Anti-Einstein
or
towards recovery of theoretical physics after the collapse in the last century

by

Jorma Jormakka

Contents:

1. Was Einstein wrong?..................................1
2. Divergence of Green functions....................15
3. Is Special Relativity tested empirically?.......35
4. The EPR problem and Bell's Theorem........53
5. Einstein is wrong, Nordström correct..........71
6. Quantization of gravity............................107
7. Solutions to Yang-Mills equations.............127
8. Annex: calculation of Ricci numbers.........177

1. Was Einstein wrong?

I took six manuscripts on theoretical physics that I have written in the recent years and made a book. None of these manuscripts were ever reviewed by any scientific journal. In fact, it was practically impossible to get anybody to read them. So, who do I imagine will read this book? I do not know. Probably nobody, or maybe somebody who wants to know what is wrong with theoretical physics, what went wrong in the last one hundred years. Because something is wrong. Theoretical physics is a typical taken over field. There are many such, that is, taken over fields. One does not talk much about it, unless it is a case in far away history. Why did I mention Einstein in the title? I do not have anything against Einstein, but now I am on pension and finally took a look at my old physics books and decided to write the study I was planning on doing long ago as a student before deciding to make a Ph.D. thesis on mathematics instead. Then I noticed that there were problems in theoretical physics and the problem usually was in a result by Einstein: both in the Special and in the General Relativity. This is not to say that I did not find problems in quantum mechanics and quantum field theory, which tracked down to Bohr and Feynman. Problems everywhere. That is why I wrote the world *collapse* to the title of this book. It looks like the field collapsed some time in the last century. But I doubt it helps to mention it. I mention it here, but I am fairly certain that this book will be ignored by the physical community. There are only rare times in any field when the field is ready to consider dissident ideas. This is not such a rare time in theoretical physics. But maybe you try to read this book and maybe it gives some ideas, then maybe you don't. Anyway, I did what I could.

1.1 The main results

Chapter 2 starts by looking at the divergence of Green functions in quantum field theories. Green functions are conditional transition probabilities between states and they should give values between zero and one. However, some Green functions do diverge. This happens both in canonical quantization and in the path integral method of quantization. The reason for this divergence is that the momentum space in the inverse Fourier transform from a Feynman propagator to a Green function is integrated over a 4-space while the integrand is a too slowly decreasing function of the momentum. Presently used solutions, renormalization methods, are not well motivated and the chapter proposes that only two ways to remove the divergencies are justified by the physical argument that momentum states are quantized. The first method leads to a minimum wavelength, which implies that the space-time is in a special way discrete. The second method is less attractive in the physical sense, but in a mathematical sense it is fine. The chapter discusses the use of the second method in the CMI Millennium Prize problem of proving that quantum Yang-Mills fields have a mass gap. In fact, it is possible to construct a quantum Yang-Mills field that does not have a mass gap. It is given in Chapter 7. In Chapter 1 this approach is explained in a simple way.

Chapter 3 starts from the suggestion of Chapter 2 that the space-time is discrete. It is not intended that the space is rigidly divided into a grid, but that the space is a sort of fluid with finite size elements. It is shown in Chapter 3 that this discrete model is possible: standard verification experiments for Special Relativity cannot not distinguish between the Minkowski space of Special Relativity and the discrete space. The discrete model has a different velocity addition formula than the Lorentz transform. The formula can be illustrated by the old example that is given when explaining addition of velocities in Special Relativity. Consider running in a train. In Newtonian physics the speed of the train and the running speed add and the total speed of the running person with respect to the ground is the sum of these speeds. In Special Relativity you get the total speed from the Lorentz transform and it is less than the sum. In the discrete model it is not possible to run in the train. If you want to run, you have to jump out of the train and run next to it. The speed of the runner (light) with respect to the ground does not depend on the speed of the train. An observer in the train and an observer on the ground get the same speed for the runner: it is the running speed c. The model gives a velocity addition formula that fully agrees with the Lorentz transform if the speeds that are added have values c. Chapter 3 also includes an argument that directly refutes Special Relativity. The customary way is to calculate the time dilation in the direction orthogonal to the movement. If the time dilation is calculated in the direction of the movement from the Lorentz transform and the time dilation in both direction is set equal, as time is a scalar, then the coefficient of the transform cannot be the one in the Lorentz transform. The coefficient in Lorentz transform is the only one giving the Minkowski space and the result that there is no preferred frame of reference. Thus, there is a preferred frame of reference. This result is not made for the discrete model. It is derived to the Special Relativity model. Thus, Special Relativity is wrong. The preferred frame of reference solves the twin paradox: there is no such paradox. Only one twin is moving.

Chapter 4 shows that Bell's theorem is incorrect. Bell's famous theorem derives a violation of a basic probability identity (!), and measurement experiments seem to support Bell's theorem. So, mathematics does not work in theoretical physics? But I show it is not so. The reason a basic probability identity fails is that detector values are normalized by Bohr's rule. Obviously, basic mathematics must not fail and therefore the normalization of detector values must be changed. After fixing the normalization there is no violation of any mathematical identities. The beginning of Chapter 4 discusses the EPR paradox from a more philosophical point of view and proposes a simulation world solution. In this solution history changes if the other alternative is that some strong rule (like the Heisenberg uncertainty principle) is violated. The simulation world solution allows miracles: the history changes, but these miracles cannot be verified from the history. There can be a miracle, but no miracle can be found by looking at the history because history has changed to lead to the situation where we are now. To illustrate this, remember the movie The Butterfly Effect. In that movie the hero asked a fellow convict to watch his hands, which had no scars, then he changed history, wounded his hands as a child, and for that reason had stigmatization scars in his hands as an adult. The fellow convict saw the scars to appear and believed. In the simulation world solution the hero may have been able to make the scars, but after the scars appeared the fellow convict could not remember that there was a time when the hero did not have scars in his hands. So, this is a nice theory for miracles: totally impossible to prove false. But the actual result in Chapter 4 is that Bell's theorem is false and can be proven false.

Chapter 5 argues that Gunnar Nordström's scalar gravitation theory is the correct one, not Einstein's General Relativity. The chapter takes a new look at the experiments that are supposed to have shown that Nordström's gravitation theory is wrong. None of the tests show that this scalar theory is wrong. Instead, the Schwarzschild solution of General Relativity is found to fail the Shapiro delay test. Of special interest is the movement of Mercury. The chapter calculates that the movement of planets is more complicated than is normally thought. Explicitly it is shown that planets do not move in ellipses (unless the ellipse is a circle) as Kepler claimed. The orbit is not ellipse (probably not even closed) and because of this and other problems one cannot verify if the movment of Mercury does or does not agree with the predictions of General Relativity and if Nordström's theory fails this test. The chapter argues that the energy-stress tensor in Einstein's theory is wrong. Indeed, stress is a force opposing movement and stress forces are mainly caused by electromagnetism. In a pure gravitation theory, like Nordström's scalar gravitation theory, there should not be stress. Only the diagonal elements of the energy-stress tensor should be nonzero. Einstein's theory has the off-diagonal stress elements, but because of that it is a combined gravitation and electromagnetism theory. Chapter 5 gives Ricci entries for a scalar field in Cartesian and spherical coordinates.

Chapter 6 makes an even better case against General Relativity and for Nordström's gravitation theory. If we calculate Ricci entries (they can be found in Chapter 5) for gravitation in a satellite orbit around the Earth, we naturally get Ricci entries that are very well approximated by Ricci entries for a scalar field corresponding to the Newtonian gravitation potential. In other words said: on the Earth and on the close vicinity of the Earth Newtonian gravitation works well. The Ricci scalar curvature R is zero, but the diagonal Ricci elements R_{aa} are not zero. Nordström's theory predicts them correctly. Einstein's General Relativity predicts that the diagonal Ricci entries should be zero like in the Schwarzschild solution. Well, they are not. I think this is a definite nail to the coffin of General Relativity. There is another contribution in Chapter 6. It proposes that Nordström' s gravitation field is simply the Goldstone boson field from the Higgs mechanism of electro-weak interactions. The Goldstone boson is a massless scalar field like Nordström's scalar gravitation field. The Higgs mechanism creates masses and the gravitation field is closely connected with masses. There must be some relation between them.

Chapter 7 is a long and difficult chapter which contains the solution I made for the Clay Millennium Prize problem of Yang-Mills fields ten years ago. Each of the seven Millennium Prize problems took me 4-6 months of hard work, the Poincaré Conjecture much longer from 1986 to 2000 (most of the time took a futile effort to find somebody to read the paper). The Yang-Mills problem took me about six months. That means, the problem was quite difficult and most probably Chapter 7 is difficult to read. Fortunately the solution in Chapter 7 has been checked. A blogger Zulfahmed wrote on one of his posts that he now considers my solution correct, and Simone Farinelli put in 2014 to Arxiv a paper where she quantized my solutions in the axiomatic way. She also must have checked the solution in my paper. We can conclude that the classical solutions for Yang-Mills equations in Chapter 7 are correct. Then the issue is only if the way I quantize the field is acceptable. This quantization part of the paper I remember very well. It is those endless lemmas in Chapter 7 that I cannot remember. I did not want to make it clear that the paper has a solution to the Millennium Prize problem as journal referees are really, really irritated by such claims and refuse to read the paper. (It did not help, no referee could be found to read and comment the paper.) So, the ending of the paper is intentionally a bit obscure, but I explained the idea hopefully in a very clear way in Chapter 1 of this book. It is a very simple idea: Green functions should converge in quantum field theories but they do not and the renormalization methods used are not justified. Thus, there must be a different way to make the Green functions converge. The way I use in Chapter 7 is that I select a set of classical solutions and declare them pure states, as they correspond to standing waves of a suitably constrained universe (which I do not need to construct). Then I impose a renormalization method that keeps these states separate and does not allow state transitions. This yields a quantized Yang-Mills field that has no mass gap.

Annex shows how to calculate Ricci entries R_{ab}. The example derives the Ricci numbers for the Schwarzschild solution to Einstein's equations. In general, one can say that calculations of Ricci entries R_{ab} as in Chapter 5 is rather long and tiresome work.

1.2 The arcgenius Einstein

Some years ago all over the web were news that Albert Einstein was only a plagiarizer. In lack of a better name, I dubbed this theory the Einstein conspiracy and as I for some ten years have been checking the validity of conspiracy theories (just checking, not inventing them, feel safe) I had to look at it. It initially seemed to me to belong to the same group of findings showing that Luis Pasteur was cheating, Isaac Newton was not really a physicist but mostly worked with the Revelation, Martin Luther King was beating prostitutes and controlled by a notorious communist and so on, but actually belongs to that genre of research which points out that the three geniuses, Einstein, Freud and Marx, were not so great after all. One can add a fourth genius to this group: Franz Boas, who started the anthropology school denying the existence of human races, and why not Milton Friedman, the Chicago boys have done quite much harm.

Freud's scientific discoveries, the basis of psychoanalysis, were debunked long time ago: there is no empirical verification of any of the basic concepts of psychoanalysis and no evidence that psychoanalysis helps better than normal conversation with the patient. The same happened with the scientific communism of Marx. The theory of prices as added value is wrong and the program only leads to totalitarism. Boas - albeit a crusader against racism - was shown wrong in more recent years: nature has importance, not only nutrure- It is a long while since his student Margaret Mead was discredited for denying sex differences in personality. And neoliberals? There are every now and then demonstrations objecting to World Bank and IMF meetings, but there are also theoretical arguments against the theory. I must write one day my arguments, already Ricardo is wrong.

Only Einstein still keeps his reputation, and that for sure irritated some conspiracy theoreticians. I guess for some people he was a wrong genius. It did not disturb me. Some people had Sibelius, Nurmi and sauna, some other people had Einstein, Freud and Marx, what is wrong with that? It looked to me that human envy is so great that these conspiracy theoreticians just have bring to dirt everything.

Or that is what I initially thought, many years ago. I did look it up and it was not quite so. There was something real in even this improbable conspiracy theory. It is not that Einstein was not a good theoretical physicist. He worked for 55 years in theoretical physics, published his first article in the age of 22, made Ph.D. at the age of 26, and published in total some 300 scientific papers (not all are scientific articles: the figure includes written reviews, comments and corrections). This is a quite good record, but he is considered as the highest human genius, not just a good researcher. Was he the highest genius? That particular field is full of some level geniuses. I have a book by Murray Gell-Mann, *The Quark and the Jaguar.* The back cover informs that Gell-Mann is the most ingenious physicist after Einstein, and I also have a book by Richard Feynman, *Surely you are joking, Mr. Feynman.* Feynman was also claimed to be the highest genius. And finally I have *Dreams of a final theory* by Steven Weinberg, another most intelligent person in the world. There must be some pattern here, or a cult of a genius. But was Einstein a plagiarizer?

Einstein's Ph.D. thesis contains a new determination of Avogadro's number. The thesis is not remarkable and the history of the thesis reflects rather typical problems in writing a Ph.D. Einstein started as a Ph.D. student of H.F. Weber, but that lasted only for the winter 1900-1901, after which Einstein wanted to change the supervisor and turned to Alfred Kleiner. Einstein submitted a manuscript for a Ph.D. in 1901 to Kleiner, but the work was not satisfactory and Einstein withdrew it in 1902. The manuscript has not survived. It might have been on thermodynamics. Einstein published two papers on thermodynamics in 1902-1903. He had also published a paper in 1901 on the capillary phenomenon. It is difficult to know what the Ph.D. thesis requirements were at that time. Traditionally a Ph.D. thesis has been defined as corresponding to five published articles. In 1903 Einstein chose himself the topic for his Ph.D. thesis. The reviewers were Kleiner and Heinrich Burkhard. The calculations in the thesis were considered impressive and the work was accepted. Burkhard checked the calculations but did not notice a significant error in them, so the thesis was accepted. The only comment was that the thesis was too short. Einstein's new method of calculating Avogadro's number gives a worse estimate than the earlier methods, so the contribution is only in finding a different method. This history is fairly normal for a dissertation, I remember similar cases, but at the same time as making the Ph.D., Einstein wrote four other papers, the Annulus Mirabilis articles, and we have to look at them to find the genius of Einstein.

Einstein's fame is based on two sets of four articles each. The first set was published in 1905, the miraculous year. These four articles deal each with a different topic: explaining the photoelectric effect, explaining the Brownian motion, formulating the theory of special relativity and the last presents the famous formula $E = mc^2$.

That is a nice set of remarkable results. But...

It turns out that the Brownian motion actually was explained mathematically by Marian Smoluchowski, who did not publish his results until July 1906, see [1]. Smoluchowski writes in the 1906 article: *I have not published hitherto the results since I wanted to verify them by the most exact experimental methods. But in the meantime the discussion on this subject was re-opened by two theoretical Einstein's papers, (the papers of 1905 and*

1906)…. In Einstein's formulas I found the part of my findings and his final result, which, though obtained by quite different method, agrees completely with mine. Therefore I publish my argumentation,

In a letter from 1909 to Jean Perrin Smuluchowski tells: *should like to state that the priority is of course due to Einstein (1905), the author whose ingenuity and talent inspire my deep respect. Its my fault that I have delayed until July 1906 the publication of my investigation on this subject, in which I was busy since 1900 (work of Mr. Exner).*

This miraculous article was then not so original, as the problem had already been solved though the solution was not published. In 1910 Einstein wrote an article of the opalecence of liquids, he mentions there an article of Smoluchewski from 1904, see [1]. There is no evidence to claim that Einstein had Smoluchowskis unpublished results before writing the article of 1905, but it is also not possible to prove that he did not.

It is not doubted that Smoluchowski obtained the solution first. The priority goes to Einstein, since he published the results first, that is the normal criteria. Therefore the formula $E = mc^2$ should not be considered as Einstein's result. This formula was known to many others, including Henry Poincaré, but it is also presented in an article published by Olinto De Pretto in 1903 in a scientific magazine and republished in 1904 by Veneto's Royal Science Institute. Certainly these, especially the last one, count as scientific publications. They predate Einstein's article from 1905.

So, that is a clear case that the result should not be Einstein's, and there is a reason to suspect that there may be more to this case. There are still two more miraculous articles.

The third miraculous article explains the photoelectric effect. Einstein received the Nobel Prize in 1921 especially from this article. It was a problem on which one of Einstein's supervisors had worked experimentally for a long time and hoped to get a Nobel Prize for it. That the photoelecric effect could be easily explained by Max Planck's concept of a quantum was probably known to all people working on that problem, but Planck's quantum was not yet an accepted concept. Planck had tried to prove a formula for blackbody radiation with his concept of a quantum, but his proof was unsound, see e.g. Alonso-Finn III [2] after equation (1.8). Therefore there was need for more experiments. In my opinion Einstein's paper is just an application of Planck's quantum. It explains the phenomenon, but by using an unverified concept. Einstein later belonged to Max Planks group of researchers, and after some years Planck got a Nobel Prize for the quantum concept, as he should have. The concept is central to quantum physics and even the blackbody energy density formula has been soundly proven. But if we go back to Einstein's times, the important drive behind Einstein's article on the photoelectric effect and his subsequent Nobel Prize was Planck's effort to get the quantum concept accepted.

Max Planck tried to reform the field of quantum physics with his new concept and he probably helped the people in his group to to have their results accepted. Here the goal is laudable, but this still is a form of a takeover and can lead to acceptance of false theories, as may have happened with Einstein's results. We can compare Planck's case to the cases of the other geniuses that I mentioned from other fields where is it quite common to view the development as a takeover by a small group supporting a special and often incorrect theory.

The clearest the takeover view is with Marxists. They tried to inflitrate the academia and the intellectual elite and get scientific Marxism accepted as the correct and final explanation of history. Yet, Marx's theory of prices is incorrect, as was shown in the practical experiment of the Soviet Union. Friedman's neoliberal theories were likewise pushed through with a planned takeover, especially in South America with the USA educating Chicago boys to lead the economic change to free markets in their own countries. Opinions of neoliberal theories are divided, but I tend to view neoliberalism and globalization as invalidated by practise and theory. Freud build a group of followers. He even had a ring society (he was dealing out rings for members of his group, something a bit esoteric). The group captured a niche in psychology. Boas managed to dominate the field and his students had much influence. I see the group of Planck in a similar way as these other groups, but let us continue with Einstein's papers.

The fourth miraculous article is the special relativity theory. This is the most famous of Einstein's results. Apparently the submitted manuscript had two authors: Einstein and his first wife Mileva Marić. The published version had only Einstein as the author. Probably Marić had made a contribution to the article, but Einstein did not give her due credit. The article builds on recent work (at that time) by Hendrik Lorentz and Henry Poincaré. Einstein directly uses the transformation given by Lorentz as the concept from which almost everything in the special relativity theory follows in a fairly simple way. Einstein drops the absolute time as was suggested by Henry Poincaré in a book that Einstein and Marić were reading. Einstein's article does not even mention Poincaré. As the main ideas behind the special relativity theory derive from Lorentx and Poincaré, Einstein's main contribution was to drop the ether and the absolute time. Poincaré wanted to keep the absolute time as he thought it would be useful in some aspects.

Was Einstein's choise an important contribution to the field? It may have been just the opposite. Though the absolute time is not needed for describing the movement of materia-energy, there may be more to this world than materia-energy. A 4-dimensional real manifold can be smoothly embedded into a 10-dimensional Riemannian space, which can have an absolute time. Dropping the absolute time directed researchers to ignore the possibility that there is more to research. There is the possibility that there is our physical, measurable time in our physical world, but there exists also the real time in the real world.

The second set of Einstein's important articles is from November 1915. In four articles he defines the general relativity theory. There is a controversy concerning the role of David Hilbert. Hilbert invited Einstein to lecture on the problem in June-July 1915. Hilbert worked on the problem since that time. Einstein's friend Marcel Grossmann pointed to Einstein that the key to his problem was the Riemann tensor. This is probably how Hilbert got involved.

On November 4 and 11, 1915 Einstein sent to Hilbert non-covariant and pseudo-Riemannian attempts to the field equations, so Einstein had not solved it yet. Einstein writes in a letter to Hilbert, dated November 18, that Hilberts article from 4. November needs to be reconsidered. Hilbert had submitted an article to a journal. There are galley proofs marked December 6, 1915, of this article. There is missing half a page but most is preserved. The equations in these proofs are non-covariant. I think that the article by Hilbert from 4. November that Einstein mentions in his letter is the submitted article in the galley proofs from December 6.

Assuming that this is the case, let us try to see who found the covariant (correct) equations first. Hilbert had not found them on November 4 and Einstein had not found then on November 4 and 11.

Einstein submitted a manuscript with covariant equations on November 25 and his article was published extremely fast in December 2. This article does not mention Hilbert. Hilbert rewrote the article in March 1916 with identical covariant equations as in Einstein's paper. This article mentions Einstein. It is only a question what happened between November 11 and November 25.

On November 16 and 17 Hilbert sent to Einstein notes from a talk he had given on November 16. On November 18 Einstein replied that Hilbert's equations are equivalent to what he himself had derived in the last four weeks. On November 20 Hilbert lectured on the problem. It follows from these dates and comments that Einstein had a new solution on November 18 and it was the same solution that Hilbert had on November 16. Einstein claims to have found it in the last four weeks, but on November 11 he had not found it. Thus, Hilbert found it first and presented it on November 16 and November 20. Einstein did not independently find it but noticed his problem from the notes Hilbert had sent to him on November 16 and 17. Einstein immediately corrected his paper, submitted it on November 25, and got it published in December 2 without any mention of Hilbert.

This is apparently what happened, so it was foul play from Einstein's side. Einstein apparently thought that Hilbert tried to steal his result. This is to be understood in the following way: Einstein had invented the problem and worked on it, but could not do math well enough. His friend Grossmann formulated the problem as Riemannian tensors. Then jumps in Hilbert, a mathematician, and tries to publish a solution, which is just the one that Einstein was hoping to find. Einstein looks desperately for an error in Hilbert's work and points out to non-covariance. Einstein decides just at that time that the equation should be covariant. Hilbert makes his solution covariant and then Einstein has no other way to save his right to be the sole creator of his theory than to fast write a paper and to publish it through his contacts in seven days while delaying Hilberts paper and making galleys of the original version of Hilbert's paper. Hilbert, already famous, does not much care and gives the arcgenius full priority to the results. Does this sound likely? I think it may be very possible.

Anyway, I would not much care of who invented the General Relativity theory because I think that Gunnar Nordström's scalar gravitation theory, predating General Relativity, is the correct theory of gravity. Certainly Einstein was the main inventor of General Relativity even if Hilbert first formulated the equations.

There was a less important but a bit similar event of foul play in Bose-Einstein statistics. Einstein received a description of a statistical model from Indian physicist Satyendra Nath Bose in 1924 and originally tried to present it as his own. He was stopped by some mathematicians who gave credit to the original inventor.

There is also a case when Einstein refused to add his name to a result. He suggested an experiment to Schrödinger and Schrödinger proposed that Einstein be a co-author to the Schrödinger gas model. Einstein declined, but maybe it is since Schrödinger showed what Einstein should have done, but did not do, with De Pretto, Poincaré, Hilbert, Bose and maybe Smuluchowski and Nordström.

Later scientific work by Einstein was not remarkable. He worked on different cosmological models and on the unified field theory. I found his unified field theory from the library of the theoretical physics department in Helsinki. I apparently was the first to read it through since nobody had cut separate the pages, they were on cheap paper and unified on one side. Needless to say, I could not make any sense of the calculations, just try it yourself. Later I learned from Magueijo's book [3] that Einstein used very obscure notations, which may have been one reason these calculations did not open to me. (Though, I am not sure Magueijo read Einstein's notes on the unified theory or just The Meaning of Relativity.)

What more Einstein did? He drew attention to the Einstein-Podolsky-Rosen paradox. The EPR problem is important and my favorite, but Einstein's explanation is in my opinion incorrect. I think the solution is that the effect goes through the history, implying that history changes, implying that the time is not real, implying that this is a simulation world, implying that there is a real world with a real time, implying that the absolute time should have been kept. That is, EPR does not only partially change the interpretation of quantum physics, it ruins the basis of Einstein's special and general relativity. But Bell's recormulation of EPR paradox is just a pseudo-paradox: there is a hidden channel in that reformulation.

In general one can say that Einstein was a good physicist, not a head taller than the others, but also not a head shorter. He was made a head taller by media and he himself seems to have been quite ambitious to have all results ascribed to him alone.

Nevertheless, in the rest of this book I argue that Einstein was wrong in quite many issues, including Special and General Relativity.

References to Chapter 1:

[1] B. Średniawa, Scientific contacts of Polish physicists with Albert Einstein, Proceedings of the 2nd ICESHS, ed. M. Kokowski, Sep 2006. Available in the web.

[2] Alonso-Finn, *Fundamental University Physics.* Vol III, Addison-Wesley, 1968. The classic physics book in many universties.

[3] J. Magueijo, *Faster than the speed of light.* Penguin books, 2003. A controversial book by yet another genius in physics.

2. Divergence of Green functions

Renormalization is a concept of such importance in quantum field theory that a Lagrangian density which cannot be renormalized is discarded as physically relevant. Yet, renormalization is mathematically unsound as it often means subtracting an infinity from another infinity and getting a finite result. Theoretical physicists readily admit that renormalization is unsound but add that it works well. In [4] p. 97 there is an example of a finestructure constant for anomalous magnetical moment where the measured value and the value calculated from Quantum Electrodynamics (using renormalization) agree to seven digits. However, there seems to be some matching involved at least in the path integral quantization method as [2] p. 90 tells that numerical values obtained for physical quantities are fully determined by the renormalization scheme that is used.

