234 (1978)

R. Bott, R. Seeley *Some remarks on the paper of Callias* Commun. Math. Phys. 62, 235-245 (1978)

A Two continuous functions defined in Chapter 1

The continuity property of the two functions defined in Chapter 1 are to be proven in this appendix. It should be made clear beforehand that I shall use the Weierstrass' definition of continuity of a function in the following text, i.e. a real function $f(x)$ defined at a point x_0 in a real domain is continuous if and only if for any small positive number ϵ there is such a small positive number δ that $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \epsilon$.

Lemma2 : The function $t^*(\Theta_i)$ defined in page 15 is continuous in its domain.

Proof. This statement is true, primarily because the function $\Theta(t, r)$ is continuous at any point (t, r) with finite distance from the origin (Theorem 6.1[18]). The interval of t is irrelevant in this proof, although it was specified in the main text where the lemma was applied.

Suppose $t^*(\Theta_i)$ is defined for Θ_i, denoted as t_0 for simplicity. We can assume without loss of generality that the following holds true: $\Theta(t_0, r = \infty, \Theta_i) = \pi/2$, $\Theta(t_0 + \epsilon, r = \infty, \Theta_i) = \pi$, and $\Theta(t_0 - \epsilon, r = \infty, \Theta_i) = 0$. Now for any small positive number ϵ, there exists such a large enough but finite r_c that $\Theta(t_0+\epsilon, r = r_c, \Theta_i) > \pi/2+2\delta$ and $\Theta(t_0-\epsilon, r = r_c, \Theta_i) < \pi/2-2\delta$, where δ is defined as in the proof of theorem 2. δ serves as the criterion so that as soon as $\pi \geq \Theta(t, r = r_c, -) \geq \pi/2 + \delta$ or $0 \leq \Theta(t, r = r_c, -) \leq \pi/2 - \delta$, we will necessarily have $\Theta(t, r = \infty, -) = \pi$ or $\Theta(t, r = \infty, -) = 0$ respectively, where the hyphen symbol stands for an arbitrary value in a small neighborhood of Θ_i. Note that the choice of δ can be made independent of t and Θ_i.

Consider $\Theta(t_0+\epsilon, r = r_c, \Theta'_i)$. This is a continuous function of Θ'_i. So there exists such a small positive number Δ_1 that $\Theta(t_0+\epsilon, r = r_c, \Theta'_i) > \Theta(t_0+\epsilon, r = r_c, \Theta_i) - \delta > \pi/2+\delta$ for any $\Theta'_i \in (\Theta_i - \Delta_1, \Theta_i + \Delta_1)$. Similarly, there exists such a small positive number Δ_2 that $\Theta(t_0 - \epsilon, r = r_c, \Theta'_i) < \Theta(t_0 - \epsilon, r = r_c, \Theta_i) + \delta < \pi/2 - \delta$ for any $\Theta'_i \in (\Theta_i - \Delta_2, \Theta_i + \Delta_2)$. Denote the smaller one between Δ_1 and Δ_2 as Δ. These facts can be summarized as follows: for any $\Theta'_i \in [\Theta_i - \Delta, \Theta_i + \Delta]$, $\Theta(t_0 + \epsilon, r = r_c, \Theta'_i) \geq \pi/2 + \delta$, thus $\Theta(t_0 + \epsilon, r = \infty, \Theta'_i) = \pi$; $\Theta(t_0 - \epsilon, r = r_c, \Theta'_i) \leq \pi/2 - \delta$, thus $\Theta(t_0 - \epsilon, r = \infty, \Theta'_i) = 0$. Therefore, we conclude that $t^*(\Theta'_i) \in (t_0 - \epsilon, t_0 + \epsilon)$ as long as $\Theta'_i \in (\Theta_i - \Delta, \Theta_i + \Delta)$, or in other words, $t^*(\Theta'_i)$ is continuous at Θ_i. □

We also observed that at $r = 0.0001$, for any $t \in [t_1, t_2]$, there is such a $\Theta^*_i \in [\Theta_{i_1}, \Theta_{i_2}]$ that $\Theta(t, r = 0, \Theta^*_i) = 0$, $\Theta(t, r = 0, \Theta^*_i + \epsilon) = \pi/2$ and $\Theta(t, r = 0, \Theta^*_i - \epsilon) = -\pi/2$ for arbitrarily small ϵ. This Θ^*_i for a given t is unique for the monotonous dependence of $\Theta(t, r = 0, \Theta_i)$ on Θ_i. Thus a function $\Theta^*(t)$ can be defined. By intuition we expect that this function is continuous. A rigorous proof is presented below.