The problem is not that calculations would not work. It is that Green functions, Fourier inverses of Feynman propagators, should be transition probabilities and a probability must always have a value between zero and one. Divergent Feynman diagrams, which these theories produce, give infinite probabilities. Renormalization reduces these probabilities to finite values but in a way that is unsound in the mathematical sense and lacks logical motivation in the physical sense.

Divergence of some Green functions is a real phenomenon and not a result of eny error in calculations. Let us see how and why two standard textbooks end up to these divergent functions and try to find the reason for divergence. The first book by Richard D.Mattuck [1] is from the 1960s and the second edition from 1970s. The book follows the canonical quantization approach as was customary at that time. The book gives an example of a divergent Green function for the Schrödinger equation with a simple driving potential. The second book by David Bailin and Alexander Love [2] is from 1980s and follows the functional path quantization method, which was new at that time and which the authors say is better suited to quantum gauge field theories. The book works out in detail a simple scalar field theory and gets a divergent Green function. Also [3] follows the canonical approach to quantization and can be compared with the path integral quantization in [2]. There is a third approach to quantization, the axiomatic quantization, but I did not find a good simple example of a divergent Green function from that approach. As both canonical and path integral quantization fulfill the axiomatic approach, it does not help in this divergence problem.

Both textbooks [1] and [2] derive a divergent Green function in a way that is (or can be made) mathematically sound (some small errors in [2] must be corrected). The reason for the divergence cannot be anywhere else than in the very final inverse Fourier transform that takes a Feynman propagator in the 4-space p, the momentum space, and transforms it to the 4-space x, the position space. In fact, if the integration were over the 3-space, the integral giving the Green function would be convergent, a Yukawa potential. But this is not the case. In the following calculations I set $hbar = c = 1$ and often work in the Euclidian space in order to simplify expressions. More precise treatment can be found from [1] and [2], it does not change the divergence problem.

2.1 Derivation of a divergent Green function in Mattuck

Mattuck [1] starts his derivation from the Schrödinger equation

$$\left(\frac{1}{2m}\nabla^2 + i\frac{\partial}{\partial t}\right)\Psi(x,t) = 0. \tag{2.1}$$

As the differential operator

$$L = \frac{1}{2m}\nabla^2 + i\frac{\partial}{\partial t} \tag{2.2}$$

is linear we can try the Green function method: in order to solve a linear differential equation

$$L\Psi(x) = f(x)$$

find a function G satisfying

$$LG(x', x) = \delta(x - x').$$

If $f(x)$ is not identically zero, the Green function G gives the solution as

$$\Psi(x) = \int dx' G(x', x) f(x') \tag{2.3}$$

since

$$\begin{aligned} L\Psi(x) = L \int dx' G(x', x) f(x') &= \int dx' LG(x', x) f(x') \\ &= \int dx' \delta(x - x') f(x') = f(x). \end{aligned}$$

In the Schrödinger equation (2.1) the function $f(x)$ is identically zero, but the Green function method can still be useful. The Green function for the operator (2.2) must satisfy

$$LG(x - x', t - t') = \delta(x - x')\delta(t - t').$$

Feynman propagators are Fourier transforms of Green functions from the x-space to the k-space, that is, $G(k, t)$ is a Feynman propagator for the Green function $G(x - x', t - t')$ if

$$G(x - x', t - t') = \int \frac{d^3k}{(2\pi)^3} e^{ik\cdot(x-x')} G(k, t - t').$$

Inserting L and Fourier transforming gives

$$\left(-\frac{k^2}{2m} + i\frac{\partial}{\partial t}\right) G(k, t - t') = \delta(t - t').$$

Notice that k is the discrete state while x is the continuous position. The solution to this equation is the free propagator $G_0(k, t - t')$

$$G = G_0 = -iu(t - t')e^{-\frac{k^2}{2m}(t-t')}$$

where $u(t)$ is the step function defined as $u(t) = 0$ is $t \leq 0$ and $u(t) = 1$ if $t > 0$. The solution can be verified by a direct calculation

$$i\frac{\partial}{\partial t} G_0 = \delta(t - t')e^{-\frac{k^2}{2m}(t-t')} + \frac{k^2}{2m} G_0$$

$$= \delta(t-t') + \frac{k^2}{2m} G_0.$$

The solutions to the Schrödinger equation (2.1) are functions

$$\Psi(r,t) = \phi_k(r) e^{-\frac{k^2}{2m}(t-t')} \quad , \quad \Psi(k,t) = \phi(k) e^{-\frac{k^2}{2m}(t-t')} \tag{2.4}$$

$$\phi_k(r) = \frac{1}{\sqrt{\Omega}} e^{ik\cdot r}$$

here Ω is the space volume.

We see that though the Green function G_0 does not give the solution Ψ by the formula (2.3), it is related to the solution: G_0 is the conditional probability that a particle is in the place x' in the time t' if it was in the place x in the time t.

If there is a driving potential $V(\nabla)$, then the Schrödinger equation is

$$\left(\frac{1}{2m} \nabla^2 + i \frac{\partial}{\partial t} - V(\nabla) \right) \Psi(x,t) = 0. \tag{2.5}$$

The Green function satisfies

$$\left(\frac{1}{2m} k^2 + i \frac{\partial}{\partial t} - V(k) \right) G(k, t-t') = \delta(t-t').$$

In this simple example we can directly Fourier transform the equation and solve it:

$$\left(\frac{1}{2m} k^2 + i\omega - V(k) \right) G(k,\omega) = e^{-i\omega t'}$$

$$G(k,\omega) = -\frac{2m}{k^2 - i2m\omega - 2mV(k)} e^{-i\omega t'}. \tag{2.6}$$

$G(k,\omega)$ is a Feynman propagator in the quantum field theory for (2.5). Inverse Fourier transform of this function gives a Green function. A Green function is a probability but the inverse transform of (2.6) leads to a divergent integral.

If V is more complicated the solution can be written in the form of a series of integrals of G_0, a perturbation series (see Mattuck p. 53):

$$G(k, t-t') = G_0(k, t-t') + \int dt'' G_0(k, t-t') V(k) G_0(k, t''-t') + \cdots$$

$$+\int\int dt''dt'''G_0VG_0VG_0+\cdots$$

The perturbation series has convolution integrals and it is convenient to Fourier transform also the time parameter t to frequency ω. The transform changes convolution integrals to products. In this simple case the resulting series, a simple geometric series, can be summed. Mattuck does explain that the perturbation series often diverges. It may in some cases be an alternating series where reordering the terms can render it convergent. Mattuck mentions an example by Kurkisuonio where this mechanism occures. However, the divergence of the Green function corresponding to (2.6) is not caused by the perturbation series: the series can be summed or the equation solved without a perturbation series. The inverse Fourier transform of (2.6) diverges because in the inverse Fourier transform the integral has a too high dimensional volume differential and the propagator is not decreasing fast enough. Mattuck does not discuss this real problem. The only direct treatment on renormalization he gives is of mass renormalization, which is a logically sound case of renormalization. Mass renormalization does not help with the divergent Green function from (2.6). Let us next look at the derivation of a divergent Green function in Bailin and Love through the path integral method.

2.2 Derivation of a divergent Green function in Bailin and Love

The free-field Green function $\mathcal{G}^{(2)}$ is derived for a simple scalar field in the first part of the book by Bailin and Love [2]. The authors start from the path integral

$$I=\int\mathcal{D}\varphi\int\mathcal{D}\pi\exp\ \ i\int_{t'}^{t''}dt\int d^3x\,(\pi\partial_0\varphi-\mathcal{H}(\pi,\phi)+J\varphi)\,.$$

It is proportional to the state transition probability

$$\langle\varphi'(x),t'|\varphi''(x),t''\rangle\sim I$$

where

$$\hat{\varphi}(x)|\varphi(x),t\rangle = \varphi(x)|\varphi(x),t\rangle \tag{2.7}$$

is the eigenvalue equation for the operator $\hat{\varphi}(x)$. After this start Bailin and Love focus to the ground state to ground state probability

$$W[J] = \int \mathcal{D}\varphi \int \mathcal{D}\pi \;\; \exp \;\; i \int dt \int d^3x \, (\pi\partial_0\varphi - \mathcal{H}(\pi,\phi) + J\varphi)$$

$$= \int \mathcal{D}\varphi \int \mathcal{D}\pi \;\; \exp \;\; i \int dt \int d^3x \, (\mathcal{L}(\pi,\phi) + J\varphi) \, . \tag{2.8}$$

Here $\mathcal{H}$ is the Hamiltonia density, $\mathcal{L}$ is the Lagrangian density and J is an external driving potential. This path integral is a generating function for Green functions

$$W[J] = \sum_{n=0}^{\infty} \frac{i^n}{n!} \int d^4x_1 \cdots \int d^4x_n \mathcal{G}^{(n)}(x_1 \ldots x_n) J(x_1) \cdots J(x_n)$$

where the n-particle Green function is

$$\mathcal{G}^{(n)}(x_1 \ldots x_n) = \langle 0|T(\hat{\varphi}(x_1) \cdots \hat{\varphi}(x_n)|0\rangle$$

and T is the time ordering operator (that is, the n events come in a certain given order in time). We can obtain the Green functions as

$$\mathcal{G}^{(n)}(x_1 \ldots x_n) = \frac{\delta^n W[J]}{\delta J(x_1) \cdots \delta J(x_n)}_{|J=0} \, .$$

The difference with this path integral method to canonical quantization is that the integrand

$$\mathcal{L}(\varphi, \partial_\mu\varphi) + J\varphi$$

is an ordinary function (real, complex, or matrix as in spinor fields) while in canonical quantization the integrand is made of creation and annihilation operators in the occupation number formalism.

Bailin and Love work out completely a very simple case of a real scalar field where

$$\mathcal{L} = \frac{1}{2}(\partial_\nu\varphi)(\partial^\nu\varphi) - \frac{1}{2}\mu^2\varphi^2. \tag{2.9}$$

This simple example suffices for showing the problem and the reason for it.

Bailin and Love make use of the Gaussian integral

$$\int dy \;\; \exp(-\frac{1}{2}ay^2 + \rho y) = (2\pi)^{\frac{1}{2}} a^{-\frac{1}{2}} \exp\left(\frac{1}{2}\frac{\rho^2}{a}\right).$$

If $A = (a_{ij})$ is a diagonal matrix ($a_{ij} = 0$ if $j \neq i$) and every $a_{ii} > 0$ then

$$\int dy_1 \cdots dy_N \;\; \exp\left(\sum_{i=1}^{N}\left(-\frac{1}{2}a_{ii}y^2 + \rho_i y_i\right)\right)$$

$$= (2\pi)^{\frac{N}{2}} \prod_{i=1}^{N} a_{ii}^{-\frac{1}{2}} \exp\left(\frac{1}{2}\sum_{i=1}^{N}\frac{\rho_i^2}{a_{ii}}\right).$$

This can be written as

$$\int dy_1 \cdots dy_N \;\; \exp\left(-\frac{1}{2}y^T A y + \rho^T Y\right) = (2\pi)^{\frac{N}{2}} (\det A)^{-\frac{1}{2}} \exp\left(\frac{1}{2}\rho^T A^{-1}\rho\right)$$

where $y = [y_1 \ldots y_N]^T$ and $\rho = [\rho_1 \ldots \rho_N]^T$. Changing y to a function φ where $\varphi(x_i) = y_i$ for $x_i = (x_{\max} - x_{\min})/N + x_{\min}$ and taking the limits $N \to \infty$, $x_{\max} \to \infty$, $x_{\min} \to -\infty$ we can write the result as a path integral

$$\int \mathcal{D}\varphi \;\; \exp\left(-\frac{1}{2}\varphi^T A\varphi + \rho^T\varphi\right) = (2\pi)^{\frac{N}{2}} (\det A)^{-\frac{1}{2}} \exp\left(\frac{1}{2}\rho^T A^{-1}\rho\right). \tag{2.10}$$

The special form of the Lagrangian allows it to be written as

$$\mathcal{L} = \frac{1}{2}(\partial_\nu\varphi)(\partial^\nu\varphi) - \frac{1}{2}\mu^2\varphi^2 = \frac{1}{2}(\partial_0\varphi)^2 - \frac{1}{2}\nabla^2\varphi - \frac{1}{2}\mu^2\varphi^2$$

$$= \frac{1}{2}(\partial_0\varphi)^2 + F(\varphi, \nabla\varphi).$$

The conjugate momentum density is

$$\pi = \frac{\partial\mathcal{L}}{\partial(\partial_0\varphi)} = \partial_0\varphi \tag{2.11}$$

and the Hamiltonian density is

$$\mathcal{H}(\pi,\varphi) = \frac{1}{2}\pi^2 - F(\varphi, \nabla\varphi).$$

In (2.11) $\pi = \partial_0\varphi$. It can cause some worry in the following integration where π and $\partial_0\varphi$ are treated as independent functions. In reality this is not a problem. The conjugate momentum π is a discrete variable because the energy levels are proportional to π^2 and there is a gap between two adjacent energy levels in a quantized theory. Therefore the path integral over π is discrete. Because of that the derivative $\partial_0\varphi$ is not a continuous derivative function, it is a discrete approximation

$$\Delta_0\varphi = (\varphi(x_{i+1} - \varphi(x_i))/(t_{i+1} - t_i).$$

States have energy levels and they do not depend on taking a limit where $t_{i+1} - t_i$ goes to zero. Thus, the conjugate momentum π has a value which is not a function of $\Delta_0\varphi$. In the following integration we do not take a limit where the number of states goes to infinity. The expression (2.10) is valid for any finite N. Let us evaluate

$$I = \int \mathcal{D}\pi \;\; \exp\left(-\int d^4x\,(-i\pi\Delta_0\varphi + \mathcal{H} - J\varphi)\right)$$

$$= \exp\left(-\int d^4x(-F - J\varphi)\right)\int \mathcal{D}\pi \;\; \exp\left(-\int d^4x\left(\frac{1}{2}\pi^2 - i\Delta_0\varphi\pi\right)\right)$$

$$= \exp\left(\int d^4x(F + J\varphi)\right)$$

$$\bullet \int \mathcal{D}\pi \;\; \exp\left(-\frac{1}{2}\int d^4x' \int d^4x\pi(x')A(x',x)\pi(x) + \int d^4x\rho(x)\pi(x)\right) \tag{2.12}$$

where

$$A(x',x) = \delta(x'-x) \quad , \quad \rho(x) = i\Delta_0\varphi$$

By (2.10) the integral is

$$= \exp\left(\int d^4x(F+J\varphi)\right)\exp\left(\frac{1}{2}\int d^4x'\int d^4x\rho(x')A^{-1}(x',x)\rho(x)\right)$$

The operator $A(x',x)$ is inverted by first Fourier transforming it, inverting and transforming back

$$\mathcal{F}[\delta(x'-x)] = \mathcal{F}[\delta(x-x')] = e^{-i\omega x'} \quad , \quad \mathcal{F}^{-1}[e^{i\omega x'}] = \delta(x+x')$$

Inserting A^{-1} to (2.12) sets $x' = -x$. For this special Lagrangian $\varphi(x) = \varphi(-x)$. $\Delta_0(x')$ is a difference described in the coordinate x'_0. Thus, when x' is set to $-x$, the time coordinate changes the sign $x'_0 = -x_0$ and $\Delta_0(x') = -\Delta_0(x)$. Therefore $\rho(x')$ becomes $-\rho(x)$ when A^{-1} is inserted. The integral gets the form

$$\begin{aligned}
&= \exp\left(\int d^4x(F+J\varphi)\right)\exp\left(\frac{1}{2}\int d^4x\rho(x'_{|x'=-x})\rho(x)\right) \\
&= \exp\left(\int d^4x(F+J\varphi)\right)\exp\left(-\frac{1}{2}\int d^4x(\rho(x))^2\right) \\
&= \exp\left(\int d^4x(F+J\varphi)\right)\exp\left(\frac{1}{2}\int d^4x(\Delta_0\varphi)^2\right). \qquad (2.13)
\end{aligned}$$

We let N be large and approximate $\Delta_o\varphi$ by $\partial_0\varphi$ and insert I to the expression of $W[J]$

$$W[J] = \int \mathcal{D}\varphi \ \exp\left(\int d^4x \ \frac{1}{2}(\partial_0\varphi)^2 + F + J\varphi\right)$$

$$W[J] = \int \mathcal{D}\varphi \ \exp\left(\int d^4x \ \mathcal{L} + J\varphi\right). \qquad (2.14)$$

We notice that if there is an error causing divergence of Green functions in [2] such an error must come after (2.14). This is because not much changes from (2.8) to (2.14). After integrating (2.8) over the variation of π the integrand in the path integral (2.14) has the same form. While it is true that in [2] the integration over the variation of π the function π is treated as a continuous function, though it is discrete, we here made the integration in the discrete form and the result is the same. Let us continue.

The path integral over φ can also be calculated for this special $\mathcal{L}$. Bailin and Love select coordinates y so that

$$(\partial_\nu\varphi)(\partial^\nu\varphi) = (\partial_0\varphi)^2 - \nabla^2\varphi = -\sum\left(\frac{\partial\varphi}{\partial y_\nu}\right)^2 \tag{2.15}$$

This trick allows writing (2.14) as

$$W[J] = \int \mathcal{D}\varphi \;\; \exp\left(-\frac{1}{2}\int d^4y\left(\sum_{i=0}^{3}\left(\frac{\partial\varphi}{\partial y_i}\right)^2 + \mu^2\varphi^2\right) + J\varphi\right)$$

We can expand the integrand into a two variable operator

$$\int d^4y\left(\sum_{i=0}^{3}\left(\frac{\partial\varphi}{\partial y_i}\right)^2 + \mu^2\varphi^2\right) = \int d^4y' \int d^4y\left(\sum_{\alpha,\beta}\frac{\partial\varphi}{\partial y_\alpha}\frac{\partial\varphi}{\partial y_\beta} + \varphi\mu^2\varphi\right)$$

$$= \int d^4y' \int d^4y\varphi(y')A(y',y)\varphi(y)$$

where

$$A(y',y) = \left(\frac{\partial}{\partial y'_\alpha}\frac{\partial}{\partial y_\beta} + \mu^2\right)\delta(y'-y).$$

The whole path integral takes the form

$$W[J] = \int \mathcal{D}\varphi \;\; \exp\left(-\frac{1}{2}\varphi^T A\varphi + \rho^T\varphi\right)$$

with

$$\rho^T\varphi = \int d^4yJ(y)\varphi(y).$$

By (2.10)

$$W[J] = \exp\left(\frac{1}{2}\rho^T A^{-1}\rho\right)$$

$$= \exp\left(\frac{1}{2}\int d^4y' \int d^4yJ(y')A^{-1}(y',y)J(y)\right).$$

It remains to invert $A(y', y)$. A can be expressed as a convolution product of two functions: A is clearly the same as

$$A(y', y) = \left(\left(\frac{\partial}{\partial y_\alpha}\right)^2 + \mu^2\right) \delta(y' - y) = f(y)g(y' - y).$$

Fourier transforms of $f(y)$ and $g(y)$ are

$$f(p) = \mathcal{F}[\left(\frac{\partial}{\partial y_\alpha}\right)^2 + \mu^2] = p^2 + \mu^2$$

$$g(p) = \mathcal{F}[\delta(y)] = 1.$$

The transform of a convolution is the product $f(p)g(p)$. Its inverse function is

$$A^{-1}(p', p) = (p^2 + \mu^2)^{-1}$$

and the inverse Fourier transform

$$\Delta(y' - y) = A^{-1}(y', y) = \int d^4 p e^{ip\cdot(y-y')}(p^2 + \mu^2)^{-1} \tag{2.16}$$

diverges. The divergence is caused by the dimension of the integration space

$$\approx \int dp|p|^3(p^2 + \mu^2)^{-1} e^{ip\cdot(y-y')}$$

In the Minkowski space the function $\Delta(y' - y)$ changes to the extent that instead of $p^2 + \mu^2$ there is $p^2 - \mu^2$ and for avoiding a pole in the complex plane when moving from the Euclidean space to the Minkowski space it is customary to add $+i\epsilon$, thus the propagator in the Minkowski space is $p^2 - \mu^2 + i\epsilon$ The divergence problem of the integral is the same in the Minkowski space as in (2.16) and we get a similar function $\Delta(y' - y)$. The free-field Green function $\mathcal{G}^{(2)}(y', y)$ is imaginary unit i times the Minkowski space function $\Delta(y' - y)$ and it is given by a divergent integral, similar to (2.16). The function $\Delta(y' - y)$ appears in Feynman diagrams for scalar field theories and it creates divergent Feynman diagrams.

I counted four errors in this short derivation as it is given in [2]. The first error is in their equation (1.4) (in [2]) where $\ln \det A = Tr \ln A$ must be $\ln \det A = Tr \ln U^T AU$ where U diagonalizes the symmetric matrix A. All eigenvalues of A must be positive. The second error is moving ln inside the integral from (1.9) to (1.10). The third error is the confusion with $\partial\varphi/\partial\bar{x}_0$ and π in (4.18) in [2]. The fourth error is in (4.22) in [2]. They do not invert A and have a sign error in (4.23). See how it is done here in (2.12-2.14). Bailin and Love have a correct formula in (4.24), it is the same as the equation (2.14) here. Fixing these small errors (and possibly introducing some new ones that I cannot see) does not remove the divergence of the Green function, so essentially the calculation in [2] is correct.

2.3 The reason for the divergence

As should be expected from an old and accepted result, the divergence of Green functions is not an error in calculations. Calculations in both books are essentially correct. Yet, Green functions are probabilities and they certainly must have values between zero and one.

There may be a normalization constant in the path integral. In Bailin's and Love's book there is such a constant, which I omitted in the short derivation of their result. If a constant is used, the constant must be the same for all Green functions. As is shown in [2] by calculating legs in Feynman diagrams, only some Green functions can diverge and most must converge for any Lagrangian. If the divergent diagrams are reduced to convergent ones by adjusting a normalization constant to an infinite value, then all convergent Green functions are reduced to zero. This is not correct and it is not so done in practical calculation: only divergent Feynman diagrams are renormalized and convergent diagrams are not changed. This may give reasonable results, but it is illogical. Adjusting a normalization constant is not a correct solution to the divergence problem.

The reason for the divergence is (and can only be) in the inverse Fourier transform at the end when a propagator is transformed to a Green function. One obvious concern is that the calculations in both books treat $k, \omega = p$ as a continuous (4-dimensional) variable, but this variable is discrete. k is the discrete state (it is meant to include all numbers needed to describe a state) of a quantized system, while ω is the energy parameter and energy values are discrete.

Because the variable p must be discrete, the transform must be discrete. It is not a Fourier transform, it is a Fourier series. Only periodic functions have a Fourier series, but a space limited function can be extended to a periodic function by adding copies of it to the empty space where the function is zero. The universe is space limited because it started expanding with a finite speed a finite time ago. Therefore we can complete the universe into a periodic function and have a discrete transform that gives quantized energy levels and discrete states. This does not fix the divergence of Green functions but it helps us towards the real reason.

Let us consider how a continuous integral, in this case one-dimensional, is changed to a discrete integral:

$$\int_0^L dx f(x) dy \to \frac{L}{N} \sum_{n=1}^{N} f((L/N)n). \tag{2.17}$$

If the variable is quantized, like the energy levels in a quantum system, there is a finite lower limit Δ to L/N. In that case N and L must grow together

$$\int_0^L dx f(x) dy \to \Delta \sum_{n=1}^{N} f(\Delta n). \tag{2.18}$$

If the integral diverges, as with divergent Green functions, the sum also diverges. One logical solution is to set a finite bound to N. After all, the universe is finite. The other logical solution is to separate the pure states (the discrete solutions) by changing the interval L/N to a smaller interval $\delta L/N$:

$$\int_0^L dx f(x) dy \to \delta \frac{L}{N} \sum_{n=1}^{N} f((L/N)n). \tag{2.19}$$

where $\delta \to 0$ fast enough so that the sum converges. If we do this, then the Fourier series does not converge to the Fourier integral, i.e., it is not as is typically required for Fourier series, but this choice can be motivated.

Following the path integral quantization we can quantize Lagrangians describing the boson field only without any fermion field. Indeed, the Lagrangian (2.8) describes a scalar boson field. The corresponding Euler-Lagrange equations are similar to the wave equation or the Klein-Gordon equation and quantized solutions of a wave equation mean standing waves, which occur if the space is limited. There is a countable number of standing wave solutions, harmonics in the wave equation. Higher frequencies imply higher energy. If we want to create a quantized boson field that has standing waves of arbitrarily small energy, we need more than one dimension (three is plenty) and we should set the diameter of the space to a different value for different directions. The diameter value should change in steps in order to get discrete energies. To each direction the lowest energy corresponds to the largest wavelength. If the space expands infinitely in one direction, we can construct a series of standing waves where the energy of the lowest standing wave goes to zero. The approach (2.19) for removing the divergence of Green functions separates these standing wave solutions and forbids state transitions. This kind of a (unphysical) quantized solution has no mass gap, i.e., there is no lowest positive energy state.