Lemma3 : The function $\Theta^*_i(t)$ defined in page 14 is continuous on the interval $[t_1, t_2]$.

Proof. This proof largely resembles that of lemma 2. At time $t_0 \in [t_1, t_2]$ suppose $\Theta^*(t_0) = \Theta_i$, without loss of generality we can say the following is true: $\Theta(t_0, r = 0, \Theta_i + \epsilon) = \pi/2$ and $\Theta(t_0, r = 0, \Theta_i - \epsilon) = -\pi/2$, for an arbitrarily small positive number ϵ. Thus, there is a small enough r_c (or large negative x, as was introduced in Eq. (1.8), page 13) that $\Theta(t_0, r = r_c, \Theta_i + \epsilon) > 2\delta$ and $\Theta(t_0, r = r_c, \Theta_i - \epsilon) < -2\delta$. Here δ is again such a criterion that if $\Theta(t, r = r_c, -) > \delta$ or $\Theta(t, r = r_c, -) < -\delta$, we must have $\Theta(t, r = 0, -) = \pi/2$ or $\Theta(t, r = 0, -) = -\pi/2$ respectively. The hyphen sign can be replaced by an arbitrary value around Θ_i.

Now let us consider $\Theta(t, r = r_c, \Theta_i + \epsilon)$. As a function of t it is continuous in the interval $[t_1, t_2]$. Thus, there exists such a small positive number Δ_1 that as soon as $t \in (t_0 - \Delta_1, t_0 + \Delta_1)$ it must be true that $\Theta(t, r = r_c, \Theta_i + \epsilon) > \Theta(t_0, r = r_c, \Theta_0 + \epsilon) - \delta > \delta$. Thus $\Theta(t, r = 0, \Theta_0 + \epsilon) = \pi/2$ for $t \in (t_0 - \Delta_1, t_0 + \Delta_1)$. Similarly, there must be such a small positive number Δ_2 that for $t \in (t_0 - \Delta_2, t_0 + \Delta_2)$ it must be that $\Theta(t, r = r_c, \Theta_i - \epsilon) < \Theta(t_0, r = r_c, \Theta_i - \epsilon) + \delta < -\delta$, thus $\Theta(t, r = 0, \Theta_i - \epsilon) = -\pi/2$. If we denote the smaller between Δ_1 and Δ_2 as Δ, the above facts then can be summarized as follows: for $t \in (t_0 - \Delta, t_0 + \Delta)$, $\Theta(t, r = 0, \Theta_i + \epsilon) = \pi/2$ and $\Theta(t, r = 0, \Theta_i - \epsilon) = -\pi/2$. Therefore, $\Theta^*(t) \in (\Theta_i - \epsilon, \Theta_i + \epsilon)$ for $t \in (t_i - \Delta, t_i + \Delta)$, i.e. $\Theta_i^*(t)$ is continuous at t_0. □

Although the lemmas in the appendix are proved for $t^*(\Theta_i)$ and $\Theta_i^*(t)$, it is clear that the exact same statements can be made for $t^*(\phi_i)$ (defined on page 17) and $\phi_i^*(t)$ (page 17). The continuity property of the two functions defined in the main text are to be proven in this appendix. It should be made clear beforehand that I shall use the Weierstrass' definition of continuity of a function in the following text, i.e. a real function $f(x)$ defined at a point x_0 in a real domain is continuous if and only if for any small positive number ϵ there is such a small positive number δ that $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \epsilon$.

Lemma2 : The function $t^*(\Theta_i)$ defined in page 15 is continuous in its domain.

Proof. This statement is true, primarily because the function $\Theta(t, r)$ is continuous at any point (t, r) with finite distance from the origin (Theorem 6.1[18]). The interval of t is irrelevant in this proof, although it was specified in the main text where the lemma was applied.