The problem of the existence of a mass gap in quantum field theories is motivated by the Coleman-Mandula theorem (see [5] p. 4) which classifies renormalizable quantum field theories assuming that there is a mass gap. The accepted renormalizable theories do not allow boson fields that realize Einstein's general relativity, but they allow Nordström's salar gravitation theory (which is just the scalar field studied and renormalized in [2]). The Coleman-Mandula theorem was the reason Julius Wess and Jonathan Bagger proposed supergravity in [5]. The question of mass gap was considered so important that the Clay Mathematics Institute proposed a Millennium Prize problem [6] of proving that quantum Yang-Mills theories have a mass gap.

The problem of a quantum Yang-Mills theory must be seen in the same sense as Bailin and Love see quantization: the goal is to quantize only the Yang-Mills boson field without any fermion field. It is not the question as in Quantum Chromo Dynamics (QCD), where the Lagrangian is (see e.g. [4] p.

114)

$$\mathcal{L} = -\frac{1}{4}F^{\alpha}_{\mu\nu}F^{\alpha\mu\nu} + \bar{q}(i\gamma_\mu D^\mu - \hat{m})q \tag{2.20}$$

where the first part is the bosonic Yang-Mills field of gluons and the second part is the fermionic field of quarks. (Here quarks have mass $\hat{m}$, though Bailin and Love write in [2] that mass breaks the gauge symmetry.)

When (2.20) is quantized through the path integral method ([2] describes quantization of QCD), there comes a new divergence problem: it is necessary to remove the gauge symmetry before integrating the path integral. It can be made by the way invented by Faddeev and Popov. In this way the gauge is fixed and the process creates artificial entities, Faddev-Popov ghosts. In Quantum Electrodynamics (QED) it is also necessary to fix the gauge for this reason, but as the U(1) group in QED is Abelian, Faddev-Popov ghosts do not need to be used (still [2] calculates the ghost terms for QED).

In 2009 there were a few attempts to solve Clay's problem [8][9][10]. They all started by fixing the gauge. I found this method unnecessary as the problem statement [1] does not say that there should be a fermion field (indeed, following the approach of Bailin and Love, there should not) and the simple method illustrated above should be sufficient. Thus, in [7] I constructed solutions to classical Yang-Mills equations such that it is possible to select an infinite sequence of solutions with energy approaching zero. Then, using the approach of (2.19) I constructed a quantized field that has no mass gap. The solution is surely not physical, but [1] is a mathematical problem motivated by a mathematical theorem, the Coleman-Mandula theorem, and [1] has an essentially mathematical formulation, more aking to the differential topological treatment in [11] than to any honest physics book [1][2][3][4][5].

In [7] the quantization approach is as follows. In the path integral method the Lagrangian is a classical field, an ordinary (in this case, matrix) function. Quantization in the path integral method means first finding solutions to the eigenvalue equations (2.7) for a suitable operator. There is only the boson field. There is no fermion field in the Lagrangian, thus there is no Schrödinger equation or any other similar equation that would give the eigenvalue equations in (2.7). The only equations we have are the the Euler-Lagrange equations, which in this case are the classical Yang-Mills equations. We can compare the situation to Maxwell equations and wave equations: the eigenstates are standing waves. They are solutions to the classical Euler-Lagrange equations.

As the system must be quantized, a discrete set of these, in a sense, standing wave solutions must be selected. The way standing waves are selected for the wave equation is that the dimensions of the space impose them. If the dimensions are finite (or the wave functions are periodic) we get a Fourier series, which is in a sense quantized. In our case of Yang-Mills equations instead of wave equations there also should be some conditions that impose a set of solutions, but we can work the other way: select a set of solutions and let them impose the conditions. Then we do not need to construct the conditions and we also do not need to construct the operators of (2.7). They exist but are not explicitly constructed as they are not needed in the path integral quantization method. Solving the Yang-Mills equations and selecting a set of solutions as pure states is the first step.

The second step is to treat the solutions as orthogonal. It gives the delta function $\delta(y-y')$ in the operator $A(y',y)$ in section 2.2. This delta is in fact (see [2] p. 6 eq. (2.10))

$$\langle q_{j+1}|\hat{H}|q_j\rangle = \delta(q_{j+1}-q_j)$$

in the Schrödinger picture when Bailin and Love derive their equations by dividing the axis into small intervals. This is later in [2] on page 42 eq. (4.20). We do not need to find a metric where the solutions are orthogonal. We can impose that they are orthogonal and give the delta function. Thus, there are so far two steps in the path integral quantization procedure: starting from a classical Lagrangian find or select solutions to the eigenvalue value equation, which amounts to solving the classical Euler-Lagrange equations, and consider these discrete solutions (states) orthogonal. The path integral method does not treat the solutions as discrete in evaluation of the path integral as the integral is not a sum.

Then follows the third step in this quantization procedure: the important step of removing divergencies. This step does not come from the operators that define eigenstates and eigenvalues. This step must be added in order to get physical values: probabilities that are between zero and one. This step may include the method of Faddeev and Popov of fixing the gauge, and it certainly involves the step of renormalization.

The way this third step is done in [7] is by following the approach of (2.19) of separating the states to a sum of copies of natural numbers (figuratively saying, lowest energy states in each direction and multiples of these energies).

This approch does not need fixing the gauge. The integral is reduced convergent by selecting δ in (2.9). Of the two possible logical solutions (2.18) and (2.19) for rendering divergent Green functions (which always appear in these quantum theories) the latter one is better suited to Yang-Mills fields. At least Bailin and Love discard renormalization by assigning a threshold to the inverse Fourier transform with the reason that it conflicts with gauge symmetry. I think this way of removing infinities is better motivated than the way infinities are removed in canonical and path integral quantization.

Bailin and Love ([2] p. 79) discuss renormalization of the scalar field theory (related to section 2.2) and mention the cut off of the momentum integration at some high level as the simplest method, which they do not recommend. They also mention that up to 1972 a commonly used renormalization method was to dump high momentum values with a function of the type (here $k = p$ in (2.16))

$$\frac{-\lambda^2}{k^2 - \lambda^2 + i\epsilon} \tag{2.21}$$

but their recommended method is dimensional renormalization where the integral is calculated in a 2ω-dimensional space (instead of 4-dimensional) and $\omega \to 2$. In this case the integral (2.16) changes to the type

$$\int d^{2\omega} \frac{k}{(2\pi)^{2\omega}} (k^2 - \mu^2 + i\epsilon)^{-n} = i(-1)^n \frac{\mu^{2\omega-2n}}{(4\pi)^{\omega}} \frac{\Gamma(n-\omega)}{\Gamma(n)}. \tag{2.22}$$

The Gamma function $\Gamma(n-\omega)$ has a pole at $\omega = 2$.

While the methods (2.21) and (2.22) work well for the purpose of calculations, they cannot be motivated in the physical situation. Energy levels in a quantized system are discrete and given by natural numbers: in the solutions (2.4) the energy values are $\frac{k^2}{2m}$. There is no justification in assuming that these energy values (related to the coordinate $p_0 = \omega$) or the state parameters (related to p_i, $i = 1, 2, 3$) are dumped or get thinner for large values of the coordinate. The discrete values of the momentum form a grid in the Euclidean space and the number of grid points grows as the ball of the Euclidean 4-space.

I conclude that the only two logical ways to stop the divergence of Green functions are the approaches in (2.18) and (2.19). In a mathematical Yang-Mills problem (2.19) was the preferred one as it does not break symmetry, but in a finite universe that has mass, like ours, (2.18) is more natural.

An upper bound for N in the $p_0 = \omega$ coordinate, time, means that there is a maximum frequency. A maximum frequency means that there is a minimum wavelength. A minimum wavelength means that in practice the space is composed of finite size units. The space-time is discrete.

There need not be any implications of the discreteness of space-time to calculations in quantum field theories because if N is very large, continuous approximations are very good. In practice the integral over the momentum space must be discrete, but treating it as a continuous space does not lead to a noticable error. Yet, in a philosophical sense there is a major difference: special relativity needs modifications if the space-time is discrete. I tried to make an Einsteinian thought experiment a discrete model as an alternative to special relativity in [12]. The text of that article is included as the next chapter in this book.

References to Chapter 2:

[1] R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem.* Dover Publications, New York, 1992 (first published 1967, 1976).

[2] D. Bailin and A. Love, *Navier-Stokes Introduction to Gauge Field Theory*, Adam Hilger, Bristol and Boston, 1986.

[3] V. Heine, *Grop Theory in Quantum Mechanics*, Dover Publications, New York, 1993 (first published 1960).

[4] D. Ebert, *Eichentheorien, Grundlage der Elementarteilchenphysics*, Akademie-Verlag, Berlin, 1989.

[5] J. Wess, J. Bagger, *Supersymmetry and Supergravity*, Princeton Univ. Press, New Jersey, 1992.

[6] A. Jaffe and E. Witten: Quantum Yang-Mills Theory. available online at www.claymath.org.

[7] J. Jormakka: Solutions to Yang-Mills equations. arxiv:1011.3962 15. Nov 2010.

[8] L. D. Faddeev: Mass in Quantum Yang-Mills Theory (comment on a Clay Millennium Problem). arxiv:0911.1013v1 5. Nov 2009.

[9] M. Frasca: Mass gap in a Yang-Mills theory in the strong coupling limit. arxiv:0511173v6 26. Jan 2007.

[10] A. Dynin: Energy-mass spectrum of Yang-Mills bosons is infinite and discrete. arxiv:0903.4727v2 20. May 2009.

[11] D.S. Freed and K. K. Uhlenbeck, *Instatons and Four-Manifolds.* Springer-Verlag, New York, 1984.

[12] J. Jormakka: Is Special Relativity really tested empirically? Available at ResearchGate.

3. Is Special Relativity tested empirically?

Newtonian physics is based on the absolute time and the coordinate transform from rest frame of reference R to a moving frame of reference R' is the Galileo transform

$$x' = x - vt \tag{3.1}$$

$$t' = t$$

$$y' = y$$

$$z' = z$$

for R' moving with the speed v with respect to R. If a signal sent from R' has the velocity u' in the x-direction the signal has the velocity

$$u = v + u' \tag{3.2}$$

in the R frame of reference. The Galileo transform has the inverse transform

$$x = x' + vt' \tag{3.3}$$

$$t = t'$$

$$y = y'$$

$$z = z'$$

i.e., the inverse is obtained by changing the sign of the velocity. This is the principle of relativity: there is no preferred frame of reference, all frames of reference moving with a constant velocity are equivalent. We can consider R' as the moving frame and R as the rest frame, or think about R' as the rest frame and R as a moving frame of reference, moving to the opposite direction with the same speed v.

The Lorentz transform for these two frames of reference is

$$x' = \gamma(x - vt) \tag{3.4}$$

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)$$

$$y' = y$$

$$z' = z$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{3.5}$$

The Lorentz transform has the inverse transform

$$x = \gamma(x' + vt') \tag{3.6}$$

$$t = \gamma\left(t' + \frac{vx'}{c^2}\right)$$

$$y = y'$$

$$z = z'$$

Also this transform has no preferred frame of reference. The value of γ is determined from the requirement that there is no preferred frame of reference: inverting (3.4) gives

$$x = \frac{1}{1 - \frac{v^2}{c^2}} \gamma^{-1}(x' + vt')$$

and setting

$$x = \gamma(x' + vt')$$

gives γ the value in (3.5). If a signal with the velocity u' in the x-direction is sent from R' the velocity observed in R is

$$u = \frac{v + u'}{1 + (vu'/c^2)} \tag{3.7}$$

It was this velocity addition formula that made me look for a discrete model.

A discrete space consists of finite size space volume elements which in the x direction have the length Δx_i and a time which is divided into discrete size time elements Δt_i. Initially we can assume that all space elements have the same length Δx and all time elements have the length Δt. Movement can proceed only to the neighboring element. A finite model has a natural maximum speed that we can assign the value c

$$c = \frac{\Delta x}{\Delta t} \tag{3.8}$$

I do not assume that the space volume elements are in a lattice and cannot move relative to each other. The model can be best compared to liquid: space volume elements have a fixed size and nearest neighbord but they can move in the space. The space is three-dimensional, thus the time is not considered a coordinate, though we can treat it as a coordinate in a coordinate transform.

It may feel that a discrete mode is unnatural and a continuous space is more natural for the reality. The motivation for a discrete model is that all interactions with the exception of gravitation have been modelled as quantum gauge field theories. Gauge fields introduce new dimensions for the symmetries. If the space is continuous, we have to add more dimensions to the space: in each point in the space there is needed a number of compactified dimensions, like small circles where the symmetries are realized. This model does not seem very natural: we have no evidence that there are more dimensions in the reality. The only way to dispense with these dimensions is to realize them as constructions in the dimensions that we have, but constructions require volume. Thus, the space has to have finite size volume elements. Our space has a maximum speed c. If the space consists of finite size space elements and there is a maximum speed, then the time must consist of finite size time units so that moving one space unit in one time unit gives the maximum speed c. This reasoning leads to the discrete model, but in this very simple analysis I will not go to the model at all.

Let us assume that the space volume elements keep a state and a test particle in such a discrete space is some kind of a construction of space elements in certain states. A test particle can move to any direction with a constant speed v. How can such a movement be realized in a discrete model? The test particle can only move in discrete steps. We could assume that the test particle keeps counters and counts how many steps it has taken to each direction in a number of time steps, but this seems unnatural. A better way is to assume that the test particle keeps a probability for a step in a given direction in a time unit. In each time step it makes a probabilistic choice whether to move to a given direction or not to move. If the space and time elements are small, this results into a fairly straight movement to a chosen direction with a given velocity v and the model does not need a more complicated state than a test mass has in a continuous space.

The velocity addition formula (3.7) seems very unnatural for a discrete model. This is so because in a discrete model there is a preferred frame of reference: the space elements have a rest frame. All movement is in reality happening in the rest frame and the test particle would have to implement (7) in some way. The formula (3.7) requires calculations that a test particle cannot be expected to do.

A natural velocity addition formula that a test particle can be expected to do is the following. In each time step the test particle has a rule to move to the nearest neighbor in a given direction or not to move, depending on the outcome of the probabilistic decision. If the test particle is sent from a moving frame of reference with the velocity u' in the x-direction it follows the same rule: it makes the same number of decisions but if the frame of reference already has a move to a given direction in a given time unit, the test particle omits this time unit. Thus, it makes a decision only for those time units when the frame of reference has no move in the given direction. I give the rule here only for a movement in the x-direction. If the frame of reference moves with the speed v in the x-direction, the proportion of time units when the frame does not move is

$$P_1 = 1 - \frac{v}{c}$$

For these time units the test particle makes the decision to move to the x-direction or not to move. The proportion of time units when it will not move is

$$P_2 = 1 - \frac{u'}{c}$$

When observed from the rest frame R the test particle does not move with the probability P_1P_2 in a given time unit and therefore it moves with the velocity

$$u = c(1 - P_1P_2) = v + u' - \frac{vu'}{c} \tag{3.9}$$

in the frame of reference R. This natural velocity addition formula gives the maximum velocity c to all test particles. It is not more complicated than keeping the velocity in a continuous space.

Let (x', y', z', t') be the coordinates in R' for the test particle sent from R' with the velocity u'. In R the coordinates are (x, y, z, t). We fix the origins for these coordinates so that the test particle is sent from the origin in each coordinates. In R' it moves with the speed u', thus $x' = u't'$, while in R it moves with the speed u, thus $x = ut$. The frame R' moves with the velocity v with respect to R, therefore

$$x' = x - vt \tag{3.10}$$

The time transform is obtained from

$$t' = \frac{x'}{u'} = \frac{1}{u'}(x - vt) = \frac{1}{u'}(ut - vt) = \frac{1}{u'}(u - v)t$$

$$= \frac{1}{u'}(v + u' - \frac{vu'}{c} - v)t = \left(1 - \frac{v}{c}\right) t$$

Coordinates transverse to the movement do not change: $y' = y$, $z' = z$. The unit of measure in R' can be different from the unit of measure in R. If the unit of measurement is different, we should introduce a multiplier and write (3.10) as

$$x' = \gamma_1(x - vt)$$

for some γ. In the Lorentz transform this change of the unit of measurement affects only the direction of the movement. We can make a similar assumption. The transform gets the form

$$x' = \gamma_1(x - vt) \tag{3.11}$$

$$t' = \gamma_1 \left(1 - \frac{v}{c}\right) t$$

$$y' = y$$

$$z' = z$$

where γ_1 is not yet determined. This transform has the inverse transform if $v \neq c$

$$x = \gamma_1^{-1} \left(x' + \frac{vc}{c - v} t'\right) \tag{3.12}$$

$$t = \gamma_1^{-1} \frac{c}{c - v} t'$$

$$y = y'$$

$$z = z'$$

The discrete model has a rest frame of reference and the inverse transform is not obtained by inverting the sign of v. It is not possible to set γ_1 to any value so that there is no preferred frame of reference. If we require that

$$x = \gamma_1 (x' + vt')$$

and equate this with the form in (3.12) we get

$$\gamma_1 = \sqrt{1 + \frac{v^2}{u'(c - v)}}$$

while if we require that

$$t = \gamma_1 \left(1 + \frac{v}{c}\right) t'$$

we get $\gamma_1 = \gamma$ in (3.5).

The transform (3.11) is clearly not the Lorentz transform (3.4) and we would expect that surely experiments have ruled out (3.11). But this is not the case. If $u' = c$ in (3.11), the transform (3.11) is exactly (3.4) if γ_1 is set to γ. Also if $v = c$ the transform (3.11) is (3.4). Classical experiments by Michelson-Morley and Kennedy-Thorndyle used light which implies that $v = u' = c$. In the Mössbauer experiment the signal is gamma rays and again

$v = u' = c$. These experiments cannot distinguish between (3.11) and (3.4). The Ives-Stilwell experiment has $v < c$ but $u' = c$ and it also cannot tell the difference. The same seems to be true also to all later experiments: they always use a signal that travels with the speed of light.

The Ives-Stilwell experiment is of special interest as it verifies the value γ in the Lorentz transform. Or does it? The problem is that the Ives-Stilwell experiment, like modern experiments of the similar type, measured the time dilation from which we get the proper time. In special relativity the proper time is

$$\Delta t' = \sqrt{1 - \frac{v^2}{c^2}} \Delta t \tag{3.13}$$

but let us calculate the time dilation directly from the Lorentz transform (3.4). We have a clock in the R' frame of reference and measure signals (x_1, t_1) and (x_2, t_2) sent by the clock. In the Ives-Stilwell experiment the clock is ions speeded to a fraction of c, in the classical experiment $v = 0.005c$ but in modern experiments v is much closer to c. These ions emit light and the experiment has a way to separate the frequency shift caused by the time dilation from the frequency shift caused by the Doppler effect. Thus, the signal is light and travels with the speed $u' = c$. Two signals come from a clock that is in rest in R' if $x'_2 - x'_1 = u'(t'_2 - t'_1)$. That is, $x'_2 = x'_1$ is not the condition that the two signals come from a clock which is at rest in R'. The positions x'_2 and x'_1 can be the same for an object at rest in R' only if $v = u'$. The condition $x'_2 - x'_1 = u'(t'_2 - t'_1)$ is always filled as $x' = u't'$, thus we always measure signals that are in rest in R'. Assigning $u' = c$ the inverse transform (3.6) gives

$$x_2 - x_1 = \gamma(x'_2 - x'_1 + v(t'_2 - t'_1) = \gamma(c + v)(t'_2 - t'_1)$$

As $x = ut$ and $u = c$ in the experiment we get

$$c\Delta t = c(t_1 - t_2) = x_2 - x_1 = \gamma(c + v)(t'_2 - t'_1)$$

Thus the time dilation is

$$\Delta t' = t'_1 - t'_2 = \gamma^{-1}(c + v)^{-1}\Delta t$$

$$= \sqrt{(1 - \frac{v}{c})(1 + \frac{v}{c})} \frac{c}{c + v} \Delta t$$

$$= \sqrt{(1 - \frac{v}{c})(1 + \frac{v}{c})} \frac{1}{1 + \frac{v}{c}} \Delta t$$

$$= \frac{\sqrt{1 - \frac{v}{c}}}{\sqrt{1 + \frac{v}{c}}} \frac{\sqrt{1 - \frac{v}{c}}}{\sqrt{1 - \frac{v}{c}}} \Delta t$$

$$= \gamma \left(1 - \frac{v}{c}\right) \Delta t \tag{3.14}$$

Clearly, (3.14) is not (3.13), which was measured in the Ives-Stilwell experiment.

This is because (3.14) is the time delay of an oscillator oscillating in the x-direction, but in the described test we measure an oscillator that oscillates in the transverse direction. We can consider oscillation to be in the y-direction. The transform (3.4) and the inverse (3.6) do not treat this case as in those formulae x' is in the x-direction, but let us assume that an oscillator moves between two positions y'_1 and y'_2 with the speed $u' = c$. This move is half an oscillation and in R' it takes the time $\Delta t' = (y'_2 - y'_1)/u'$. The frame R' moves with the velocity v, thus in Δt it has moved the distance $v\Delta t$ along the x-axis in R. The y coordinate is unchanged in (3.4), thus $y_2 - y_1 = y'_2 - y'_1$ and this distance is along the y-axis in R. The total distance the signal has travelled in R is from the Pythagoras theorem

$$L = \sqrt{(y_2 - y_1)^2 + (v\Delta t)^2}$$

In R the signal travels with the speed u, thus

$$\Delta t = \frac{L}{u}$$

Inserting L and $y_2 - y_1 = y'_2 - y'_1 = c\Delta t'$ and solving gives

$$(\Delta t')^2 = \frac{u^2}{u'^2} \left(1 - \frac{v^2}{u^2}\right) (\Delta t)^2 \tag{3.15}$$

In the Ives-Stilwell experiment, as in all later experiments, $u' = u = c$, thus we get the equation (3.13)

$$(\Delta t')^2 = \left(1 - \frac{v^2}{c^2}\right)(\Delta t)^2$$

The parameter γ in (3.4) does not appear in this calculation. The calculation uses the Pythagoras theorem and assumes that in the y direction there is no length change.

For the transformation (3.11) time dilation in the x-direction is calculated in the same way as for the Lorentz transform. Inserting the always valid expression $x'_2 - x'_1 = u'(t'_2 - t'_1)$ to the inverse formula (3.12) yields the relation of $\Delta t'$ and Δt and it corresponds to the time ditation of a clock that is in rest in R' and oscillates in the x-direction. Thus

$$x_2 - x_1 = \gamma_1^{-1}(x'_2 - x'_1 + \frac{vc}{c-v}(t'_2 - t'_1))$$

$$= \gamma_1^{-1}\left(u' + \frac{cv}{c-v}\right)(t'_2 - t'_1)$$

Inserting $x = ut = (v + u' - \frac{vu'}{c})t$ yields

$$\gamma_1\left(v + u' - \frac{vu'}{c}\right)(t_2 - t_1) = \frac{c(v+u')}{c-v}(t'_2 - t'_1)$$

Thus

$$\frac{t'_2 - t'_1}{t_2 - t_1} = \gamma_1 \frac{c-v}{c(v+u') - vu'} \frac{c(v+u') - vu'}{c} = \gamma_1\left(1 - \frac{v}{c}\right)$$

We get a result similar to (3.14)

$$\Delta t' = \gamma_1\left(1 - \frac{v}{c}\right)\Delta t \tag{3.16}$$

The derivation of (3.15) only needs the Pythagoras theorem and that there is no length change in the transverse direction. It follows that (3.14) is valid also for the transform (3.11). We cannot insert the u' from velocity summation formula (3.9) to (3.15) because in (3.15) the velocities v and u'

are orthogonal and (3.9) only gives the summation for parallel velocities. Let us define summation of orthogonal velocities in the transform (3.11) by the formula

$$u^2 = v^2 + u'^2 - \frac{v^2 u'^2}{c^2} \tag{3.17}$$

This is similar to (3.9) and also limits the maximal velocity to c. Solving

$$\frac{1}{u'^2} = \frac{1}{u^2 - v^2}\left(1 - \frac{v^2}{c^2}\right)$$

and inserting to (3.15) gives

$$\Delta t'^2 = \frac{u^2}{u^2 - v^2}\left(1 - \frac{v^2}{c^2}\right)\frac{u^2 - v^2}{u^2}\Delta t^2 = \left(1 - \frac{v^2}{c^2}\right)\Delta t^2$$

Thus, it gives (3.13) for every u, not only for $u = c$. It is very good that the result does not depend on u since if the time dilation is a real phenomenon it cannot depend on u, the speed of the signal sent from R'. The velocity summation formula (3.17) is the only formula giving (3.13) for all values of u. The summation formula (3.17) is not the same as (3.7) in the Lorentz transform.

The time delay has been experimentally measured many times and can be accepted as a real phenomenon. Time is a scalar, therefore the time dilation in the x-direction must be the same as the time delay in the y-direction. Thus, (3.14) must give the same time dilation as (3.13). We get the equation

$$\sqrt{1 - \frac{v^2}{c^2}} = \gamma\left(1 - \frac{v}{c}\right)$$

Thus

$$\gamma = \sqrt{1 - \frac{v^2}{c^2}}\left(1 - \frac{v}{c}\right)^{-1}$$

$$= \frac{\sqrt{1 + \frac{v}{c}}}{\sqrt{1 - \frac{v}{c}}} \tag{3.18}$$

This value of γ does not give the form (6) for the inverse transform and therefore it rules out the principle that there is no preferred frame of reference.

This observation solves the twin paradox: a twin, who travels with a speed close to c ages slower than the one who stays at home, but if R can be considered moving and R' at rest, the result is the opposite. The solution is that for speeds v close to c we cannot consider R' to be at rest and R as moving.

We can set γ_1 to the value in (3.18)

$$\gamma_1 = \gamma = \frac{\sqrt{1+\frac{v}{c}}}{\sqrt{1-\frac{v}{c}}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left(1+\frac{v}{c}\right)$$

Inserting this value of γ_1 to (11) and (12) shows that the inverse transform is almost obtained by changing v to $-v$

$$x' = \frac{\sqrt{1+\frac{v}{c}}}{\sqrt{1-\frac{v}{c}}}(x-vt) \tag{3.19}$$

$$t' = \frac{\sqrt{1+\frac{v}{c}}}{\sqrt{1-\frac{v}{c}}}\left(1-\frac{v}{c}\right)t$$

$$x = \frac{\sqrt{1-\frac{v}{c}}}{\sqrt{1+\frac{v}{c}}}\left(x' + vt'\left(1-\frac{v}{c}\right)^{-1}\right) \approx \frac{\sqrt{1-\frac{v}{c}}}{\sqrt{1+\frac{v}{c}}}(x'+vt')$$

$$t = \frac{\sqrt{1-\frac{v}{c}}}{\sqrt{1+\frac{v}{c}}}\left(1-\frac{v}{c}\right)^{-1}t' \approx \frac{\sqrt{1-\frac{v}{c}}}{\sqrt{1+\frac{v}{c}}}\left(1+\frac{v}{c}\right)t'$$

The length contraction in (3.11) is the same as in the Lorentz transform assuming γ has the same value. For two points (x'_2, t') and (x'_1, t') at the same time the length in R' and in R relate as

$$\Delta x' = x'_2 - x'_1 = \gamma(x_1 - x_2) \tag{3.20}$$

For the transform (3.11) holds

$$t'^2 - c^{-2}x'^2 = t^2 - c^{-2}x^2 + \frac{2vt}{c^2}(x-ct) \tag{3.21}$$

Thus, for $u' = c$ holds

$$ds^2 = dt'^2 - c^{-2}dx'^2 - c^{-2}dy'^2 - c^{-2}dz'^2 = t^2 - c^{-2}dx^2 - c^{-2}dy^2 - c^{-2}dz^2 = ds^2$$

and the transform leaves the space element invariant, but this is not true for $u' < c$. The invariant property for (3.11) is

$$x' - ct' = x - ct$$

The transform (3.11) does not describe the geometry of a flat Minkowski space.

As $ds'^2 \neq ds^2$ and the inverse transform is not obtained by changing the sign of the velocity, there is no argument that in some way proves that the Lorentz transform (3.4) is the correct one.

It has been known for a long time that the mass depends on the velocity and becomes infinite when the velocity approaches c. These experiments were started by Thomson (1893) and Searle (1897) and later continued by Kaufman and others. The formula for the moving mass

$$\frac{m_t}{m} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{3.22}$$

was derived by Lorentz from his theory where an electron gets deformed in the direction of movement. The index t in m_t refers to the transversal mass as the growing mass seemed to be transversal to the movement. Now it is called the moving mass and Einstein's totally different derivation of this formula is considered as the correct one, but perhaps it should be reconsidered.

In Kaufman's experiments the mass growth was observed by measuring the trajectory of an electron in different speeds. The charge of an electron does not seem to depend on the velocity. The trajectory shows that the mass grows as in (3.22) and it must be in the frame of reference of the moving electron as it is measured from the equations of motion of the electron. That is, the equations of motion are given in the local coordinates of the moving particle. The mass grows in R', not necessarily in R. Indeed, if mass is conserved in R, mass cannot grow in R.

There are several ways of deriving (3.22), the first derivation being given by Lorentz. I suggest the following simple explanation, which I will not work out to a proof in this simple analysis. In the discrete model I consider space as liquid consisting of space volume elements which can flow and gravitation not as geometry where point masses move along geodesic lines but as a flow where space volume elements flow to point masses and disappear to a hole in the point mass. This model means that there must be some universe where the space volume elements go and that there must be antigravitation because otherwise we run out of space volume elements. Antigravitation would not appear as massive objects. It would appear as new space volume being created. Such sources of space elements would be as common as mass points in the universe, but stars in the universe are far apart and there is no reason to assume that there would be a source of space volume elements anywhere close to us. Some phenomena, possibly supernovas, would be explainable with antigravitational sources if this model were correct.

Starting from this simple model, a point mass eats space volume elements proportionally to its transverse size. If it has a higher speed it meets and eats more space volume elements. The higher speed can also be understood as a shorter time unit: because of the time dilation the time unit in R' is smaller in the same proportion as in (3.22). The smaller time unit makes it possible for the mass point to eat space elements faster. This is seen in the equations of motion as a larger mass. Because the mass gain is caused by the time dilation in the transverse direction it has the same dependency as (3.13).

This explanation I leave only on the idea level, but let us look at the usual derivation how from (3.22) we get to $E = mc^2$. If the mass m and the velocity v are both variable, the differential of the force can be expressed as

$$F = \frac{d}{dt}(mv) = m\frac{dv}{dt} + v\frac{dm}{dt}$$

The differential of the kinetic energy is then

$$dW_K = F ds = m\frac{dv}{dt}ds + v\frac{dm}{dt}ds = m\frac{ds}{dt}dv + v\frac{ds}{dt}dm = mvdv + v^2 dm$$

Assuming that we show

$$c^2 dm = mvdv + v^2 dm \tag{3.23}$$

it only remains to integrate from the rest mass m_0 to the moving mass m

$$W_K = \int_0^{W_K} dW_k = \int_{m_0}^{m} c^2 dm = (m - m_0)c^2$$

Assigning

$$W_0 = m_0 c^2$$

we get

$$E = W_k + W_0 = mc^2$$

Naturally it was known from atomic masses long before Einstein that there is mass loss which can be connected to binding energy and atoms may contain energy. The formula $E = mc^2$ was discovered before Einstein, thus he had a reason to expect that $W_0 = m_0 c^2$ holds. It remains to derive (3.23). It is derived from (3.22)

$$m_0 = m\sqrt{1 - \frac{v^2}{c^2}}$$

Squaring and multiplying by c^2

$$m_0^2 c^2 = m^2 c^2 - m^2 v^2$$

Then this expression is differentiated as

$$0 = 2mc^2 dm - 2mv^2 dm - m^2 2v dv \qquad (3.24)$$

that is

$$c^2 dm = mv dv + v^2 dm$$

That seems initially like a correct way to get (3.23), but in what frame of reference is the differentiation (3.24) done? The frame R' moves with the speed v. If v is changing there are some forces seen in R', the accelerating frame experiences these forces as gravitation. Nothing in this calculation indicates that it is done in R'. Indeed, it cannot be done in R' because $v = 0$ in R' all the time even if v is not constant. It must be that the differentiation is done in the rest frame R. The problem is that the mass grows as (3.22) in R', not in R. Experiments that show how mass grows are experiments where the growth of mass is seen from the trajectory of the particle. The trajectory is determined by the forces in the local coordinates of the moving particle, in R'.

It is very possible that the mass of the particle does not change in R. Mass is conserved in R and the reasons why the mass grows in R' can be that it is caused by the time unit becoming smaller, or the length grows in R'. In any case, (3.22) does not describe what happens to the mass in R. Therefore the differentiation cannot be made in R. If not in R and not in R', then the step (3.24) is not justified.

There is a simple way to prove $E = mc^2$ in the discrete model. The discrete space with only local interactions necessarily has a maximum speed:

$$c = \frac{s_u}{t_u}$$

where s_u is the length of a space unit and t_u is the time unit. The maximum acceleration a_u is speeding a mass from zero speed to the speed c in a single time unit t_u:

$$a_u = \frac{c}{t_u} = \frac{s_u}{t_u^2}$$

The parameters s_u, t_u, c and a_u are universal constants in this model. Space can be considered as incompressible liquid.

Force fields are created by another set of parameters: F_u, p_u, m_u and E_u where the index u refers to *unit* and F is force, p is pressure, m is mass and E is energy. These parameters are not universal constants: they get their value from the whole mass of the universe. It is possible to think of space volume elements flowing into holes in point masses. As the gravitation force points towards a point mass and a force grows to the direction where the pressure decreases it is better to think of the pressure as decreasing when approaching a point mass. Any of these parameters F_u, p_u, m_u, E_u can be taken as the cause of the others. I take the pressure as the cause. Thus

$$F_u = p_u s_u^3$$

$$m_u = \frac{F_u}{a_u} = \frac{p_u s_u^2}{a_u}$$

$$\frac{p_u}{m_u} = \frac{a_u}{s_u^2}$$

$$\frac{E_u}{m_u} = \frac{p_u s_u^3}{m_u} = a_u s_u = c^2$$

So $E = mc^2$. This calculation should not be interpreted in the way that the mass of the space element turns into energy. It should be understood as describing that if one space element is crushed in some way, the pressure created by of all outside space elements releases energy. This energy causes the fallout from the crushed space element to accelerate from the zero initial speed to the speed of c and to escape as a photon or some other particle.

We can compare energy released in nuclear reactions to energy released in a collapse of demolished building. When one floor is crushed e.g. by colliding the atom nucleus with a neutron (or a skyscraper with an airplane) the outside space elements (or the upper floors) fill the space left after the collapse. The energy released is the potential energy of the outside space elements and not some binding energy that used to keep the collapsed space element together. In a similar way, the energy that is released when a building collapses is not some binding energy that the concrete and steel of a destroyed floor had stored in their materia. Some energy is needed to destroy the structures of a floor, just like it is necessary to bombard atoms with neutrons, but the released energy is from the potential energy the upper floors had before they fell, i.e., it is the pressure times the changed volume.

Let us draw some conclusions from this simple thought experiment. The Lorentz transform is not the correct one and it has not been empirically verified. The alternative transform (3.11) has not been ruled out and it has the summation formula (3.17) which is the only formula that gives the correct proper time (3.13) so that it does not depend on the velocity u' of the signal sent from the moving frame. The principle that there is no preferred frame of reference is incorrect. The derivation of the mass growth and the $E = mc^2$ in the special relativity may be questioned and they are not the only ways to derive those properties. Geometrization is not the correct paradigm for gravitation.

References to Chapter 3:

For a lucid explanation of Special and General Relativity I recommend the small book by Einstein:

Abert Einstein, *The Meaning of Relativity.* Princeton 1922.

4. The EPR problem and Bell's Theorem

The EPR problem is one of the most famous paradoxes in theoretical physics. The basic worry of Einstein is very understandable: in quantum mechanics wavefunctions collapse in the whole space instantaneously. For Einstein this was not possible: all interactions can only proceed with the speed of light. Instantaneous action from distance was the main argument Einstein had against both quantum mechanics and extrasensory perception (ESP). Not unsurprisingly, some quantum physicists, like Robert Mattuck, were open to the idea that ESP might be explained by the instantaneous collapse of wavefunctions in quantum mechanics.

The Heisenberg uncertainty principle of quantum mechanics says that there are pairs of properties, in the original EPR problem position and momentum, that cannot both be measured precisely at the same time. In the EPR paradox these two particles are created at time zero from photons. As the particles are created as a pair, they have opposite momentums and they move with the same speed to opposite directions. Measuring the position precisely from one particle and the momentum precisely from the other particle gives a precise measurement of both position and momentum for one particle violating Heisenberg's uncertainty principle. It follows that if the position is measured precisely for one particle, it is impossible to measure the momentum precisely for the other particle. Thus, the second particle gets to know in some way which property was measured for the first particle. If the particles are far away, the fastest way information can move from one particle to the other is the speed of light, but according to quantum mechanics wavefunctions collapse instantaneously and the second particle knows what property can be measured from it immediately after the first particle was measured. This should be so even if the particles have moved light years away from each other. Einstein believed there must be a hidden channel how this information can move and that quantum mechanics violates the principle that nothing can move faster than the speed of light.

Though Einsten's thought experiment has not been made in the exact way as described in the EPR paradox, similar setups have been measured and the consensus opinion from these experiments is that the phenomenon is real: particles that once were together are connected in some way for all time in the future. Quantum mechanics predicts correctly, the explanation to the EPR paradox is not that quantum mechanics is an incomplete theory and the phenomenon does not happen. There also does not seem to be any hidden channel. A hidden channel could not be faster than light and it cannot solve the paradox.

There are at least two proposed solutions to this paradox, but both are unsatisfactory. The first is the Copenhagen interpretation, which essentially says that humans cannot understand quantum physics and should be happy that calculations work and stop thinking about it. If this solution is applied to other problems (do not think, it is enough that equations work), it destroys the whole science. Certainly everything in science must be understandable for at least some humans.

The second explanation is Hugh Everett's many worlds solution where each measurement splits the world to several worlds corresponding to different outcomes of the measurement. Humans also split in every measurement. This explanation sort of works but it has the absurdity that for each run (the world as we see it) the total energy (mass+energy) is conserved while in the whole world, all possible worlds, the total energy grows exponentially. Why should the total energy be conserved in one run if it can grow exponentially in the whole system?

I once was trying to build a more precise simulator for teletraffic models and found one explanation for the EPR problem that seemed to be more logical.

In simulators simulation time is discrete, but unless the time unit is very coarse or the system is very small, it is too slow to simulate by jumping from one time unit to the next. You have to use events and put events to an event queue which is read by the simulator. So, for instance, if telephone calls have a duration distribution e.g. negative exponential, you cannot check at every time unit if the call ended. In order to run the simulation in a reasonable time you have to take from a pseudorandom generator the duration of the call and put to the event queue an event scheduled to the time in the future when the call ends. It may happen that something changes the time when the call ends. For instance, electrical power may go down in the district. Then the call ends unexpectedly and the original event for the ending of the call must be cancelled in the event queue. Thus, every event that is put to the event queue is a guess and the guess may turn out to be wrong because of other events and if this happend they must be cancelled. Assuming that the simulator servicing the event queue reads two events that occur at the same time but are contradictory (like one saying that the power went down, the other saying it did not) then the simulator has a problem. What typically would happen in event driven simulation is that the event that was first in the event queue is considered first, so it is treated as if it happened before the other event. But this is wrong as the events had the same time. What logically should be done is to go back to the time when one of these events were created and to select another time for the event so that there will not be a contradiction in the future. But this is foreknowledge of future. Yet, the simulation must go on. We conclude: if the simulator must check that some rules are observed and there are no violations to the rules (contradictions), the simulator must sometimes change the history. It must go back to the time when an event was created and modify the future event with foreknowledge of the future.

Restating: in a simulation world, if the simulator imposes strict rules (like the Heisenberg uncertainty principle) that cannot be broken, then the simulator must at some cases change the history.

In this explanation there is no need to move information faster than the light. It is history that changes. In the explanation the simulator makes only one run (though it changes the run by returning to an earlier time) not all possible runs like Everett's many world solution.

The explanation also has interesting philosophical implications. Going back and changing the history is possible only in the simulation time. In the real time history cannot change because what once happened cannot be changed as a record of history does not exist in any event queue. If there is a simulation time there necessarily also must be the real time and the real world because a simulator cannot run without a real time. The simulation time is simply a counter. And finally, there is something wrong in a system that can lead to contradictions with strict rules. Axiomatic systems do not lead to contradictions. In the EPR problem the reason why a contradiction appears is that observers can make measurements. These observers do not belong to the model. They are external elements that can break the rules. Naturally, these observers might be living beings. They do not belong to the material world. Clearly, the simulator explanation offers some deep insight to the reality of the reality.

So, this was kind of a nice theory, but much later I looked at the actual evidence that is claimed to show that this mysterious phenomenon actually has been experimentally verified. The claims of experimental verification boiled down to Bell's theorem. I read the proof of Bell's theorem and found it incorrect. In fact, the problem is caused by Bohr's rule of counting probabilities: according to Bohr's rule scaling of the probabilities of observing a particle is done so that the sum of the observed probabilities is one. This may seem natural, but assuming that particles cannot always be observed, then the sum of probabilities should be less than one. We can explain it with a squarrel on a tree. You may see the squarrel on any branch of the tree. These are the observable states and they have some probabilities. But the squarrel can be behind the tree and not observable. A correct way for normalizing the probabilities would be that the sum is less than one. It is customary to calculate probabilities for elementary particles with Bohr's rule and when the quantum mechanical machinery has been built for this normalization everything works. The problem appeared when Bell applied the same Bohr's rule to detector values instead of elementary particles. By doing so Bell managed to derive a result that contradicts elementary probability theory. This is the famous Bell's theorem. Clearly, if your model contradicts basic mathematics, you should correct your model and not to claim that mathematics does not work in quantum mechanics.

That observation in a way decreased my faith in the experimental verification of the mysterious result of the EPR experiments. But there may be other verification experiments and certainly wavefunctions collapse in quantum mechaniscs instantaneously, which may lead to changing of history solution. I still hope there is something real behind these mysterious events, and still consider the simulation world solution quite interesting. It would be the idealistic illusion world explanation for the reality. Now, let us continue to Bell's theorem.

4.1. Bell's Theorem

Bell's Theorem [1] considers measurements of spin for entangled particles. The entangled wave function studied in this theorem is

$$|\phi\rangle = \frac{1}{\sqrt{2}}(|\psi_{z+}\rangle \otimes |\psi_{z-}\rangle - |\psi_{z-}\rangle \otimes |\psi_{z+}\rangle)$$

where $|\psi_{j+}\rangle$ and $|\psi_{j-}\rangle$, $j \in \{x, y, z\}$, are the eigenvectors of the Pauli matrices σ_j corresponding to the eigenvalues 1 and -1 respectively. The spin of the first particle is detected with two detector directions a and a' and the spin of the second particle is detected in the second measurement with two detector parameters b and b'. Measurement of the spin means applying operators

$$(\sigma \cdot a) \otimes Id \quad , \qquad Id \otimes (\sigma \cdot b).$$

In both cases the spin measurement can only give the results spin up or spin down and collapses the wave function to direct products of eigenvectors of Pauli matrices. The quantum correlation

$$C(a, b)_q = \langle \phi | Id \otimes (\sigma \cdot b)(\sigma \cdot a) \otimes Id | \phi \rangle$$

gives the expected value for the empirical correlation

$$C(a, b)_e = \frac{N_{++} + N_{--} - N_{+-} - N_{-+}}{N_{++} + N_{--} + N_{+-} + N_{-+}}$$

where $N_{\alpha,\beta}$, $\alpha, \beta \in \{+,-\}$, is the number of cases when the first particle is measured spin α and the second spin β. A direct calculation shows that

$$C(a,b)_q = -b \cdot a.$$

The measurements of entangled pairs in the directions a, b, a, b', a', b and a', b' form four time series that have only values ± 1. We can define binary probability variables A, B, A' and B' and assign to them probabilities of being 1 or -1 from these time series. The correlation between these variables is then the expectation value of the product of the variables, thus

$$C(a,b)_p = E(AB).$$

Binary variables satisfy certain inequalities, called Bell's inequalities. The CHSH inequality is a convenient Bell's inequality for proving Bell's theorem:

$$C(a,b)_p + C(a,b')_p + C(a',b)_p - C(a',b')_p \leq 2. \tag{4.1}$$

Assigning detector directions as

$$a = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a' = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad b = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}, \quad b' = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{4.2}$$

gives

$$C(a,b)_q + C(a,b')_q + C(a',b)_q - C(a',b')_q = 2\sqrt{2} > 2 \tag{4.3}$$

showing that $C(a,b)_q \neq C(a,b)_p$. Bell's inequality violations have been observed in experiments. However, the reason is quite simple. The detector values are normalized to give

$$a \cdot b = a' \cdot a' = b \cdot b = b' \cdot b' = 1 \tag{4.4}$$

which may initially seem correct, but it is not. The correct scaling is

$$\sum_i |a_i| = \sum_i |a'_i| = \sum_i |b_i| = \sum_i |b'_i| = 1. \tag{4.5}$$

Scaling the directions according to (4.5) removes the Bell's inequality violation in (4.2).

In order to show that the scaling (4.4) is incorrect, let the directions be scaled as in (4.4) and let us consider the first particle

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\psi_{z+}\rangle - |\psi_{z-})$$

without focusing on the entanglement. It is measured by applying the operator $\sigma \cdot a$. The wave function $(\sigma \cdot a)|\psi\rangle$ collapses to one of the eigenvectors $|\psi_{m\alpha}\rangle$, $m \in \{x, y, z\}$, $\alpha \in \{+, -\}$, of Pauli matrices and the corresponding eigenvalue is α. If the first particle collapses to $|\psi_{m\alpha}\rangle$ the second particle collapses to $|\psi_{m\beta}\rangle$, $\beta \neq \alpha$. In the second measurement this collapsed second particle collapses to one of the eigenvectors $|\psi_{n\gamma}\rangle$, $n \in \{x, y, z\}$, $\gamma \in \{+, -\}$. Because $|\langle\psi_{n\gamma}|\psi_{m\beta}\rangle|^2 = \frac{1}{2}$ if $n \neq m$, the probability of $|\psi_{m\alpha}\rangle$ collapsing to $|\psi_{n+}\rangle$ is the same as the probability that it collapses to $|\psi_{n-}\rangle$. As the eigenvalues for $|\psi_{n+}\rangle$ and $|\psi_{n-}\rangle$ are opposites, these contributions to the correlation of the first and the second particle cancel. There remains the collapse of $|\psi_{m\beta}\rangle$ to $|\psi_{m\gamma}\rangle$, $\gamma \in \{+, -\}$. As $|\langle\psi_{m\gamma}|\psi_{m\beta}\rangle|^2 = \delta_{\beta=\gamma}$ if can only collapse to $|\psi_{m\beta}\rangle$.

In the scaling (4.4) the wave function $|\psi\rangle$ of the first particle collapses to $|\psi_{m\alpha}\rangle$ with the probability a_m. Because of entanglement, the second particle collapses after the first measurement to $|\psi_{m\beta}\rangle$, $\beta \neq \alpha$, with the same probability a_m. In the second measurement this collapsed second particle collapses to either $|\psi_{n+}\rangle$ or $|\psi_{n-}\rangle$ with the weight $|b_n|$. The sum of these weights must be one. Thus, the probability of the collapse is

$$\frac{b_n}{\sum_i |b_i|.}$$

Thus, the quantum correlation is not $-b \cdot a$. It is

$$C(a, b)_q = \frac{-b \cdot a}{\sum_i |b_i|}. \tag{4.6}$$

Correcting the quantum correlation removes Bell's inequality violation in (4.2).

However, this is not the logical way to correct it. We certainly want that

$$\langle\psi|(\sigma\cdot a)(\sigma\cdot a)|\psi\rangle = a\cdot a.$$

This holds if we scale as in (4.5):

$$\sum_i |a_i| = 1.$$

Adopting this scaling we see that

$$\sum_i a_i^2 < 1 \tag{4.7}$$

and may wonder where the missing probability is. It is in the contributions that cancelled in the calculation of quantum correlation and which also cancel in the calculation of autocorrelation (4.6). Indeed, the probability of $|\psi\rangle$ collapsing to either $|\psi_{m+}\rangle$ or $|\psi_{m-}\rangle$ is not a_m^2. That is only the part that is seen. Calculating autocorrelation (4.6) can be understood as two measurements, as with the correlation of two particles. The first measurement collapses $|\psi\rangle$ to eigenvectors $|\psi_{m+}\rangle$ or $|\psi_{m-}\rangle$ with the probability $|a_m|$ in the scaling (4.5). In the second measurement these eigenvectors are further collapsed to $|\psi_{n\gamma}\rangle$ and the probability of collapses to either $|\psi_{n+}\rangle$ or $|\psi_{n-}\rangle$ is $|a_n|$. Thus, the sum of the probabilities of eigenvectors to which $|\psi_{m+}\rangle$ or $|\psi_{m-}\rangle$ collapse is

$$|a_m|(|a_x|+|a_y|+|a_z|) = |a_m|$$

of which we see only the part $|a_m|^2$ as the other parts cancel in the measurement. The total probability of scaling (4.5) is one, though it appears, as in (4.7), that probabilities do not sum to one. This phenomenon explains why experiments have verified violations of Bell's inequality. In these experiments the probabilities have been derived from the numbers of detected particles and their sum has been scaled to 1 as in (4.7). This ignores the probability of contributions that cancel.

Bell's Theorem is sometimes explained by the following example. Assume that in a test of entangled particles the detectors in both sides are perfectly aligned, $a_z = b_z = 1$. We see perfect anticorrelation. Then move the detector of the first particle in the (x,z)-plane to a small angle $\alpha = \gamma/2$ so that there are 1% errors in detection. Moving the detector of the second particle in the (x,y)-plane to the angle $\beta = -\gamma/2$ must also introduce 1% errors. Thus, there should be 2% errors, but according to quantum mechanics, and experiments, there will be 4% errors. The number of errors was $\sin^2(\gamma/2)$ when the (small) angle between the detectors was $\gamma/2$ and it grows to

$$\sin^2(\gamma) = (2\sin(\gamma/2)\cos(\gamma/2)^2 \approx 4\sin^2(\gamma/2)),$$

to four times as large when the angle between the detectors is γ.