Suppose $t^*(\Theta_i)$ is defined for Θ_i, denoted as t_0 for simplicity. We can assume without loss of generality that the following holds true: $\Theta(t_0, r = \infty, \Theta_i) = \pi/2$, $\Theta(t_0 + \epsilon, r = \infty, \Theta_i) = \pi$, and $\Theta(t_0 - \epsilon, r = \infty, \Theta_i) = 0$. Now for any small positive number ϵ, there exists such a large enough but finite r_c that $\Theta(t_0 + \epsilon, r = r_c, \Theta_i) > \pi/2 + 2\delta$ and $\Theta(t_0 - \epsilon, r = r_c, \Theta_i) < \pi/2 - 2\delta$, where δ is defined as in the proof of theorem 2. δ serves as the criterion so that as soon as $\pi \geq \Theta(t, r = r_c, -) \geq \pi/2 + \delta$ or $0 \leq \Theta(t, r = r_c, -) \leq \pi/2 - \delta$, we will necessarily have $\Theta(t, r = \infty, -) = \pi$ or $\Theta(t, r = \infty, -) = 0$ respectively, where the hyphen symbol stands for an arbitrary value in a small neighborhood of Θ_i. Note that the choice of δ can be made independent of t and Θ_i.

Consider $\Theta(t_0 + \epsilon, r = r_c, \Theta_i')$. This is a continuous function of Θ_i'. So there exists such a small positive number Δ_1 that $\Theta(t_0 + \epsilon, r = r_c, \Theta_i') > \Theta(t_0 + \epsilon, r = r_c, \Theta_i) - \delta > \pi/2 + \delta$ for any $\Theta_i' \in (\Theta_i - \Delta_1, \Theta_i + \Delta_1)$. Similarly, there exists such a small positive number Δ_2 that $\Theta(t_0 - \epsilon, r = r_c, \Theta_i') < \Theta(t_0 - \epsilon, r = r_c, \Theta_i) + \delta < \pi/2 - \delta$ for any $\Theta_i' \in (\Theta_i - \Delta_2, \Theta_i + \Delta_2)$. Denote the smaller one between Δ_1 and Δ_2 as Δ. These facts can be summarized as follows: for any $\Theta_i' \in [\Theta_i - \Delta, \Theta_i + \Delta]$, $\Theta(t_0 + \epsilon, r = r_c, \Theta_i') \geq \pi/2 + \delta$, thus $\Theta(t_0 + \epsilon, r = \infty, \Theta_i') = \pi$; $\Theta(t_0 - \epsilon, r = r_c, \Theta_i') \leq \pi/2 - \delta$, thus $\Theta(t_0 - \epsilon, r = \infty, \Theta_i') = 0$. Therefore, we conclude that $t^*(\Theta_i') \in (t_0 - \epsilon, t_0 + \epsilon)$ as long as $\Theta_i' \in (\Theta_i - \Delta, \Theta_i + \Delta)$, or in other words, $t^*(\Theta_i')$ is continuous at Θ_i. □

We also observed that at $r = 0.0001$, for any $t \in [t_1, t_2]$, there is such a $\Theta_i^* \in [\Theta_{i_1}, \Theta_{i_2}]$ that $\Theta(t, r = 0, \Theta_i^*) = 0$, $\Theta(t, r = 0, \Theta_i^* + \epsilon) = \pi/2$ and $\Theta(t, r = 0, \Theta_i^* - \epsilon) = -\pi/2$ for arbitrarily small ϵ. This Θ_i^* for a given t is unique for the monotonous dependence of $\Theta(t, r = 0, \Theta_i)$ on Θ_i. Thus a function $\Theta^*(t)$ can be defined. By intuition we expect that this function is continuous. A rigorous proof is presented below.

Lemma3 : The function $\Theta_i^*(t)$ defined in page 14 is continuous on the interval $[t_1, t_2]$.