In reality, there is no mystery. The number of errors is related to the x-coordinate of the detector as $\sin^2(\gamma/2) = a_x^2$ thus $a_x = \sin(\gamma/2)$. In a communication system analogy we can think that the fraction a_x of the bits have errors. When β is set to $\gamma/2$ the errors in the communication analogy grow to about $2a_x$. The number of noticed errors is thus about $(2a_x)^2$, that is, four times as many errors as when the angle was only in one side.

The mathematical form of the quantum correlation $-b \cdot a$ can be expressed with the angle θ between the detectors as

$$-b \cdot a = -b_z a_z - b_x a_x = -|b||a|\cos(\alpha)\cos(\beta) - |b||a|\sin(\alpha)\sin(\beta)$$

$$= -\cos(\alpha - \beta) = -\cos(\theta)$$

as $|a| = |b| = 1$ by the norming (4.4), which is used when deriving this mathematical form of the correlation and also in experiments that have confirmed the form $-\cos(\theta)$.

The function $-\cos(\theta)$ is -1 if $\theta = 0$, zero if $\theta = \pi/2$ and 1 if $\theta = \pi$. Sometimes it is argued that as a classical correlation should be a linear function and the linear function fitting to these three points is $\frac{2}{\pi}\theta$, but as experiments confirm that $-\cos(\theta)$ is the correct form, this is a demonstration that quantum mechanics differs from classical physics.

Quantum mechanics certainly differs from classical physics, there is e.g. the collapse of wave functions, but this particular mathematical form of quantum correlation does not touch those issues. The correlation should indeed classically be a linear function, but a linear function of $|a_x|$. As $|a_x| = \sin(\theta)$ we have to look for a linear function agreeing on those three points. Two free parameters is not enough to fit the three points. We have to still take a linear shift of θ in the inside function $\sin(\theta)$. Thus, we look for a solution of the type

$$C(a,b)_q = k\sin(\theta + \gamma) + r$$

where k, r and γ are to be determined. From the three points we get three equations and the solution is $k = -1$, $r = 0$ and $\gamma = \frac{\pi}{2}$ giving the correct quantum correlation $C(a,b)_q = -\cos(\theta)$.

For clarity, I will go through the problem in Bell's theorem again. We have expressed a vector in a base e_i as $a = \sum_i a_i e_i$. The square a_i^2 is interpreted as a probability. Then this probability is divided into parts by weights $p_{ij} \geq 0$:

$$a_i^2 = \sum_j p_{ij} a_i^2$$

These weights are real and nonnegative numbers and their sum must be one. The square root of the probability is also divided into parts by these weights

$$a_i = \sum_j p_{ij} a_i$$

and again the weights must sum to one or the probability does not stay constant. In Bell's theorem the weights p_{ij} are the settings b_j of the other detector. These b_j are real and nonnegative numbers. They divide the probability a_i^2 and they must sum to one in order to the probability to stay constant. In Bell's theorem the numbers a_i and b_j are scaled as wave functions by the Born rule and division of probability is made by a projection. The projection is done correctly, the way projections are usually done, but a projection in a complex space is not a valid way of dividing probability and here the question is division of probability: a violation of Bell's inequality is caused by the loss of probability in this division of probability by taking a projection.

It may seem natural that taking a projection should be the correct way as it is similar to the Born rule, but doing so leads to a contradiction: an elementary theorem from the probability theory is violated. Elementary mathematics cannot be violated in quantum mechanics just like they cannot be violated in any other field. Indeed, if basic mathematics would turn out false, we could just as well throw away all science as it depends on basic mathematics. The usual explanation is that the elementary probability theorem cannot be applied because the particles coming to the detectors do not have probabilities. This explanation is false, as I show in [2]. There are two base vectors (one was forgotten by Bell) and the particles coming to the detectors have probabilities in the normal way. The reason for the failure of the Bell's inequality can only be the scaling of the detector values and we can exactly see that the problem comes when dividing a_i by the weights b_j. The problem appears if the particles have probabilities when coming to the detectors and it appears in the same way if the particles do not have probabilities, so the usual explanation does not solve the contradiction.

In Born's rule a wave function is expressed in the base e_i and the squares of the norms of the coefficients sum to one. No contradicion has arisen from this rule and it is logical: the wave function is expressed in the basis of wave functions. Consequently a wave function is zero if all coefficients in this basis are zero. That is, the probability of a wave function giving the zero vector in this basis is zero. The detector values are not expressed in the basis of detector values. They are expressed in the basis of wave functions. The sum of components of the detector value in the basis of wave functions is a projection of the detector value to the basis of wave functions. The probability of the zero vector need not be zero and therefore the sum of the squares of the norms in this basis does not need to be one. If in doubt that some rule in quantum mechanics states that detector values must be scaled as wave functions try this: turn the detector head away from the beam of particles. Then the projection of the detector value in the basis of wave functions is zero, but the detector value has some value and is not zero. Keep the head in this position for half of the experiment and ask the audience what might be the probability of a zero vector for a detector value in this basis.

There is no valid reason to claim that detector values must be scaled as wave functions. The scaling of detector values must be selected and in Bell's experiment it must be selected in such a way that probability is not lost and an elementary probability theorem holds.

The reason for the failure of Bell's inequalities seems to be incorrect scaling of detector directions and it has nothing to do with hidden parameters in the EPR paradox, which is what we will next look at.

4.2. The EPR Problem

Albert Einstein, Boris Podolsky and Nathan Rosen proposed a thought experiment where conjugate properties are measured from two entangled particles in a paper [2] published in 1935. The authors claimed that the position could be measured precisely from one particle and the momentum from another. By Heisenberg's uncertainty principle position and momentum cannot both be precisely measured, thus the authors concluded that the measurements of entangled particles must be correlated: if the position is precisely measured from one particle, the momentum cannot be precisely measured from the other particle because of some mechanism. The authors suggested that the mechanism of such correlation can either be faster than light transfer of information from one particle to the other one, which Einstein naturally rejected, or that the particles have in some way agreed on how to answer to future measurements. This agreement would be stored in some variables in the particles. As such variables did not seem to exist in quantum mechanics, they become called hidden variables in the EPR paradox.

In 1964 John Bell reformulated the paradox as a measurement of the spin of two entangled particles and proved Bell's Teorem in [1]. His proof seemed to show that local hidden variables could not be a solution to the EPR paradox. However, as we saw, Bell's proof is not valid: Bell's inequality violations are caused by an incorrect normalization of detector directions. Experiments that have confirmed Bell's Theorem use the same incorrect scaling of detector directions and naturally get the same result as Bell.

The present short note shows that the EPR paradox in the form proposed by Bell can be solved by what he called hidden parameters, but these parameters are already in the formulation of quantum mechanics and therefore not in any way hidden. Notably it will be shown that the entangled particle system studied in the EPR paradox may be in two mixed states after breaking up, but in both states each particle has a definite spin angular momentum value.

In one of these states the first particle has the spin + in the x-direction and the second has -, while in the second mixed state the spins in the x-direction are the opposites. Assuming, as in the hidden parameter solution, that the particles do have definite spins, the wave functions are not mixes of these two states but one or the other.

As spin is conserved, spin directions are not changed by the following measurement in either particle. This simple mechanism gives the observed anticorrelation of the spins of the two particles. The argument does not show that this is the way it happens, but it shows that this is a possible logical explanation to the EPR paradox.

The entangled wave function studied in the EPR paradox in [2] is of the type

$$|\phi\rangle = c_1|\psi_{z+}\rangle \otimes |\psi_{z-}\rangle + c_2|\psi_{z-}\rangle \otimes |\psi_{z+}\rangle$$

where $|\psi_{j+}\rangle$ and $|\psi_{j-}\rangle$, $j \in \{x, y, z\}$, are the eigenvectors of the Pauli matrices σ_j corresponding to the eigenvalues 1 and -1 respectively and c_1, c_2 are complex numbers. The wave function $|\phi\rangle$ is created by a break-up of a spin zero single state and must have spin zero. We have to derive the spin zero subspace.

The first particle is

$$|\psi\rangle = c_1|\psi_{z+}\rangle + c_2|\psi_{z-}\rangle.$$

As

$$\langle\psi_{x+}|\psi\rangle = \frac{1}{\sqrt{2}}(c_1 + c_2)$$

$$\langle\psi_{x-}|\psi\rangle = \frac{1}{\sqrt{2}}(c_1 - c_2)$$

$$\langle\psi_{y+}|\psi\rangle = \frac{1}{\sqrt{2}}(c_1 - ic_2)$$

$$\langle\psi_{y-}|\psi\rangle = \frac{1}{\sqrt{2}}(c_1 + ic_2)$$

$$\langle\psi_{x+}|\psi\rangle = c_1$$

$$\langle\psi_{z-}|\psi\rangle = c_2$$

the probabilities of measuring the eigenvalues $|\psi_{n\alpha}\rangle$, $m \in \{x, y, z\}$, $\alpha \in \{+, -\}$, are the real numbers

$$|\langle\psi_{x+}|\psi\rangle|^2 = \frac{1}{2}(1 + (c_1^* c_2 + c_1 c_2^*))$$

$$|\langle\psi_{x-}|\psi\rangle|^2 = \frac{1}{2}(1 - (c_1^* c_2 + c_1 c_2^*))$$

$$|\langle\psi_{y+}|\psi\rangle|^2 = \frac{1}{2}(1 - i(c_1^* c_2 - c_1 c_2^*))$$

$$|\langle\psi_{y-}|\psi\rangle|^2 = \frac{1}{2}(1 + i(c_1^* c_2 - c_1 c_2^*))$$

$$|\langle\psi_{z+}|\psi\rangle|^2 = c_1^* c_1$$

$$|\langle\psi_{z-}|\psi\rangle|^2 = c_2^* c_2.$$

The wave function $|\psi\rangle$ can be expressed in a basis $|\psi_{m+}\rangle$, $|\psi_{m+}\rangle$ as

$$|\psi\rangle = \langle\psi_{m+}|\psi\rangle|\psi_{m+} + \langle\psi_{m-}|\psi\rangle|\psi_{m-}.$$

The second particle

$$|\psi'\rangle = c_2|\psi_{z+}\rangle + c_1|\psi_{z-}\rangle.$$

has similar formulas with c_1 and c_2 interchanged, but as it is moving in the opposite direction, the spin is inverse. Thus, the total spin + in the x-direction for the two particles is

$$\frac{1}{2}\hbar\frac{1}{2}(1 + (c_1^* c_2 + c_1 c_2^*)) + \frac{1}{2}\hbar\frac{1}{2}(1 - (c_2^* c_1 + c_2 c_1^*)) = \frac{1}{2}\hbar$$

The total spin - in the x-direction for the two particles is

$$-\frac{1}{2}\hbar\frac{1}{2}(1-(c_1^*c_2+c_1c_2^*))-\frac{1}{2}\hbar\frac{1}{2}(1+(c_2^*c_1+c_2c_1^*))=-\frac{1}{2}\hbar.$$

The total spin of the two particles in the x-direction is the sum of these numbers, i.e., it is zero for any c_1, c_2.

The total spin + in the y-direction for the two particles is

$$\frac{1}{2}\hbar\frac{1}{2}(1-i(c_1^*c_2-c_1c_2^*))+\frac{1}{2}\hbar\frac{1}{2}(1+i(c_2^*c_1-c_2c_1^*))$$

$$=\frac{1}{2}\hbar(1-i(c_1^*c_2-c_1c_2^*))$$

The total spin - in the y-direction for the two particles is

$$-\frac{1}{2}\hbar\frac{1}{2}(1+i(c_1^*c_2-c_1c_2^*))-\frac{1}{2}\hbar\frac{1}{2}(1-i(c_2^*c_1-c_2c_1^*))$$

$$=-\frac{1}{2}\hbar(1+i(c_1^*c_2-c_1c_2^*))$$

The total spin of the two particles in the y-direction is the sum of these numbers,

$$\frac{1}{2}\hbar i(-c_1^*c_2+c_1c_2^*).$$

This number is zero if

$$c_2=\pm\frac{c_1}{|c_1|}|c_2|.$$

The total spin + in the z-direction for the two particles is

$$\frac{1}{2}\hbar(c_1^*c_1+c_2^*c_2)=\frac{1}{2}\hbar$$

if the norm of the wave function is set to one. The total spin - in the z-direction for the two particles is

$$-\frac{1}{2}\hbar(c_2^*c_2+c_1^*c_1)=-\frac{1}{2}\hbar.$$

Summing the numbers shows that the total spin in the z-direction is always zero.

The total spin must be zero to all directions. Therefore $c_1^* c_2 - c_1 c_2^* = 0$. It implies that

$$|\langle \psi_{y+} | \psi \rangle|^2 = |\langle \psi_{y-} | \psi \rangle|^2 = \frac{1}{2}$$

i.e., the spin in the y-direction is zero for each particle.

We can set

$$c_1 = \frac{1}{\sqrt{2}}$$

as a basis vector can be multiplied by any complex number. We do not necessarily need to have $|c_1| = |c_2|$ but can select the basis wave functions so that this is true. Thus,

$$c_2 = \pm \frac{1}{\sqrt{2}}.$$

As a conclusion, there are two basis vectors for the subspace spin zero:

$$|\phi_1\rangle = \frac{1}{\sqrt{2}}(|\psi_{z+}\rangle \otimes |\psi_{z-}\rangle - |\psi_{z-}\rangle \otimes |\psi_{z+}\rangle)$$

and

$$|\phi_2\rangle = \frac{1}{\sqrt{2}}(|\psi_{z+}\rangle \otimes |\psi_{z-}\rangle + |\psi_{z-}\rangle \otimes |\psi_{z+}\rangle).$$

Both of these basis vectors give total spin zero to all directions for the two particle system and both give spin zero in y and z directions to each particle separately, but they do not give spin zero in the x-direction for each particle. Indeed, for $|\phi_1\rangle$ the first particle has the spin

$$\frac{1}{2}\hbar\frac{1}{2}(1 + (c_1^* c_2 + c_1 c_2^*)) - \frac{1}{2}\hbar\frac{1}{2}(1 - (c_2^* c_1 + c_2 c_1^*))$$

$$\frac{1}{2}\hbar(c_1^* c_2 + c_1 c_2^*) = -\frac{1}{2}\hbar$$

in the x-direction and the second particle has the opposite spin. For $|\phi_2\rangle$ it is inversely.

In the hidden parameter solution each particle has a definite spin after the single state has broken up. Thus, $|\phi\rangle$ is not a linear combination of the basis vectors $|\phi_1\rangle$ and $|\phi_2\rangle$. It is one or the other with equal probabilities.

The total spin of a particle cannot change from + to - in a measurement, thus, if the first particle has the spin + in the x-direction, it is also the total spin of this particle, and the first measurement can only collapse it to $|\psi_{m+}\rangle$ eigenvectors. It follows that the second particle must have - spin in the x-direction and the second measurement can only collapse it to $|\psi_{m-}\rangle$ eigenvectors. A similar conclusion is true if the first particles has spin -.

The basis vectors $|\phi_1\rangle$ and $|\phi_2\rangle$ are both mixed states and measurements collapse these mixed stated to pure states of eigenvectors $|\psi_{m\alpha}\rangle$, thus measurements have this somewhat mysterious property of collapsing wave functions, but the correlation is caused by the conservation of angular momentum, in this case, spin angular momentum. This mechanism gives the observed anticorrelation between the measurements without assuming neither instantaneous long distance information transfer nor any missing variables in quantum mechanics.

The original EPR paradox in [2] has a different solution. The authors of [2] assume that both position and momentum could be measured precisely from two entangled particles unless there is some mechanism causing correlation of measurements. There is such a mechanism in the case of spin or polarization measurements, but with conjugated properties in Heisenberg's uncertainty principle no such correlation mechanism is needed: particles are waves and they do not have precise values for conjugated properties.

A point mass has both a precise position and a precise momentum. The momentum can be obtained by measuring the mass and averaging the velocity over a long distance. Because the momentum is conserved, the velocity measured over any short distance equals the average over a long distance.

A wave packet, instead, has some minimum distance over which velocity can be measured so that it still closely equals the average velocity over a long distance. Trying to measure velocity over a shorter distance faces the problem that the wave extends outside this distance, i.e., the mass is not the whole mass. Therefore the momentum measurement becomes necessarily unprecise if the location is very precise. Indeed, a wave packet does not have a precise position in the sense a point mass has.

There is another problem in the problem setting in [2]. The authors assume that it is possible for all particles to measure what you want to measure, but is this necessarily true? Consider a case where balls can be red or green. They are sent in red-green pairs one ball going to one direction and the other to the opposite direction. Every now and then a tester observes a ball. With the green filter he can see the red ball but the green goes unnoticed, and with the red filter it is the opposite situation. The tester believes that his measurement with the green filter forces the ball to take the red color, and if he uses the red filter the ball takes the green color. The tester is surprised how the ball going in the opposite direction always has the opposite color and concludes that either the theory is wrong or information moves faster than light.

References for section 4:

[1] John Bell,*On the Einstein Podolsky Rosen Paradox. Physics.* 1 (3):195-200, 1964.

[2] Albert Einstein, Boris Padolsky, and Nathan Rosen, *Can Quantum-Me- chanical Description of Physical Reality be Considered Complete? Physical Review.* 47 (10):777-780, 1935.

5. Einstein is wrong, Nordström correct

Gunnar Nordstöm's two scalar theories of gravitation were published in the the set of articles [1]. A historical overview of the development and rejection of his theories explaining the arguments of that time is given in [2]. A fairly recent and very interesting scientific paper exploring the final for of Nordström's theory is in [3]. It explores the unique property that Nordström's theory shares with the General Relativity: in both theories the gravitational mass of the universe equals the inertial mass.

The field equation in Nordström's second gravitation theory is

$$\Phi \Box \Phi = -4\pi\rho \tag{5.1}$$

Φ is a continuous scalar field defined in the flat Minkowski four-space. ρ is scalar and defined in the four-space, but not necessarily continuous: in many cases mass can be replaced by a set of point masses, thus ρ is often best treated as a set of singularities. The gravitation constant G and the speed of light c are both set to one in this equation and $\Box$ is the D'Alembertian operator.

The field Φ in a flat Minkowski space gives the line element in Cartesian coordinates as (c is set to one.)

$$ds^2 = \Phi^2 dt^2 - \Phi^2 dx^2 - \Phi^2 dy^2 - \Phi^2 dz^2 \tag{5.2}$$

Einstein noticed (see [2][3][4]) that (5.2) can be written as

$$ds^2 = g^{ab} dx^a dx^b \tag{5.3}$$

for $g_{00} = \Phi^2$, $g_{ii} = -\Phi^2$ for $i > 0$ and $g_{ab} = 0$ if $a \neq b$. Thus, (5.2) can be interpreted as a line element of a curved Lorentz four-space. The Ricci scalar of the curved Lorentz space satisfies

$$R = -6\Phi^{-3} \Box \Phi \tag{5.4}$$

and (5.1) can be expressed as a geometric equation

$$\Phi^{-3}\Box\Phi = 24\pi GT \tag{5.5}$$

where T describes the mass-energy distribution. A bit later, in 1915-1916, Einstein formulated the General Relativity field equations

$$R_{ab} - \frac{1}{2}R = k_0 T_{ab} - \lambda g_{ab} \tag{5.6}$$

where $k_0 = 8\pi G/c^4$, T_{ab} is the stress tensor and λ is the cosmological constant. Nordström's field equation (5.1) can be obtained from Einstein's field equations by taking a trace provided that the metric tensor g_{ab} has the special form as in (5.3).

In the last form of Nordström's theory ρ in the field equation (5.1) was understood as T in (5.5) and as equal to the trace of the energy-stress tensor T_{ab} of the General Relativity, but in earlier forms of the theory ρ and T are not exactly related in this way. I will use the form (5.5) for Nordström's theory and will also adopt the interpretation of ρ in (5.1) as related to the stress-energy tensor T_{ab} as in (5.5), but will not consider the T_{ab} of Nordström's theory to be exactly the same as the T_{ab} in the General Relativity. The difference between the T_{ab} in the two theories is that at least in early forms of Nordström's theory the diagonal elements T_{aa} included the energy of the gravitational field and were not zero in a vacuum outside a point mass.

The line element in Nordström's theory has to be of the form (5.2) because if the field has the value $\Phi(x)$ at a point $x = (x^0, x^1, x^2, x^3)$ and it is continuous, then in a close distance from the point x in every direction the field has almost the same value $\Phi(x)$. It follows that the line element must be in Cartesian coordinates

$$ds^2 = \Phi^2 dt^2 - \Phi^2 dx^2 - \Phi^2 dy^2 - \Phi^2 dx^2$$

and in spherical coordinates

$$ds^2 = \Phi^2 dt^2 - \Phi^2 dx^2 - r^2\Phi^2 d\theta^2 - r^2 \sin^2(\theta)\Phi^2 d\psi^2$$

In Nordström's theory the field Φ is a continuous scalar in the Minkowski space and though the field equation can be calculated from geometry as in (5.5) the metric in (5.2) does not admit in a natural geometric interpretation: close to a mass the field Φ is stronger. It means that if ds^2 is a line element, the volume of the line element grows as Φ^3 when we go closer to a point mass. The space expands and the punctuated vacuum extends to the infinity when we approach the mass.

Schwarzschild's exact solution to Einstein's equations has the line element

$$ds^2 = B(r)dt^2 - A(r)dr^2 - r^2 d\theta^2 - r^2 \sin^2(\theta) d\phi^2 \tag{5.7}$$

where

$$A(r) = \left(1 - \frac{2GM}{rc^2}\right)^{-1} \quad , \quad B(r) = 1 - \frac{2GM}{rc^2}$$

It is not of the type (5.2) and thus Schwarzschild's solution is not a field. In Schwarzschild's solution the space elements elongate in the radial direction when we approach a mass and the solution can be imagined as (a four-dimensional version of) a membrane where a mass bends the geometry. In Schwarzschild's solution gravitation is curved geometry.

In orthogonal coordinates the nonzero Christoffel symbols are ($b \neq a$, no summation)

$$\Gamma^a_{aa} = \frac{1}{2} g^{aa} g_{aa,a} \qquad \Gamma^a_{ab} = \frac{1}{2} g^{aa} g_{aa,b} \qquad \Gamma^a_{bb} = -\frac{1}{2} g^{aa} g_{bb,a}$$

The Ricci curvature tensor is defined as $R_{bd} = R^a_{bad}$ where

$$R^a_{bcd} = \Gamma^a_{bd,c} - \Gamma^a_{bc,d} + \Gamma^e_{bd}\Gamma^a_{ec} - \Gamma^e_{bc}\Gamma^a_{ed}$$

is the Riemann curvature tensor. The Ricci scalar is $R = g^{ab} R_{ab}$. The Ricci tensor has the symmetry $R_{ab} = R_{ba}$ and consequently there are ten distinct entries.

Directly calculating from these definitions we get for orthogonal coordinates (explicit summing, $j \in \{0, 1, 2, \}$)

$$R_{jj} = f_j - \sum_{\substack{i=0 \\ i \neq j}}^{4} \left\{ \frac{1}{4} g^{ii} g_{jj,i} \left(\sum_{\substack{k=0 \\ k \neq j}}^{4} g^{kk} g_{kk,i} - g^{jj} g_{jj,i} \right) + \frac{1}{2} \partial_i (g^{ii} g_{jj,i}) \right\} \tag{5.8}$$

where

$$f_j = \sum_{\substack{i=0 \\ i\neq j}}^{4} \left\{ \frac{1}{4} g^{ii} g_{ii,j} (g^{jj} g_{jj,j} - g^{ii} g_{ii,j}) - \frac{1}{2} \partial_j (g^{ii} g_{ii,j}) \right\} \tag{5.9}$$

and the cross terms are

$$R_{ij} = \frac{1}{4} g^{jj} g_{jj,i} (g^{kk} g_{kk,j} + g^{mm} g_{mm,j}) + \frac{1}{4} g^{kk} g_{kk,i} (g^{ii} g_{ii,j} - g^{kk} g_{kk,j})$$

$$+ \frac{1}{4} g^{mm} g_{mm,i} (g^{ii} g_{ii,j} - g^{mm} g_{mm,j})$$

$$- \frac{1}{2} \partial_i (g^{ii} g_{ii,j}) - \frac{1}{2} \partial_j (g^{ii} g_{ii,i} + g^{kk} g_{kk,i} + g^{mm} g_{mm,i}) \tag{5.10}$$

where $j > i$, and $k, m \notin \{i, j\}$.

The equation (5.5) is a wave function, so all its solutions in polar coordinates can be constructed from product form solutions. The wave function in spherical coordinates is solved by separating all variables (r, θ, ψ, t). These product form solutions are products of spherical Hankel functions and spherical harmonics. The metric tensor in Schwarzschild's solution also has the product form. Finally, it is very difficult to see what else but a product form could zero all six R_{ij}, $j > i$, in (5.10). For these reasons I assume that the solution of vacuum space outside a point mass, i.e., when $R_{jj} = 0$, $j > i$, and $R = 0$, has a product form. For a product form

$$g_{jj} = A_{j0}(x^0) A_{j1}(x^1) A_{j2}(x^2) A_{j3}(x^3) \tag{5.11}$$

the function

$$y_{ij} = g^{ii} g_{ii,j}$$

is a function of x^j only and f_j in (5.9) depends only on x^j. Inserting (5.11) to R_{jj} gives

$$R_{jj} = f_j(x^j) - \sum_{\substack{i=0 \\ i\neq j}}^{4} \frac{A_{jj}}{A_{ij}} G_{ji}$$

where G_{ji} is a function of other coordinates than x^j. In order for the solution to be separating variables, we must be able to separate x^j from this equation.