Proof. This proof largely resembles that of lemma 2. At time $t_0 \in [t_1, t_2]$ suppose $\Theta^*(t_0) = \Theta_i$, without loss of generality we can say the following is true: $\Theta(t_0, r = 0, \Theta_i + \epsilon) = \pi/2$ and $\Theta(t_0, r = 0, \Theta_i - \epsilon) = -\pi/2$, for an arbitrarily small positive number ϵ. Thus, there is a small enough r_c (or large negative x, as was introduced in Eq. (1.8), page 13) that $\Theta(t_0, r = r_c, \Theta_i + \epsilon) > 2\delta$ and $\Theta(t_0, r = r_c, \Theta_i - \epsilon) < -2\delta$. Here δ is again such a criterion that if $\Theta(t, r = r_c, -) > \delta$ or $\Theta(t, r = r_c, -) < -\delta$, we must have $\Theta(t, r = 0, -) = \pi/2$ or $\Theta(t, r = 0, -) = -\pi/2$ respectively. The hyphen sign can be replaced by an arbitrary value around Θ_i.

Now let us consider $\Theta(t, r = r_c, \Theta_i + \epsilon)$. As a function of t it is continuous in the interval $[t_1, t_2]$. Thus, there exists such a small positive number Δ_1 that as soon as $t \in (t_0 - \Delta_1, t_0 + \Delta_1)$ it must be true that $\Theta(t, r = r_c, \Theta_i + \epsilon) > \Theta(t_0, r = r_c, \Theta_0 + \epsilon) - \delta > \delta$. Thus $\Theta(t, r = 0, \Theta_0 + \epsilon) = \pi/2$ for $t \in (t_0 - \Delta_1, t_0 + \Delta_1)$. Similarly, there must be such a small positive number Δ_2 that for $t \in (t_0 - \Delta_2, t_0 + \Delta_2)$ it must be that $\Theta(t, r = r_c, \Theta_i - \epsilon) < \Theta(t_0, r = r_c, \Theta_i - \epsilon) + \delta < -\delta$, thus $\Theta(t, r = 0, \Theta_i - \epsilon) = -\pi/2$. If we denote the smaller between Δ_1 and Δ_2 as Δ, the above facts then can be summarized as follows: for $t \in (t_0 - \Delta, t_0 + \Delta)$, $\Theta(t, r = 0, \Theta_i + \epsilon) = \pi/2$ and $\Theta(t, r = 0, \Theta_i - \epsilon) = -\pi/2$. Therefore, $\Theta^*(t) \in (\Theta_i - \epsilon, \Theta_i + \epsilon)$ for $t \in (t_i - \Delta, t_i + \Delta)$, i.e. $\Theta_i^*(t)$ is continuous at t_0. □

Although the lemmas in the appendix are proved for $t^*(\Theta_i)$ and $\Theta_i^*(t)$, it is clear that the exact same statements can be made for $t^*(\phi_i)$ (defined on page 17) and $\phi_i^*(t)$ (page 17).

Acknowledgements

Now that I am at the end of my PhD program at KIT, I am once again reminded of that time, three years ago, when I contemplated applying for a PhD position.

Those conversations with my friend Peter removed my doubts about the future and encouraged me to focus on what I wanted to do.

My girlfriend Renée's understanding and willingness to cope with the long-hour journeys by train at first, and then learning the new language and having no hagelslag and dropjes, makes me feel sorry for her. I really appreciate the big heart she has.

Three years ago, after a nice conversation on the phone, Professor Klinkhamer decided to offer me the position. In the past three years, Professor Klinkhamer gave me the freedom to explore the area of the topology of gauge fields. This led me to realize that there are so many beautiful aspects about gauge fields that I did not know of, and that I would perhaps never have known of had I not been given this position and this freedom.

During the three years program, I had many interesting discussions about physics problems with colleagues in our small team, as well as in the institutes of ITP, TTP and TKM at KIT. They helped me become aware of my lack of knowledge in particle physics and corrected some wrong thoughts of mine. I will miss those discussions.

When writing this thesis, I received a lot of help on polishing the language. Renée, Pascal, Kumar, Ben and Hamzeh must have gone through terrible pain when proof-reading the draft text. Their corrections have made this thesis more accessible to my readers. Pascal also took the burden to translate the abstract into the German version, the Zusammenfassung. I am grateful to all the help.