It can be done in two ways, either A_{ij} is a constant times A_{jj}, or the terms G_{ji} disappear.

A stationary spherically symmetric solution has $A'_{j0} = A'_{j2} = A'_{j3} = 0$ for every j. This guarantees that $R_{ij} = 0$ for every case of $j > i$ and it also causes $G_{ji} = 0$ for all i and j for a product form solution. For this kind of solution holds:

$$g^{00}R_{00} = -\frac{1}{4}g^{11}g^{00}g_{00,1}(-g^{00}g_{00,1} + g^{11}g_{11,1} + g^{22}g_{22,1} + g^{33}g_{33,1}) \quad (5.12)$$

$$-\frac{1}{2}g^{00}\partial_1(g^{00}g_{00,1})$$

$$g^{22}R_{22} = -\frac{1}{4}g^{11}g^{22}g_{22,1}(g^{00}g_{00,1} + g^{11}g_{11,1} - g^{22}g_{22,1} + g^{33}g_{33,1})$$

$$-\frac{1}{2}g^{22}\partial_1(g^{22}g_{22,1})$$

$$g^{33}R_{33} = -\frac{1}{4}g^{11}g^{33}g_{33,1}(g^{00}g_{00,1} + g^{11}g_{11,1} + g^{22}g_{22,1} - g^{33}g_{33,1})$$

$$-\frac{1}{2}g^{33}\partial_1(g^{22}g_{22,1})$$

$$g^{11}R_{11} = \frac{1}{4}g^{11}g^{00}g_{00,1}(g^{11}g_{11,1} - g^{00}g_{00,1})$$

$$+\frac{1}{4}g^{11}g^{22}g_{22,1}(g^{11}g_{11,1} - g^{22}g_{22,1}) + \frac{1}{4}g^{11}g^{33}g_{33,1}(g^{11}g_{11,1} - g^{33}g_{33,1})$$

$$-\frac{1}{2}g^{11}\partial_1(g^{00}g_{00,1}) - \frac{1}{2}g^{11}\partial_1(g^{22}g_{22,1}) - \frac{1}{2}g^{11}\partial_1(g^{33}g_{33,1})$$

Summing these terms gives

$$R = -\frac{1}{2}g^{00}g_{00,1}g^{22}g_{22,1} \quad (5.13)$$

$$-\frac{1}{2}g^{00}g_{00,1}g^{33}g_{33,1} - \frac{1}{2}g^{22}g_{22,1}g^{33}g_{33,1}$$

$$-\frac{1}{2}(g^{00}+g^{11})\partial_1(g^{00}g_{00,1}) - \frac{1}{2}(g^{22}+g^{11})\partial_1(g^{22}g_{22,1}) - \frac{1}{2}(g^{33}+g^{11})\partial_1(g^{33}g_{33,1})$$

Taking the form for a gravitational field as in (5.2) and writing $A = \Phi^2$ the nonzero elements of the metric tensor are $g_{00} = A$, $g_{11} = -A$, $g_{22} = -r^2A$

and $g_{33} = -r^2\sin^2(\theta)A$. Assuming that $A = A(r) = A_{j1}(x^1)$ for every j the diagonal Ricci tensor entries from (5.12) are

$$R_{00} = \frac{1}{2}A''A^{-1} + \frac{1}{r}A'A^{-1} \tag{5.14}$$

$$R_{11} = -\frac{3}{2}A''A^{-1} + \frac{3}{2}(A')^{-2} + \frac{1}{r}A'A^{-1}$$

$$R_{22} = -\frac{1}{2}r^2A''A^{-1} - 2rA'A'^{-1}$$

$$R_{33} = \sin^2(\theta)R_{22}$$

and the Ricci scalar from (5.13) as

$$R = 3A''A^{-2} - \frac{3}{2}(A')^2A^{-3} + \frac{6}{r}A'A^{-2} \tag{5.15}$$

The equation $R = 0$ gives

$$A''A^{-\frac{1}{2}} - \frac{1}{2}(A')^2A^{-\frac{3}{2}} + \frac{2}{r}A'A^{-\frac{1}{2}} = 0$$

Denoting $y = A'A^{-1}$ the equation for Nordström's theory in the vacuum outside a point mass is

$$y' + \frac{1}{r}y = 0$$

Thus $y = kr^{-2}$ for some k and

$$\Phi(r)^2 = A = (b - \frac{k}{2r})^2$$

for some b. We get the field potential

$$\Phi(r) = b - \frac{k}{2r} \tag{5.16}$$

from which b and k can be identitied as $b = 0$ and $k/2 = GM$. Thus, $\Phi = \Phi(r)$ gives the Newtonian potential. This is what it should give since by (5.5) the equation $R = 0$ in this vacuum reduces to

$$\Box\Phi = 0$$

In spherical coordinates D'Alembertian is

$$\square\Phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Phi}{\partial\theta}\right)$$

$$+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\psi^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$$

The solution for $\Phi = \Phi(r)$ is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) = \frac{2}{r}\Phi' + \Phi'' = 0$$

Thus $\Phi' = kr^{-2}$ and $\Phi = -kr^{-1}$ giving the same result as (5.16).

The constant $k = GM$ but we for simplicity set it to one and insert $A_{j1} = r^{-2}$ to (5.12). The result is $R_{00} = -1$, $R_{11} = 1$, $R_{22} = R_{33} = 0$ and $R_{ij} = 0$ if $i \neq j$. This shows that Einstein's equations are not satisfied by a metric tensor g_{ab} as in (5.3). The diagonal Ricci tensor entries are not all zeros for g_{ab} derived from a field Φ. However, they do satisfy $R = 0$ in a vacuum outside a point mass. This means that if gravitation is a field, Einstein's equations are not satisfied and Nordström's equations are.

5.2. Ricci tensor entries in Nordström's field equation

In the vacuum the field equation of Nordström's second gravitation theory is the wave equation. The wave equation is linear and therefore linear combinations of product form solutions also fill it, but these linear combinations are usually not solutions to time dependent Nordström's field equations in spherical coordinates because Nordström's field equation is not linear outside the vacuum and this restricts the set of acceptable solutions.

The product form solutions to (5.2) are found by separating variables. We can look for product form solutions of the form

$$g_{00} = A, \quad g_{11} = -A, \quad g_{22} = -r^2A, \quad g_{33} = -r^2\sin^2(\theta)A$$

$$A = A_0(t)A_1(r)A_2(\theta)A_3(\psi)$$

From (5.10) we obtain

$$R_{0j} = \frac{1}{2}A'_j A_j^{-1} A'_0 A_0^{-1}, \qquad j = 1,2,3$$

$$R_{12} = \frac{1}{2}A'_2 A_2^{-1}\left(A'_1 A_1^{-1} + \frac{2}{r}\right) \qquad R_{13} = \frac{1}{2}A'_3 A_3^{-1} A'_1 A_1^{-1}$$

$$R_{23} = \frac{1}{2}A'_3 A_3^{-1}(A'_2 A_2^{-1} + 2\cot\theta)$$

From these expressions we see that for the time-dependent case the cross entries $R_{0,j}$, $j > 0$, are not zero for g_{ab} derived from a field Φ. These entries indicate that there is a flow of mass-energy and momentum. A physical interpretation for such a flow can for instance be a planet moving in an orbit around the sun: the planet slightly disturbs the gravitational field of the sun.

For a stationary field we derived the Newtonian potential $\Phi(r) = k/r$. If a field Φ is time dependent, the radial dependency is different. This is because when r is separated from θ, ψ, t a constant must be added, i.e.,

$$f(r) + g(\theta,\psi,t) = 0 \quad \text{gives} \quad f(r) = C \quad \text{and} \quad g = -C$$

Looking at the forms of R_{ii} and R for the time dependent case of Nordström's theory given below make it obvious that the solution to the equation $R = 0$ does not usually satisfy the equations $R_{aa} = 0$ for all a. (They may satisfy all equations, for instance if $\Phi = 0$.) The most general $\Phi = \Phi(r,\theta,\psi,t)$ has the following nonzero diagonal Ricci tensor entries:

$$R_{00} = \frac{1}{2}A^{-1}\left\{\frac{\partial^2 A}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 A}{\partial\theta^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 A}{\partial\phi^2} - 3\frac{\partial^2 A}{\partial t^2}\right\} \tag{5.17}$$

$$+\frac{1}{2}A^{-1}\left\{+\frac{2}{r}\frac{\partial A}{\partial r} + \frac{1}{r^2}\cot\theta\frac{\partial A}{\partial\theta}\right\} + \frac{3}{2}A^{-2}\left(\frac{\partial A}{\partial t}\right)^2$$

$$R_{11} = \frac{1}{2}A^{-1}\left\{-3\frac{\partial^2 A}{\partial r^2} - \frac{1}{r^2}\frac{\partial^2 A}{\partial\theta^2} - \frac{1}{r^2\sin^2\theta}\frac{\partial^2 A}{\partial\phi^2} + \frac{\partial^2 A}{\partial t^2}\right\}$$

$$+\frac{1}{2}A^{-1}\left\{-\frac{2}{r}\frac{\partial A}{\partial r} - \frac{1}{r^2}\cot\theta\frac{\partial A}{\partial\theta}\right\} + \frac{3}{2}A^{-2}\left(\frac{\partial A}{\partial r}\right)^2$$

$$R_{22} = \frac{1}{2}A^{-1}\left\{-r^2\frac{\partial^2 A}{\partial r^2} - 3\frac{\partial^2 A}{\partial \theta^2} - \frac{1}{sin^2\theta}\frac{\partial^2 A}{\partial \phi^2} + r^2\frac{\partial^2 A}{\partial t^2}\right\}$$

$$+\frac{1}{2}A^{-1}\left\{-4r\frac{\partial A}{\partial r} - \cot\theta\frac{\partial A}{\partial \theta}\right\} + \frac{3}{2}A^{-2}\left(\frac{\partial A}{\partial \theta}\right)^2 + 1$$

$$R_{33} = \frac{1}{2}A^{-1}\left\{-r^2\sin^2\theta\frac{\partial^2 A}{\partial r^2} - \sin^2\theta\frac{\partial^2 A}{\partial \theta^2} - 3\frac{\partial^2 A}{\partial \phi^2} + r^2\sin^2\theta\frac{\partial^2 A}{\partial t^2}\right\}$$

$$+\frac{1}{2}A^{-1}\left\{-4r\frac{\partial A}{\partial r} + 3\sin\theta\cos\theta\frac{\partial A}{\partial \theta}\right\} + \frac{3}{2}A^{-2}\left(\frac{\partial A}{\partial \phi}\right)^2 - \sin^2\theta$$

and the Ricci scalar is

$$R = 3A^{-3}\left\{\frac{\partial^2 A}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 A}{\partial \theta^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 A}{\partial \phi^2} - \frac{\partial^2 A}{\partial t^2} + \frac{2}{r}\frac{\partial A}{\partial r} + \frac{2}{r}cot\theta\frac{\partial A}{\partial \theta}\right\}$$

$$+3A^{-3}\left\{\frac{1}{2}A^{-1}\left((\frac{\partial A}{\partial t})^2 - (\frac{\partial A}{\partial r})^2 - \frac{1}{r^2}(\frac{\partial A}{\partial \theta})^2 - \frac{1}{r^2\sin^2\theta}(\frac{\partial A}{\partial \phi})^2\right)\right\}$$

In spherical coordinates D'Ambertian $\Box\Phi$ for $A = \Phi^2$ gives

$$\Box\Phi = -\frac{1}{2}A^{\frac{-1}{2}}\left\{\frac{2}{r}\frac{\partial A}{\partial r} + \frac{\partial^2 A}{\partial r^2} - \frac{1}{2}A^{-1}(\frac{\partial A}{\partial r})^2 - \frac{\partial^2 A}{\partial t^2} + \frac{1}{2}A^{-1}(\frac{\partial A}{\partial t})^2\right\}$$

$$-\frac{1}{2}A^{\frac{-1}{2}}\left\{+\frac{1}{r^2}\cot\theta\frac{\partial A}{\partial \theta} + \frac{1}{r^2}\frac{\partial^2 A}{\partial \theta^2} - \frac{1}{r^2}\frac{1}{2}A^{-1}(\frac{\partial A}{\partial \theta})^2\right\}$$

$$-\frac{1}{2}A^{\frac{-1}{2}}\left\{+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 A}{\partial \phi^2} - \frac{1}{r^2\sin^2\theta}\frac{1}{2}A^{-1}(\frac{\partial A}{\partial \phi})^2\right\}$$

Thus

$$R = -6A^{-\frac{3}{2}}\Box\Phi = -6\Phi^{-3}\Box\Phi$$

In Cartesian coordinates

$$g_{00} = \Phi^2 \quad , \quad g_{ii} = -\Phi^2 \quad , \quad g_{ab} = 0 \quad \text{if} \quad a \neq b$$

the diagonal elements of the Ricci curvature tensor are

$$R_{00} = -\Phi^{-1}\Box\Phi + \Phi^{-2}\sum_{j=1}^{3}(\frac{\partial \Phi}{\partial x^j})^2 + 3\Phi^{-2}(\frac{\partial \Phi}{\partial t})^2 - 2\Phi^{-1}\frac{\partial^2 \Phi}{\partial t^2}$$

and for $i = 1, 2, 3$

$$R_{ii} = -\Phi^{-1}\Box\Phi - \Phi^{-2}\sum_{j=1}^{3}(\frac{\partial\Phi}{\partial x^j})^2 + \Phi^{-2}(\frac{\partial\Phi}{\partial t})^2$$

$$-2\Phi^{-1}\frac{\partial^2\Phi}{\partial(x^i)^2} + 4\Phi^{-2}(\frac{\partial\Phi}{\partial x^i})^2$$

As $g^{ab} = \Phi^{-2}\eta^{ab}$ in Cartesian coordinates, we get

$$R = g^{ab}R_{ab} = -4\Phi^{-3}\Box\Phi + 4\Phi^{-4}\sum_{j=1}^{3}(\frac{\partial\Phi}{\partial x^j})^2 + 0$$

$$-2\Phi^{-3}\Box\Phi - 4\Phi^{-4}\sum_{j=1}^{3}(\frac{\partial\Phi}{\partial x^j})^2$$

that is, we get Equation (5.4)

$$R = -6\Phi^{-3}\Box\Phi$$

but every R_{aa} is not zero in a vacuum outside a point mass.

5.3. Does Nordström's field theory fail the classical tests?

Evaluation which of the field theories is correct is today left to the empirical tests of the General Relativity. Does Nordström's theory fail three of the four tests, or is it Einstein's theory that fails three out of four? See the Wikipedia entry [4] for a calculation that concludes that Nordström's theory fails three of the four tests.

I see certain problems in the calculations of [4]. It is stated that Nordström's second field theory comes from the Lagrangian

$$L = \frac{1}{8\pi}\eta^{ab}\Phi_{,a}\Phi_{,b} - \rho\Phi$$

but calculating

$$\frac{\partial L}{\partial\Phi} = -\rho$$

$$\frac{\partial L}{\partial \Phi_{,a}} = \frac{1}{4\pi}\eta^{aa}\Phi_{,a}$$

gives the Euler-Lagrange equations

$$\frac{\partial L}{\partial \Phi} - \sum_{\mu} \partial_{\mu} \frac{\partial L}{\partial \Phi_{,\mu}}$$

$$= -\rho - \frac{1}{4\pi}\sum_{\mu} \eta^{\mu\mu} \frac{\partial^2 \Phi}{\partial (x^{\mu})^2} = -\rho - \frac{1}{4\pi}\Box\Phi$$

and the Lagrangian produces the field equation of Nordström's first theory

$$\Box\Phi = -4\pi\rho$$

This is a minor issue since the second theory can be obtained from a very similar Lagrangian

$$L = \frac{1}{8\pi}\eta^{ab}\Phi_{,a}\Phi_{,b} - \frac{1}{4}\rho\Phi^4$$

but the calculations in [4] seem to use the geodesic Lagrangian, which is given as

$$L = \Phi^2 \eta_{ab}\dot{u}^a\dot{u}^b$$

and said to produce the equation of motion for Nordström's second theory

$$\Phi\dot{u}_a = -\Phi_{,a} - \dot{\Phi}u_a \tag{5.18}$$

However, the expression is not in a form of a Lagrangian. It can be obtained by inserting $\Phi\dot{u}_a = \Phi_{,a}$ into the Lagrangian

$$L = \Phi^2\eta^{ab}\Phi_{,a}\Phi_{,b} = \eta^{ab}\Phi\dot{u}_a\Phi\dot{u}_b$$

which after raising and lowering indices comes to

$$= \Phi^2\eta_{ab}\dot{u}^a\dot{u}^b$$

but if this is the way it is derived, it is for a stationary field, that is, $\dot{\Phi}u_a = 0$ in the equations of motion (5.18). Though is it correct to look at the stationary field in most of the tests of General Relativity (when Nordström's theory gives the Newtonian potential $\Phi = \Phi(r)$), the calculations in [4] take a time dependent solution for the wave function and derive properties from it, though the equations of motion seem to be for the stationary case.

I make different calculations and the conclusions in the subsections are different from those in [4]. I hope my calculations are more transparent than the ones in [4].

5.3.1 Frequency shift in gravitational fields

Gravitational redshift has been demonstrated in the Pound-Rebka experiment. This redshift is caused by the equivalence principle, which Nordström's theory satisfies, see [2]. The equations of motion of Nordström's second theory in a vacuum outside a point mass are Lorentz invariant forms of equations of motion in classical Newtonian theory as the potential is the Newtonian potential and the field equation is the wave equation. Thus (5.18) reduces in this case to

$$\dot{u}_a = -\Phi_{,a}$$

where

$$-\Phi_{,1} = -\frac{\partial \Phi}{\partial r} = -\frac{GM}{r^2}$$

The stationary case $\Phi = \Phi(r)$ describes the gravitational field in the Pound-Rebka experiment and the radial direction is the only relevant one. The radial acceleration is thus

$$\dot{u} = -\frac{GM}{r^2}$$

where $u = u_r$ is the radial velocity of a photon. The derivation is with respect to the proper time τ. In the special relativity

$$\tau = \sqrt{1 - \frac{u^2}{c^2}}dt \tag{5.19}$$

We can use this definition of proper time in this test as the proper time is actually not needed: the final result is wave lengths in the external time. From the metric tensor in Nordström's field theory follows that

$$\frac{dr}{dt} = \frac{cAdr}{Adt} = c$$

where the speed of light c is shown explicitly for clarity. Then

$$\frac{du}{d\tau}=\frac{du}{dr}\frac{dr}{dt}\frac{dt}{d\tau}=c\frac{1}{\sqrt{1-\frac{u^2}{c^2}}}\frac{du}{dr}$$

and

$$\int_{r_1}^{r_2}\left(-\frac{GM}{r^2}\right)dr=\int_{u_1}^{u_2}c\frac{1}{\sqrt{1-\frac{u^2}{c^2}}}du=\int_{u_1/c}^{u_2/c}c\frac{1}{\sqrt{1-y^2}}dy$$

giving

$$\frac{GM}{c^2}\frac{r_2-r_1}{r_2r_1}=\arcsin\frac{u_1}{c}-\arcsin\frac{u_2}{c}=\frac{u_2-u_1}{c\sqrt{1-\frac{u_1^2}{c^2}}}=\frac{\Delta u}{c\sqrt{1-\frac{u_1^2}{c^2}}}$$

The change of the speed u is here expressed as $\Delta u/\tau$, that is, the time is the proper time. In the external time the equation is

$$\frac{GM}{c^2}\frac{r_2-r_1}{r_2r_1}=\frac{\Delta u}{c}$$

Setting $r_2=R+h$, $r_1=R$ and approximating

$$\frac{GM}{c^2}\frac{h}{(R+h)R}\approx\frac{GM}{c^2}\frac{h}{R^2}$$

and changing to wave lengths

$$\frac{\Delta u}{c}=\frac{\Delta\lambda f}{\lambda f}=\frac{\Delta\lambda}{\lambda}$$

gives the redshift in the Pound-Rebka experiment

$$\frac{GM}{c^2}\frac{h}{R^2}=\frac{\Delta\lambda}{\lambda}$$

Einstein's field theory also fills the equivalence principle and passes this test if the special relativity definition is used for the proper time. In this definition proper time does not run at all for photons which travel with the speed of light. In the frame moving with photons the photons do not move at all. Thus, in that frame no space is moved and no time is used: the speed of light in the proper time is still c. In General Relativity there is another formula given in (5.20) for the proper time difference caused by a gravitational field. I leave it to a comment in the next section to consider what implications it has to the Pound-Rebka redshift for Schwarzschild's solution.

5.3.2 Shapiro time delay in gravitational fields

In the Shapiro time delay test a radar signal is sent from the Earth to another planet, like Venus, and echoed back to the Earth. A longer delay is measured if the sun is close to the path than if the sun is far. The expression of the additional delay contains three distances: the straight line connecting the Earth and the planet is divided into two parts, one of length x_e between the Earth and the connection point and one of length x_p between the planet and the connection point. The sun is on the distance d on the straight line from connection point orthogonal to the line from the Earth to the planet. The delay is

$$\Delta t \approx \frac{2GM}{c^3} \ln \frac{4x_p x_e}{d^2}$$

This delay is mainly a result of the variable speed of light in a gravitational field. The delay formula was derived from Schwarzshild's solution.

Light bends in gravitational fields but the formula for the Shapiro time delay does not have any parameters for the orbit: the orbit is hyperbolic (or at least very close to hyperbolic) and we would expect to see the parameters a and b of a hyperbole

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

but they do not appear. This is because the delay caused by bending of the path is much smaller than the delay caused by the gravitational redshift as in the Pound-Rebka experiment. We can ignore the bending of the path in the first order approximation and think about the path as a straight horizontal line. The sun is at the point $(0, d)$, the Earth is at $(-x_e, 0)$ and Venus at $(x_p, 0)$. The distance d is approximately sun's radius $d \approx 0.6957 * 10^9 m$. The distance x_e is almost the same as the distance of the Earth from the sun and can be taken as $x_e \approx 148 * 10^9 m$. The distance x_p is almost the distance of the planet from the sun. For Venus it is about $x_p \approx 108 * 10^9 m$. The distances x_e and x_p are approximations as the distances vary in the orbits, but they are quite sufficient for this test.

In the General Relativity theory there is the following formula for the proper time difference caused by a gravitational field

$$\Delta\tau = \int_{t_1}^{t_2} \frac{1}{c}\sqrt{g_{00}}dt \tag{5.20}$$

This formula is scaled in a specific way and the metric tensor must be rescaled in order to use the formula. Thus, for Nordström's theory in the vacuum

$$\sqrt{g_{00}} = \Phi(r) = -\frac{GM}{r}$$

and as it has the quality $\frac{m^2}{s^2}$, the expression for $\Delta\tau$ is

$$\Delta\tau = \int_{t_1}^{t_2} \frac{1}{c^2}\sqrt{g_{00}}dt$$

that is, otherwise we do not get seconds. In Schwarzschild's solution

$$\sqrt{g_{00}} = \sqrt{B(r)}$$

is a plain number and we have to use the formula as

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{g_{00}}dt$$

in order to get seconds.

We can calculate the proper time for Nordström's theory

$$\Delta\tau = \int_{t_1}^{t_2} \frac{1}{c^2}\Phi dt$$

and from the metric tensor $Adt = cAds$ for a line element ds of any path we get $dt = cds$ and

$$\Delta\tau = \int_{s_1}^{s_2} \frac{1}{c^3}\frac{GM}{r}ds.$$

The integral is easily calculated:

$$\int_0^x \frac{1}{r}dx = \int_0^x \frac{1}{\sqrt{1+\frac{x^2}{d^2}}}\frac{dx}{d}$$

$$= \int_0^{x/d} \frac{1}{\sqrt{1+y^2}}dy = \ln\left(\sqrt{1+\frac{x^2}{d^2}}+\frac{x}{d}\right)$$

$$\approx \ln(2\frac{x}{d})$$

as $\frac{x}{d} >> 1$. For any smooth f changing $x=-y$ shows that

$$\int_{-x}^0 f(x^2)dx = -\int_y^0 f(y^2)dy = \int_0^y f(y^2)dy$$

Thus

$$\int_{-x_e}^{x_p} \frac{1}{r}dx \approx \ln(2\frac{x_p}{d}) + \ln(2\frac{x_e}{d})$$

$$= \ln(4\frac{x_e x_p}{d^2})$$

The result for Nordström's theory for the proper time difference is

$$\Delta\tau = -\frac{GM}{c^3}\ln\left(\frac{4x_e x_p}{d^2}\right)$$

This $\Delta\tau$ is negative: the proper time goes forward slower than the external time. It means that light moves slower than c. That is, if light travels with the speed pc, $0 < p < 1$, light travels the distance $-x_e + x_p$ in the time $t = \frac{-x_e+x_p}{pc}$. The Shapiro time delay measured by an external clock is

$$\Delta t = \frac{-x_e + x_p}{c}\left(\frac{1}{p} - 1\right)$$

In the proper time in Nordström's theory light travels with the speed c. Thus, the proper time is $\tau = \frac{-x_e+x_p}{c}$ and the difference between the external time and proper time is

$$\Delta\tau = \tau - t = -\frac{-x_e + x_p}{c}\left(1 - \frac{1}{p}\right) = -\Delta t$$

For a roundtrip path we get the delay

$$\Delta t = 2\frac{GM}{c^3}\ln\left(\frac{4x_e x_p}{d^2}\right)$$

which is exactly the expression in the Shapiro time delay. Inserting numbers we get the roundtrip delay as $240\mu s$, which agrees with observations.

Let us calculate the Shapiro time delay for Schwarzschild's solution in a similar way. We start from

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{B}dt$$

Next we have to change the integration from time to a space variable. From the metric tensor of Schwarzschild's solution

$$\sqrt{B}dt = c\sqrt{A}dy \quad , \quad \sqrt{B}dt = dz$$

where the (y,z)-coordinates are Cartesian coordinates selected so that y is parallel to r and z is orthogonal to r. That is, the first expression we get by considering a move to a direction when $dr \neq 0$ and $d\theta = d\psi = 0$. The second expression corresponds to $d\theta \neq 0$ and $dr = d\psi = 0$. The reason for not using the polar coordinates is that $r^2 d\theta$ causes unnecessary complications in a simple calculation.

As we assume that the path is closely approximated by a horizontal line we can write dy and dz with the line element ds as

$$dy = \cos(\alpha)ds = \frac{x}{\sqrt{x^2+b^2}}ds$$

$$dz = \sin(\alpha)ds = \frac{b}{\sqrt{x^2+b^2}}ds$$

where α is the angle between the horizontal line and the line from (x,y) on the path to the sun. Thus

$$\Delta\tau = \frac{1}{c}\int_{s_1}^{s_2}\left(\frac{x}{\sqrt{x^2+b^2}}\frac{1}{\sqrt{1-\frac{2GM}{c^2 r}}} + \frac{b}{\sqrt{x^2+b^2}}\right)ds$$

Assuming that the path is a horizontal line and inserting

$$\frac{1}{\sqrt{1-\frac{2GM}{c^2r}}}\approx 1+\frac{GM}{c^2r}$$

the integral gives

$$\Delta\tau=\frac{1}{c}\int_{-x_e}^{x_p}\frac{x}{\sqrt{x^2+b^2}}dx$$

$$+\frac{1}{c}\frac{GM}{c^2}\int_{-x_e}^{x_p}\frac{x}{x^2+b^2}dx+\frac{b}{c}\int_{-x_e}^{x_p}\frac{1}{\sqrt{x^2+b^2}}dx$$

which is approximated by

$$\frac{b}{c}\int_{-x_e}^{x_p}\frac{\frac{x}{b}}{\sqrt{1+\frac{x^2}{b^2}}}\frac{dx}{b}$$

$$+\frac{GM}{c^3}\int_{-x_e}^{x_p}\frac{\frac{x}{b}}{1+\frac{x^2}{b^2}}\frac{dx}{b}+\frac{b}{c}\ln(4\frac{x_ex_p}{d^2})$$

$$\approx\frac{1}{c}(-x_e+x_p)+\frac{b}{c}\ln\left(\frac{4x_ex_p}{d^2}\right)+\frac{GM}{c^3}\ln\left(\frac{x_ex_p}{d^2}\right)$$

This $\Delta\tau$ does not agree with the Shapiro time delay and we notice that it is positive: the proper time goes forward faster than the external time. It means that light moves faster than c. The proper time takes into account the gravitational redshift also in the case of Schwarzschild's solution: it is the $\sqrt{g_{00}}$ term. Without this delay the speed of light in Scwarzschild's geometry would be even faster.

The reason for exceeding the speed of light is the geometry of the ball in Schwarzschild's geometry. The ratio of the space element to the time element gives the speed of light. In a flat Minkowski space in Cartesian coordinates the line element is

$$ds^2=\frac{1}{c^2}dt^2-dx^2-dy^2-dz^2$$

The speed c corresponds to moving the space distance dx in the time $\frac{1}{c}dt$. Likewise, in spherical coordinates moving dr in the time $\frac{1}{c}dt$ is moving with the speed c. In polar coordinates with θ we get the space element $r^2 d\theta$ but this still means moving with the speed of light, only the tangential space element is longer in polar coordinates.

When the geometry is changed as in Schwarzschild's solution, moving the distance of a radial space element $\sqrt{A}dr$ in the time element $\frac{\sqrt{B}}{c}dt$ means the speed $c\sqrt{A/B}$. As in Schwarzschild's solution AB is a constant, it means that in a gravitational field light moves faster than c. In proper time it is less, because of the redshift the speed in the proper time is $c\sqrt{A}$. This is still above c. For most of the time the path is mostly to the direction of the polar angle. The speed of light is also faster in this direction since the polar angle element is $r^2 dr$ and the time element is Bdt. Instead of $r^2 dr/dt = cr^2$, which means moving with the speed c, we get $r^2 dr/\sqrt{B}dt$. In the proper time it is $r^2 dr/dt$ but this already includes the redshift. Therefore there is no gravitational redshift in the polar angle direction in Schwarzschild's solution.

A comment on the Pound-Rebka experiment was promised. It is that if we use the General Relativity definition of proper time (5.20) instead of (5.19), Schwarzschild's solution gives a different redshift in the Pound-Rebka experiment than Nordström's theory: the movement is radial and the integration is over $c\sqrt{A}$. So, in fact, Einstein's theory fails the gravitational redshift test.

Notice also that for Schwarzschild's solution the calculation does not give the proper time difference but $\Delta\tau$ includes the one-way delay from the Earth to Venus. It is because the proper time formula is understood differently in Schwarzschild's solution. The logic in Schwarzschild's solution follows the geometric paradigm: in a flat Minkowski space we can think of the functions A and B in Schwarzschild's solution as having the value one. When there is a gravitational field, the field is

$$\Phi = \sqrt{g_{00}} = \sqrt{1 - \frac{2GM}{r}} \approx 1 - \frac{GM}{r}$$

the Newtonian potential added to the potential of a Minkowski space. In the field paradigm we do not think in this way: there is no gravitational potential in an empty Minkowski space and the proper time formula gives only the proper time difference, not the one-way delay.

As a conclusion, Nordström's theory passes the Shapiro time delay test, but Einstein's theory fails it.

5.3.3 The motion of Mercury

Influence of other planets had been carefully studied with Newtonian physics long before Einstein's time, even if everything was not yet known such as that the sun creates a cloud or a field around itself. The known effects did not explain the precession of the perihelion of Mercury and Einstein proposed a relativistic explanation for it.

I did not name this test the precession of the perihelion of Mercury because there is a bigger problem in the movement of planets: if the central force is a stationary Newtonian gravitation force, an elliptic orbit is not possible because it violates conservation of energy. A direct calculation from equations of motion confirms this conclusion as will be shown in what follows. In reality, planets circulate around the sun on ellitic orbits or at least very close to elliptic orbits. This means that there must be some mechanism by which energy is lost so that the planes do not escape to the space. Failing to explain this issues should mean failing the test.

Let us consider the movement of planets around the sun in a fully classical way. By conservation of the angular momentum and momentum the orbits of two masses m_1 and m_2 attracted by a central force can only be circles, ellipses or hyberbolas around the center of mass in a system where the only force is a central force. The result does not require that the central force has the $\frac{1}{r^2}$ dependency from the distance. However, this result does not yet imply that all of these solutions satisfy other concervation laws when the central force has a particular form. Indeed, the elliptic orbit does not satisfy conservation of energy for a stationary Newtonian gravitational force.

To see the energy problem let us take two masses m_1 being Mercury and m_2 the sun. The center of mass is in the line connecting the masses. The distance from m_i to the center of mass being r_i. The center of mass is at the point where $r_2 = r_1 m_1/m_2$.

The two-body system can be modelled in such a way that there is a stationary central Newtonian force in the center of mass. The coordinates can be so selected that the center of mass does not move. The movement is in two dimensions only and we need two coordinates: r and θ with the origin at the center of mass. The velocities v_i of the masses can be divided into radial and angular components $v_{i,r}$ and $v_{i,\theta}$. The masses move symmetrically around the center of mass as it stays fixed. The velocities in the θ direction must satisfy

$$v_{2,\theta} = \frac{m_i^2}{m_2^2} v_{1,\theta}$$

because the center of mass stays fixed. This is conservation of the angular momentum. The velocities in the r direction must also satisfy

$$v_{2,r} = \frac{m_i^2}{m_2^2} v_{1,r}$$

because the total momentum must be zero if the center of mass does not move. The total kinetic energy is

$$E_k = \frac{1}{2}\left(m_1 v_{1,r}^2 + m_2 v_{2,r}^2 + m_1 v_{1,\theta}^2 + m_2 v_{2,\theta}^2\right)$$

as r and θ are orthogonal. We get

$$E_k = \frac{1}{2}\frac{m_1}{m_2}(m_1 + m_2)(v_{1,r}^2 + v_{1,\theta}^2)$$

By Kepler's law of areas, which follows from the conservation of the angular momentum,

$$v_{1,\theta} = \frac{r_{1,min}}{r_1} v_{1,\theta,max}$$

where $r_{1,min}$ is the minimum distance of m_1 from the center of mass and $v_{1,\theta,max}$ is the tangential velocity m_1 has at this distance. The velocity is on the θ-direction as the radial velocity $v_{1,r,max}$ is zero at the minimum distance. We get

$$E_k(r) = \frac{1}{2}\frac{m_1}{m_2}(m_1 + m_2)v_{1,r}^2 + \frac{1}{2}\frac{m_1}{m_2}(m_1 + m_2)\left(\frac{r_{1,min}^2}{r_1^2} - 1\right) v_{1,\theta,max}^2$$

The difference between $E_k(r_1)$ and $E_k(r_{1,min})$ is

$$\Delta E_k = E_k(r_{1,min}) - E_k(r_1)$$

$$= \frac{1}{2}\frac{m_1}{m_2}(m_1+m_2)v_{1,r}^2 + \frac{1}{2}\frac{m_1}{m_2}(m_1+m_2)\left(\frac{r_{1,min}^2}{r_1^2} - 1\right)v_{1,\theta,max}^2$$

The masses are at the opposite sides of the center of mass. Their distance is $r_1 + r_2 = r_1(m_1+m_2)/m_2$. The Newtonian gravitational force between them is

$$F = m_1 m_2 G \frac{1}{(r_1+r_2)^2} = m_1 m_2 G \frac{m_2}{m_1+m_2}\frac{1}{r_1^2}$$

where G is the gravitational constant. The difference in gravitational potential energy between the situations when the masses are at (r_1, r_2) and at $(r_{1,min}, r_{2,min})$ is

$$\Delta E_p = -\frac{m_1 m_2^2}{m_1+m_2} G \left(\frac{1}{r_1} - \frac{1}{r_{1,min}}\right)$$

By conservation of energy $\Delta E_k + \Delta E_p = 0$, thus

$$\frac{1}{2}\frac{m_1}{m_2}(m_1+m_2)v_{1,r}^2 + \frac{1}{2}\frac{m_1}{m_2}(m_1+m_2)\left(\frac{r_{1,min}^2}{r_1^2} - 1\right)v_{1,\theta,max}^2$$

$$+\frac{m_1 m_2^2}{m_1+m_2} G \left(\frac{1}{r_1} - \frac{1}{r_{1,min}}\right)$$

Thus

$$v_{1,r}^2 = -\left(\frac{r_{1,min}^2}{r_1^2} - 1\right)v_{1,\theta,max}^2 + 2\frac{m_1 m_2^3}{(m_1^2+m_2)^2} G \left(\frac{1}{r_1} - \frac{1}{r_{1,min}}\right) \quad (5.21)$$

Derivating with respect to r_1 and setting $\frac{\partial v_{1,r}}{dr_1} = 0$ at $r_1 = r_{1,min}$ gives

$$0 = 2\frac{r_{1,min}^2}{r_1^3}v_{1,\theta,max}^2 - 2\frac{m_1 m_2^3}{(m_1^2+m_2)^2} G \frac{1}{r_{1,min}^2}$$

as $v_{1,r}$ is zero at $r_{1,min}$. Thus

$$v_{1,\theta,max}^2 = \frac{m_2^3}{(m_1^2+m_2)^2} G \frac{1}{r_{1,min}}$$

Let us notice as a check-up that this equation gives the centrifugal force as the left side of

$$m_1 v_{1,\theta,max}^2 \frac{1}{r_{1,min}} = m_1 m_2 G \frac{1}{(r_{1,min}+r_{2,min})^2}$$

which is a correct formula as the rotation is around the central point at the distance $r_{1,min}$ from m_1 and the gravitation force is between the masses having a distance $r_{1,min}+r_{2,min}$ between them.

Inserting this expression to the equation (5.21) of $v_{1,r}^2$ yields

$$v_{1,r}^2 = 2\frac{m_1 m_2^3}{(m_1^2+m_2)^2} G \frac{1}{-r_{1,min} r^2}(r_1 - r_{1,min})r^2 \qquad (5.22)$$

There is a double zero at $r_1 = r_{1,min}$. That means that m_1 approaches m_2, gets to the minimum and then distances from m_2. It does not have another zero at $r_1 = r_{1,max}$ and thus an elliptic orbit is not possible. The orbit is either a circle, and then $r_1 = r_{1,min}$ all the time and $v_{1,r} = 0$ for every r, or the orbit is a hyperbole. In order to get an elliptic orbit we need a different central force or other forces. Some mechanism is necessary for reducing energy so that the radial velocity can have two zeros.

Let us check this result in another way. We can by a direct calculation see if the equation of motion for m_1 can be an ellipse if there is a central Newtonian force at the focal point of the ellipse.

Let us take an ellipse

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

At (x_0, y_0) the ellipse has the tangent

$$\frac{x_0}{a^2}x + \frac{y_o}{b^2}y = 1$$

The rotation

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{c} \begin{bmatrix} a^2 y_0 & -b^2 x_0 \\ b^2 x_0 & a^2 y_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad c = \sqrt{a^4 y_0 + b^4 x_0}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{c} \begin{bmatrix} a^2 y_0 & b^2 x_0 \\ -b^2 x_0 & a^2 y_0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

takes z_2 to a line parallel to the tangent of the ellipse and z_1 is orthogonal to this line. The rotation has the determinant one and therefore does not change the distances. In coordinates (z_1, z_2) the equation of ellipse is

$$z_2^2 - 2hz_1z_2 + ez_1^2 - m = 0$$

where

$$h = x_0 y_0 \frac{a^2 b^2 x_p^2}{b^6 x_0^2 + a^6 y_0^2}$$

$$e = \frac{a^4 b^4}{b^6 x_0^2 + a^6 y_0^2}$$

$$m = \frac{(a^4 y_0^2 + b^4 x_0^2) a^2 b^2}{b^6 x_0^2 + a^6 y_0^2}$$

$$x_p = \sqrt{a^2 - b^2}$$

The rotation takes (x_0, y_0), $x_0 > 0$, $y_0 > 0$, to (z_{10}, z_{20}) where

$$z_{10} = \frac{1}{c} x_0 y_0 x_p^2$$

$$z_{20} = \frac{1}{c} a^2 b^2$$

Solving the elliptic equation gives

$$z_2'(z_1) = \frac{dz_2}{dz_1} = h + (h^2 - e) z_1 (z_2 - hz_1)^{-1}$$

$$z_2''(z_1) = \frac{h^2 - e}{z_2 - hz_1} \left(1 - \frac{(h^2 - e) z_1^2}{(z_2 - hz_1)^2} \right)$$

A calculation shows that

$$\frac{h^2 - e}{z_{20} - hz_{10}} = -\frac{ca^2b^2}{b^6x_0^2 + a^6y_0^2}$$

$$\frac{z_{10}^2}{z_{20} - hz_{10}} = \frac{1}{c}\frac{x_0^2y_0^2x_p^4}{a^2b^2}\frac{b^6x_0^2 + a^6y_0^2}{b^6x_0^2 + a^6y_0^2 - x_0^2y_0^2x_p^4}$$

Thus

$$1 - \frac{(h^2 - e)z_{10}^2}{(z_{20} - hz_{10})^2} = \frac{b^6x_0^2 + a^6y_0^2}{b^6x_0^2 + a^6y_0^2 - x_0^2y_0^2x_p^4}$$

and

$$z_2''(z_{10}) = -\frac{ca^2b^2}{b^6x_0^2 + a^6y_0^2 - x_0^2y_0^2x_p^4}$$

The left focal point is in the point $(-x_p, 0)$ in (x, y)-coordinates. Let r be the distance from the focal point to (x_0, y_0)

$$r^2 = (x_0 + x_p)^2 + y_0^2$$

Thus, it is also the distance between the focal point in coordinates (z_1, z_2) and (z_{10}, z_{20}). We mark the focal point in coordinates (z_1, z_2) by (z_{1p}, z_{2p}). The focal point is at

$$z_{1p} = -\frac{1}{c}a^2y_0x_p$$

$$z_{2p} = -\frac{1}{c}b^2x_0x_p$$

We can express x_0 as a function of r by using $y_0^2 = b^{-2}(1 - x_0^2/a^2)$. The result is

$$x_0 = \frac{a}{x_p}(r - a)$$

As $x_0 > 0, y_0 > 0$ we have $r > a$. Likewise we solve

$$y_0 = b^2(1 - (r - a)^2x_p^{-2})$$

$$c = ab\sqrt{a^2 - (r - a)^2}$$

Inserting these to the expression of $z_2''(z_{10})$ we get

$$z_2''(z_{20}) = -\frac{ab\sqrt{a^2-(r-a)^2}}{a^4+2b^2(r-a)^2+(r-a)^4}$$

Let there be a point mass in the focal point and let it exercise gravitation force to the mass m_1 moving on the elliptic orbit. The sun is not exactly in the focal point, but we can take the mass as the mass of the center of mass and that is in the focal point. The angle θ between the horizontal line ($z_2 = 0$) in the (z_1, z_2)-coordinates and the line connecting (z_{1p}, z_{2p}) to (z_{10}, z_{20}) has the tangent

$$\tan\theta = \frac{-z_{2p}+z_{20}}{-z_{1p}+z_{10}}$$

Notice that $z_{1p} < 0$ and $z_{2p} < 0$ because $x_0 > 0, y_0 > 0$. Inserting expressions of these points as functions of x_0 and y_0 we get

$$\tan\theta = \frac{b^2}{y_0 x_p} = \frac{b}{\sqrt{a^2-(r-a)^2}}$$

The gravitation force F between the planet with mass m_1 and the center of mass can be divided into two components: F_{tan} tangential to the orbit (that is, parallel to the z_1-axis) and F_{ort} orthogonal to the tangent (that is, parallel to the z_2-axis). The z_2-axis does not point from (z_{10}, z_{20}) to the focal point. We have to take a projection

$$F_{ort} = F\sin\theta \quad , \quad F_{tan} = F\sin\theta$$

The equation of motion in the orthogonal direction is that the acceleration along the z_2-axis causes the displacement

$$\frac{1}{2}z_2''(z_{10})(dz_1)^2 = \frac{1}{2}a_{ort}(dt)^2$$

that is

$$z_2''(z_{10})(dz_1)^2 = \frac{F_{ort}}{m_1}(dt)^2$$

$$z_2''(z_{10})\frac{dz_1}{dt}^2 \sin^{-1}\theta = \frac{F}{m_1} = \frac{mG}{r^2} \tag{5.23}$$

where m is the equivalent mass that we should use for the center of mass. As

$$F = \frac{m_1 m_2 G}{(r + r_2)^2}$$

and $r_2 = rm_1/m_2$ is the distance of the sun from the center of mass we get

$$F = \frac{m_1}{r^2} m_2 \left(\frac{m_2}{m_1 + m_2}\right)^2$$

Thus, $m = m_2 \left(\frac{m_2}{m_1+m_2}\right)^2$. It is practically m_2 for the sun and Mercury. In case we believe that we know the velocity

$$v(r) = \frac{dz_1}{dt}$$

from conservation laws, we can insert the expression here. But this should not be done without considerations since the previous argument shows that the conservation of energy does not allow an elliptic orbit.

It is better to solve the velocity from tangential acceleration. The tangential velocity $v(z_1)$ is the velocity of the mass m_1 at z_{10} as the mass cannot have velocity orthogonal to the tangent of its orbit. In the tangential direction the equation of motion is

$$v(z_1) = v(z_{10}) + v'(z_{10})dz_1$$

$$v'(z_{10})dz_1 = \frac{F_{tan}}{m}dt = \cos(\theta)\frac{F}{m}dt$$

Since

$$v'(z_{10})\frac{dz_1}{dt} = v'(z_{10})v(z_{10})$$

and

$$v'(z_1)v(z_1) = \frac{1}{2}\frac{d}{dt}v(z_1)$$

we get an equation

$$v(z_{10})^2 = 2\int_{z_{1s}}^{z_{10}} \frac{F(z_1)}{m}\cos(\theta)dz_1$$

where the lower bound z_{1s} can be selected in a suitable way. The lower bound only changes the initial value of the tangential velocity and does not affect the coefficients of the Fourier series of the velocity.

The equation (5.23) for the orthogonal direction gave

$$v(z_{10})^2 = z''(z_{10})^{-1}\sin(\theta)\frac{F(z_1)}{m} \tag{5.24}$$

Let us write

$$B(r) = z''(z_{10})\sin(\theta)^{-1} = \frac{a(a^2-(r-a)^2)}{a^4+4b^2(r-a)^2+(r-a)^4}$$

and notice that as dz_1 is in the horizontal direction in the (z_1, z_2) coordinates and the focal point is down and left of (z_10, z_{20}) in angle θ with the horizontal level

$$dz_1 = \cos(\theta)dr$$

This gives the equation

$$B(r)^{-1}F(r) = 2\int_{z_{1s}}^{z_{10}} F(r)\cos^2(\theta(r))dr \tag{5.25}$$

Assuming that z_{1s} is sufficiently small, $z_{1s} > z_{10} > 0$, the value

$$\epsilon = \frac{r-a}{a}$$

is small and we can sufficiently well evaluate the sides of this equation as power series of ϵ to some chosen degree. Changing $r - a = a\epsilon$ gives

$$B(r) = \frac{1}{a}\frac{1-\epsilon^2}{1+2\frac{b^2}{a^2}\epsilon^2+\epsilon^4}$$

$$F(r) = mG\frac{1}{r^2} = mG\frac{1}{a^2(1+\epsilon)^2}$$

Thus

$$B(r)^{-1}F(r) = amG\frac{1+2\frac{b^2}{a^2}\epsilon^2+\epsilon^4}{1-\epsilon^4}$$

$$= amG(1 + 2\frac{b^2}{a^2}\epsilon^2 + \epsilon^4)(1 + \epsilon^4) + O(\epsilon^8)$$

$$= amG(1 + 2\frac{b^2}{a^2}\epsilon^2 + 2\epsilon^4 + 2\frac{b^2}{a^2}\epsilon^6) + O(\epsilon^8) \qquad (5.26)$$

In the right side of the equation (5.25) we have

$$\cos(\theta) = \frac{\sqrt{x_p^2 - (r-a)^2}}{\sqrt{b^2 + x_p^2 - (r-a)^2}}$$

Therefore

$$\cos^2(\theta) = \frac{1 - \epsilon^2 - \frac{b^2}{a^2}}{1 - \epsilon^2}$$

and

$$F(r)\cos^2(\theta) = \frac{mG}{a^2}\frac{1 - \epsilon^2 - \frac{b^2}{a^2}}{1 - \epsilon^4}$$

$$= \frac{mG}{a^2}\frac{1 - \frac{b^2}{a^2} - \epsilon^2}{1 + \epsilon^4} + O(\epsilon^8)$$

$$= \frac{mG}{a^2}(1 - \frac{b^2}{a^2} - \epsilon^2 + (1 - \frac{b^2}{a^2})\epsilon^4 - \epsilon^6) + O(\epsilon^8)$$

There remains the integration

$$2\int_{z_{1s}}^{z_{10}} F(r)\cos^2(\theta(r))dr$$

$$= 2\frac{mG}{a^2}(C - \frac{b^2}{a^2}\epsilon - \frac{1}{3}\epsilon^3 + \frac{1}{5}(1 - \frac{b^2}{a^2})\epsilon^5 - \frac{1}{7}\epsilon^7) + O(\epsilon^8) \qquad (5.27)$$

where C is some constant and it includes the initial value at z_{1s}.

Equation (5.25) is not filled: (5.26) and (5.27) do not match in powers of ϵ. The elliptic orbit is not a solution to a two body problem with a Newtonian gravitation force in this fully classical calculation. However, if we set $a = b = R$, which implies that $x_p = 0$, $r = a$, $\sin(\theta) = 1$, $\cos(\theta) = 0$, a solution is obtained: $v_{1,r}$ is constant, $\epsilon = 0$, (5.25) reduces to $aMG=$ constant and (5.24) reduces to

$$v(z_{10})^2 = a\frac{mG}{r^2} = \frac{mG}{r}$$

which is the velocity of a mass on a circular orbit. Thus, a circle is a solution in this calculation. Also the energy calculation allows a circular orbit as for a circle $r_1 = r_{1,min}$ and the radial velocity is zero for all times. In both calculations the center of the circle is the center of mass.

An ellipse is changed into a hyperbole by replacing b by ib. The formulae for an ellipse give formulae for a hyperbole if b^2 is replaced by $-b^2$. The calculation for an ellipse shows that a stationary Newtonian central force does not produce a mathematical hyperbole, in the same way as it cannot produce a mathematical elliptic orbit. But there is a difference: an orbit very similar to a hyperbole is possible because it does not contradict energy conservation: only one double root for $v_{1,r}^2$ agrees well with the hyperbolic orbit.

Kepler said that planets follow an elliptical orbit with the sun (almost) at one focal point and that the area law holds. These statements are correct, but the mathematical explanation of Kepler's laws for the orbits of planets by the conservation of the angular momentum and the momentum ignores the problem there is with conservation of energy. The elliptic orbit cannot be produced by a single stationary Newtonian central force.

A friction force would make an elliptic orbit possible, but there is little friction in space. The double root (5.22) can be broken into two roots also by modifying the gravitational potential. Let us see if Schwarzschild's solution can solve the energy problem. The force in Schwarzschild's solution is slightly larger than Newtonian gravitation force. We can change the gravitational potential to

$$E_p = -\frac{k}{r} + \frac{\alpha}{r^2}$$

For one value of α this gives the gravitational potential in Schwarzschild's solution.

The term $(r_1 - r_{1,min})^2$ in Equation (5.22) changes to the form

$$r_1^2 - \frac{2r_{1,min}}{1 + 2\alpha r_{1,min}^{-1}} r_1 + \frac{r_{1,min}(r_{1,min} - 2\alpha)}{1 + 2\alpha r_{1,min}^{-1}} = 0$$

In order to get two roots, $r_{1,min}$ and $r_{1,max}$, the equation must equal

$$r_1^2 - (r_{1,min} + r_{1,max})r_1 + r_{1,min}r_{1,max} = 0$$

Matching the parameters yields

$$\alpha = -\frac{r_{1,min}(r_{1,max} - r_{1,min})}{2(r_{1,min} + r_{1,max})}$$

For the orbit of Mercury $r_{1,min} \approx 45.9 * 10^9 m$ and $r_{1,max} \approx 69.9 * 10^9 m$. Inserting these values we get

$$\alpha \approx -4.7$$

This is not the value proposed by Schwarzschild's solution, thus it cannot explain the elliptic orbit of Mercury.

I will not try to calculate any predictions for Mercury's orbit from Nordström's second theory in this article as I consider it too difficult. I will only briefly discuss the problematics quantitatively. The suggestion is that when Mercury is close to the sun it slightly disturbs the stationary field $\Phi(r)$ and forces a time dependent solution $\Phi(r, \theta, t)$ (or even $\Phi(r, \theta, \psi, t)$). The wave equation is still almost $R = 0$ but the entries $R_{0,j}$, $j > 0$, are not zeros. The potential Φ has a slightly different r dependency than kr^{-1} coming from the constant needed to separate r from the other coordinates in the solution of the wave equation. As some energy is used for waves, less energy is left to the gravitational potential and the orbits of planets can be ellipses. The problem in calculating this is that if one planet disturbs the sun's potential, they all do, and the multibody problem becomes difficult.

Nordström's theory does not fail this test: the result is inconclusive. Einstein's theory fails this test as Schwarzschild's solution does not give an elliptic orbit.

5.3.4 Bending of light from stars by the Sun

Light bends in Nordström's theory: Nordström accepted Einstein's special relativity and considered light to have mass and the speed of light c to be the maximal velocity. Consequently, light in Nordström's theory behaves as a test mass and is attracted by gravity. Nordström's theory gives the Newtonian potential for the vacuum, $\Phi = \Phi(r)$. This stationary gravitational field of the sun may be disturbed by planets as suggested in 3.3, but the path is at least very close to a hyperbole. I consider calculating the amount of the bending of the light too difficult to be done in this article. The result of this test remains inconclusive for Nordström's theory. What I can do is to discuss the problematics from a theoretical point of view.

The argument in [4] that that light does not bend in gravitational fields in Nordström's field theory is based on the following reasoning: Electro-magnetic fields in Einstein's theory are described by a stress-energy tensor which is traceless. Nordström's field equation can be expressed in the geometric form (5). In that form T on the right side is the trace of T_{ab}. Consequently, light does not bend in Nordström's theory.

The caveat in this argument is that T_{ab} in Nordström's theory was originally not the T_{ab} in Einstein's theory, though in the last version Nordström accepted Einstein's proposals. The right side in Nordström's theory was $4\pi\rho$ or $4\pi T_{\mathrm{matter}}$ and this entity contains all mass-energy of the system, including mass-energy of electro-magnetic fields. We cannot assume that the T_{ab} in Nordström's theory are exactly the same as in Einstein's theory.

The argument in [4] can be turned against Einstein's theory: assuming that Nordström's theory correctly describes gravitation, the entity T must include the mass-energy of electro-magnetic fields. If T is the trace of T_{ab} from Einstein's theory, then T does not include the mass-energy electro-magnetic fields as it should. Therefore T_{ab} in Einstein's theory must be wrong.

Nordström's theory does not consider the problem how electro-magnetic fields are included to the field equation. According to [2] Laue discarded Nordström's theory because he could not couple electro-magnetic fields into it, but maybe this should be reconsidered. A time dependent Φ in Nordström's theory causes R_{ij}, $j \neq i$, to differ from zero. These elements may couple electro-magnetic fields to Φ in Nordström's theory even though the field equation is only the trace. Nordström's theory did not say anything of

R_{ij}, $j \neq i$, but the theory can be augmented by adding the requirements for these cross entries from Einstein's theory. Einstein's theory may be incorrect in requiring $R_{ii} = 0$, but it can pose correct requirements for the cross entries.

5.4. Conclusions

My arguments for questioning the superiority of Einstein's field theory over Nordström's second gravitation theory are three:

1) Nordström's theory is an ordinary field theory, not a geometry. Gravitation is a field in this theory and in this respect similar to other interactions, potentially making unification of the interactions easier. Einstein's theory is a geometry and the geometry in Schwarzschild's solution is not even quasiregular to the Euclidian geometry. The result is that in a gravitational field the speed of light is exceeded even if only very slightly. Light can go slower than c in Nordström's theory as the gravitational redshift shows, but if light moves faster than c the theory contradicts special relativity.

2) I do not see any strong reasons for Einstein's requirement that each R_{ab} be zero. In the Newtonian potential $\Phi = \Phi(r)$ in Nordström's theory for a stationary spherically symmetric field the elements R_{aa} are not all zero. This issue can be solved by changing the definition of T_{ab}: in the vacuum around a point mass the gravitational field has energy and it should be reflected somewhere, such as in nonzero elements R_{aa}.

Consider a point mass and the vacuum outside it. Placing the point mass in the origin of spherical coordinates there is a (static) gravitational field in this vacuum and the field has energy. A spherically symmetric static field does not depend on the polar angle θ, the azimuthal angle ψ or on the time t. As a result the Ricci curvature tensor entries R_{ab}, $a \neq b$, are zero and we may conclude that T_{ab}, $a \neq b$, must be zero in such a vacuum. However, the diagonal elements R_{aa} do not disappear for a metric tensor g_{ab} as in (5.3). Therefore T_{aa} are not all zero in Nordström's theory. I suggest that in Nordström's theory in this vacuum case $T_{ab} = 0$ for $a \neq 0$ and T_{aa}

fulfill the condition

$$g^{00}T_{00} = -g^{11}T_{11} - g^{22}T_{22} - g^{33}T_{33}$$

If there is mass-energy in the system, there is a similar difference between the stress-energy tensors of Nordström's and Einstein's theory: the diagonal elements T_{aa} are different as the energy of the gravitational field is included in Nordström's T_{ab} and is missing in Einstein's but the trace T is the same for both theories. This way of understanding T describes the ideas of Nordström's field theory as an extension of Newtonian gravitation.

3) Nordström's second field theory does not fail any of the four tests of General Relativity. It passes the redshift and Shapiro time delay tests and remains inconclusive in the Mercury and light bending tests. Einstein's theory fails the Shapiro time delay test and if the proper time is defined as in General Relativity it also fails the redshift test. Additionally, Einstein's theory fails to explain the motion of Mercury because Schwarzschild's solution does not give an elliptic orbit.

In the 1980ies I asked my supervisor for a mathematical topic with physical connections and I was given a geometric topic on quasiregular mappings between low-dimensional manifolds. The supervisor mentioned that Schwarzschild's solution is very odd: physical fields are images of conformal mappings but the ball in Schwarzschild's solution is not even quasiregular to our ball. I read at that time a book of black holes [5] and concluded that physics apparently can be consistently built on Schwarzschild's solution and did not look at the issue deeper, fortunately, as trying to pass a paper showing Einstein's field theory wrong would hardly have been accepted as a Ph.D. thesis. I made the thesis on quasiregular mappings between closed orientable 3-manifolds. But now, since I'm retired, I can investigate the geometry of General Relativity and there does seem to be some problems associated with it.

If Nordström's second field theory turns out to be the correct theory for gravitation it has some implications. For instance, the discovery of gravitational waves has been recently questioned, see [6]. In Einstein's theory gravitational waves are caused by the Weyl tensor and the way of finding the waves used patterns for waves derived from Einstein's theory. If the correct theory is Nordström's second theory, then the patters are different: one should look for gravitational waves from the Ricci tensor elements R_{ab}, $a \neq b$.

References to chapter 5:

[1] G. Nordström, Phys. Zeit. 13,1126 (1912); G. Nordström, Ann. d. Phys. 40, 856 (1913); G. Nordström, Ann. d. Phys. 42, 533 (1913); A. Einstein and A. D. Fokker, Ann. d. Phys. 44, 321 (1914); A. Einstein, Phys. Zeit. 14, 1249 (1914).

[2] J.D. Norton, in The Genesis of General Relativity Vol.3: Theories of Gravitation in the Twilight of Classical Physics. Part I. , Jörgen Renn (ed.) Kluwer Academic Publishers (2005),

www.pitt.edu/jnorton/papers/Nordstroem.pdf

[3] N. Deruelle, Nordström's scalar theory of gravitation and the equivalence principle. arXiv:1104.4608, 2011.

[4] The current (Dec 2 2018) Wikipedia page on Nordström's gravitational theory:

https://en.wikipedia.org/wiki/Nordström's_theory_of_gravitation.

[5] K. S. Thorne, R. H. Price and D.A. MacDonald (eds), *Black Holes: The Membrane Paradigm.* Yale University Press, 1986.

[6] M. Brooks, Did we really find gravitational waves: breakthrough physics result questioned. New Scientist, 3 Nov 2018.

6. Quantization of gravity

Before Einstein presented his General Relativity Theory (GR) Gunnar Nordström tried to formulate a relativistic scalar theory of gravitation. His starting point were two equations from Newton's gravitation theory. The first was the equation of motion:

$$\frac{d}{dt}u_a = -\partial_a \phi \tag{6.1}$$

which follows directly by combining $ma = F$ and $F = -\nabla\phi$. In (6.1) the function $u(t, \bar{r})$ is the velocity of the test mass. The second equation was the field equation

$$\nabla\phi = 4\pi G\rho \tag{6.2}$$

where G is the gravitation constant and ρ is mass density. This field equation can be explained by setting the gravitation potential as

$$\phi = \phi(r) = -\frac{G\rho}{r} \tag{6.3}$$

where the mass density ρ is constant. Then, as ϕ depends only on r

$$\Lambda\phi = \nabla\cdot\nabla\phi = r^2\frac{d}{dr}\left[r^2\frac{d}{dr}\phi\right] = r^2\frac{d}{dr}\left[r^2\frac{d}{dr}\left(-\frac{G\rho}{r}\right)\right] = \frac{d}{dr}\left[G\rho\right] = 0. \tag{6.4}$$

The equation (6.2) is a generalization of (6.4). It may not be so obvious that if we let the mass density depend on r, $\rho = \rho(r)$, the function ρ in (6.2) and (6.3) are not the same function. Let

$$\phi = \phi(r) = -\frac{G\rho_1(r)}{r} \tag{6.5}$$

then

$$\Delta\phi = r^2\frac{d}{dr}\left[r^2\frac{d}{dr}\phi\right] = r^2\frac{d}{dr}\left[r^2\frac{d}{dr}\left(-\frac{G\rho_1(r)}{r}\right)\right] = -G\frac{1}{r}\rho_1''(r) \tag{6.6}$$

where ρ_1'' is the second derivative of $\rho_1(r)$. Thus, (6.2) implies that

$$\rho_1'' = 4\pi\rho(r). \tag{6.7}$$

Using the same letter in (6.2) and (6.3) can cause confusion. Nordström tried to find a relativistic form that generalizes the equation of motion (6.1) and the field equation (6.2). Abrams had already suggested

$$\dot{u}_a = -\partial_a \phi \tag{6.8}$$

for the equation of motion, but it did not work. Here the dot above the velocity component in the direction of the coordinate a means derivation with respect to the proper time, the time of the moving test mass. Nordström made two proposals. The first was

$$\dot{u}_a = -\partial_a \phi - \dot{\phi} u_a \quad \text{and} \quad \Box \phi = -4\pi G \rho. \tag{6.9}$$

The D'Alembertian in Cartesian coordinates in (6.9) has the signature (+,-,-,-)

$$\Box = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2. \tag{10}$$

The second proposal was

$$\phi \dot{u}_a = -\partial_a \phi - \dot{\phi} u_a \quad \text{and} \quad \phi^{-1} \Box \phi = -4\pi G \rho \quad \text{or} \quad \phi \Box \phi = -4\pi G \rho \tag{6.11}$$

depending on the source. The Wikipedia [1] gives the second version. In 1915 Einstein and Fokker published a paper stating that Nordström's equation of motion in (6.11) comes from the egodesic Lagrangian of a curved Lorentzian manifold with $g_{ab} = \eta_{ab}\phi^2$. This is strange because it does not. The geodesic equation is

$$\ddot{x}^\mu + \Gamma^\mu_{ab} \dot{x}^a \dot{x}^b = 0. \tag{6.12}$$

Here Γ^μ_{ab} are the Christoffel symbols of the metric g_{ab}, x^a is a coordinate and $\dot{x}^a$ is the velocity in the inertial frame of reference of the moving test mass. The geodesic Lagrangian can be derived by minimizing the action, but Einstein claimed that the geodesic Lagrangian is

$$L = g_{ab} \dot{x}^a \dot{x}^b = 0$$

which in the chosen metric is

$$L = \eta_{ab}\phi^2\dot{x}^a\dot{x}^b = 0. \tag{6.13}$$

In fact, (6.12) does not come from (6.13). We can see this easily. The Euler+Lagrange equations that apply here are

$$\frac{d}{ds}\frac{\partial}{\partial \dot{x}^\mu}L(s, x(s), \dot{x}(s)) - \frac{\partial}{\partial x^\mu}L(s, x(s), \dot{x}(s)) = 0. \tag{6.14}$$

As

$$\frac{\partial}{\partial \dot{x}^\mu}L(s, x(s), \dot{x}(s)) = \eta_{\mu\mu}\phi^2 2\dot{x}^\mu$$

and

$$\frac{\partial}{\partial x^\mu}L(s, x(s), \dot{x}(s)) = \eta_{\alpha\beta}(2\phi\partial_\mu\phi)\dot{x}^\alpha\dot{x}^\beta$$

equation (6.14) takes the form

$$\phi^{-2}\frac{d}{ds}(\psi^2\dot{x}^\mu) + (-\eta^{\mu\mu}\eta_{\alpha\beta}\phi^{-1}\partial_\mu\phi)\dot{x}^\alpha\dot{x}^\beta = 0 \tag{6.15}$$

i.e.,

$$\ddot{x}^\mu + (-\eta^{\mu\mu}\eta_{\alpha\beta}\phi^{-1}\partial_\mu\phi)\dot{x}^\alpha\dot{x}^\beta = -2(\psi^{-1}\dot{\psi})\dot{x}^\mu.$$

Assuming that ϕ does not depend on the proper time s we almost have the geodesic equation (6.12), but not quite, and it is not only a matter of a sign: though the terms

$$\Gamma^\mu_{\mu\mu} = \eta^{\mu\mu}\eta_{\mu\mu}\phi^{-1}\partial_\mu\phi \quad \text{and} \quad \Gamma^\mu_{\alpha\alpha} = \eta^{\mu\mu}\eta_{\alpha\alpha}\phi^{-1}\partial_\mu\phi$$

are correct, there are nonzero terms in (6.12) that do not appear in (6.15)

$$\Gamma^\mu_{\alpha\mu} = \eta^{\mu\mu}\eta_{\alpha\alpha}\phi^{-1}\partial_\alpha\phi \neq \eta^{\mu\mu}\eta_{\alpha\alpha}\phi^{-1}\partial_\mu\phi \quad \text{if} \;\; \alpha \neq \mu.$$

Thus, the Lagrangian (6.13) is not quite correct and we cannot use it to derive a geodesic. Additionally, (6.15) is not the equation of motion in (6.11) or (6.9).

Einstein was correct in stating that the equation of motion Nordström was looking for is the equation of a geodesic of a Lorentzian manifold with the metric $g_{ab} = \eta_{ab}\phi^2$. As Nordström was trying to create a scalar field theory, the metric could not be anything else than $g_{ab} = \eta_{ab}\phi^2$. It is because a scalar field is a conformal mapping (outside isolated singularities) and at each point all sides dx_0, dx_1, dx_2, dx_3 of a small cube are multiplied by the same number $\phi(x)$. The correct equation of motion follows easily from the equivalence principle: all masses fall with the same speed (accelerate equally much, or a particle falling in a gravitational field does not have acceleration in the inertial frame of reference), as Galileo Galilei noticed long ago. A mathematical consequence of the equivalence principle is that test masses move along geodesics.

According to [1] Einstein said that Nordström's second field equation, i.e., the one in (6.11), comes from the Lagrangian

$$L = \frac{1}{8\pi}\eta^{ab}\phi_{,a}\phi_{,b} - \rho\phi. \tag{6.16}$$

(The notation $\phi_{,a}$ means $\partial\phi/\partial x_a$.)

It does not. The Euler-Lagrange equation in this case is

$$\partial_\mu\left(\frac{\partial}{\partial\psi_{,a}}L(\phi,\phi_{,a})\right) - \frac{\partial}{\partial\phi}L(\phi,\phi_{,a}) = 0$$

where

$$\frac{\partial}{\partial\psi_{,a}}L(\phi,\phi_{,a}) = \frac{1}{4\pi}\eta^{\mu\mu}\phi_{,\mu} \quad \text{and} \quad \frac{\partial}{\partial\phi}L(\phi,\phi_{,a}) = -\rho.$$

Thus, equation (6.17) is

$$\frac{1}{4\pi}\eta^{\mu\mu}\phi_{;\mu} + \rho = 0 \quad \text{i.e.,} \quad \Box\phi = -4\pi\rho. \tag{6.18}$$

(The notation $\phi_{;a}$ means the second partial derivative $\partial^2\phi/\partial x_a^2$.)

Equation (6.18) is the field equation of Nordström's first theory, not of the second theory in (6.11). Einstein was for a change correct when he told that the metric $g_{ab} = \eta_{ab}\phi^2$ yields the equation

$$R = -6\phi^{-3}\Box\phi \tag{6.19}$$

where R is the Ricci scalar curvature. Consequently, (6.19) can be obtained from the Lagrangian

$$L = \eta^{ab}\phi_{,a}\phi_{,b} - \frac{R}{12},\phi^4 \tag{6.20}$$

which is interesting since the corresponding scalar quantum theory is renormalizable. Einstein also gave the following expression to Nordström's second field equation

$$\phi\Box\phi = -4\pi G T_{matter} \tag{6.21}$$

where T_{matter} is related to the trace T of the energy-stress tensor. One notices that field equation may have on the left side $\phi\Box\phi$ or $\phi^{-1}\Box\phi$ depending on what is on the right side. It is simply to multiply the equation with some power of ϕ.

By (6.19)

$$\phi\Box\phi = -\frac{R}{6}\phi^4 = -4\pi G T_{matter}, \tag{6.22}$$

thus $R = 24\pi T$ and $T = G\phi^{-4}T_{matter}$ gives the desired form of the trace of the energy-stress tensor. These considerations were later important in the development of the General Relativity Theory. Einstein convinced Nordström to include the energy-stress tensor to the theory, which is quite good as it puts emphasis to the fact that there are four Ricci entities R_{aa}, which means that there are implicitly four equations in (6.19). It may look like there is only one equation in (6.19) and that there could be a linear combination of solutions for the component terms $g^{aa}R_{aa}$. Such linear combinations appear e.g. with degenerated energy levels, but here it is not such a case. There cannot be any linear combination. Equation (6.19) implies an equation for each a that has to be fulfilled. Three of the equations are have identical forms if (6.19) is expressed in Cartesian coordinates. One, the time, is a bit different, but also that equation is completely determined by (6.19). For the three space coordinates, the equation for coordinate b can be obtained from the equation for coordinate a by swapping a and b. This similarity of the form naturally does not imply that the functions for a and the functions for b have the same values. If the mass distribution ρ is not symmetric, the field ϕ has different values for space coordinates. Only the form is symmetric.

Nordström's gravitation theory in fact has ten equations in (6.19), just likethe General Relativity Theory has ten equations. However, in Nordström's gravitation theory all Ricci entries R_{ab}, $a \neq b$, are automatically zero. This follows from the metric for a scalar theory. There remains four equations, one equation for each diagonal Ricci entry. The diagonal entries R_{aa} are completely determined by the Ricci scalar R. If the Ricci scalar R is zero, usually (unless e.g. $\phi = 0$) the diagonal entries R_{aa} are not zero. In [2] I calculated the Ricci entries for Nordström's theory in Cartesian and spherical coordinates and one can verify that these entries are fully determined by R and nonzero if $R = 0$. There are four unknowns to be determined: one can e.g. determine $\partial_a \phi$ or ϕ_a. It means that there are four unknowns and four equations, very much as it should be. In Einstein's General Relativity there are ten equations for four unknowns.

6.1 Why the scalar theory is the correct one?

Already in 1915 it was known at least to Einstein, and after Einstein and Fokker had published their paper and Einstein had discussed with his close co-worker Nordström it must have been known also to Nordström that the correct equation od motion in the scalar theory of gravitation is the geodesic (6.12). It is therefore curious that some modern day evaluations of the scalar theory of gravitation (Nordström's theory of gravitation) take one of the original proposals Nordström initially made for the equation of motion, that is, (6.9) or (6.10). Even without any evaluations these equations of motion are known to be false, and they were known to be false in 1915. Yet, they do not in any way imply that the scalar theory of gravity with the equation of motion in (6.12) is false. Following this tendentious evaluation of Nordström's theory, some evaluations start from the geodesic Lagrangian (6.13) that Einstein's claimed gives Nordström's theory. This Lagrangian is incorrect. It does not give a geodesic and it does not give either of Nordström's theories. It seems that evaluators of alternatives for General Relativity may not be the most honest researchers in the field.

We can see how the Wikipedia pages for Nordström's theory [1] evaluate the scalar theory of gravitation. The calculations make use of the general solution of ϕ in vacuum. This general solution is derived from the field equation in vacuum (that is, $\rho = 0$)

$$\Box\phi = 0. \tag{6.23}$$

In theory one can proceed in this way, but I doubt that the general solution in [1] is correct. We can theoretically solve (6.23) by using the equivalence principle. Or at least Einstein said that the equation of motion (a geodesic) can be solved from the field equation for the vacuum. Einstein never explained how he thought this would be derived and some philosopher-physicists have doubted whether it is so. However, Einstein's thought is very simple to reconstruct, and it is correct. First notice that

$$\phi(\bar{r}) = -\frac{1}{||\bar{r} - \bar{a}||} \tag{6.24}$$

($\bar{a}$ is a constant vector) is a solution to the time-independent equation

$$\Delta\phi = 0. \tag{6.25}$$

Let $\bar{a}$ be a function of time. In (6.24) the vector $\bar{a}$ can be freely selected in the 3-space. Each selected $\bar{a}(t)$ defines a path in the 3-space starting from the point $\bar{a}(0)$. Let this path be a geodesic defined by the punctuated vacuum (with a masspoint in the origin). Then $\bar{a}(t)$ is a path of a test mass falling in the gravitational field in this punctuated vacuum. In local coordinates the test mass feels no gravitation field, thus $\Box\phi = 0$ in the moving coordinates. Making a coordinate transfer from moving coordinates to laboratory coordinates we get the solution in laboratory coordinates. When this time dependent $\bar{a}(t)$ is inserted to (6.25) we have a time dependent field. Does this field fulfill the field equation for the vacuum as in (6.24), or for some nonzero mass distribution? It must give the solution to the field equation in the vacuum because $\bar{a}(t)$ describes how a test mass moves in a (punctuated) vacuum field. Thus, we get the time dependent solution to the field equation in the vacuum from the geodesic equation in the vacuum field. Yet, it is very difficult to solve the field equation for the vacuum directly. The equation for $\bar{a}(t)$ becomes nonlinear and complicated. I am not convinced of the evaluation of the scalar theory in [1] because of the way it treats the general solution of (6.23).

The scalar theory passes the redshift and Pound-Rebka experimental tests because it fulfills the equivalence principle by construction. It must be noted that the equation of motion in such a calculation must be (6.12), not (6.9) or (6.11),and that the geodesic Lagrangian is not (6.13). It seems that in [1] the geodesic is derived from (6.13) in the calculation of the Shapiro delay and the precession of periastria. If so, then the calculations do not evaluate the past 1915 version of the scalar theory.

Earlier I have seen the following argument against the scalar field theory: the claim that light does not bend in Nordström's theory. This may be so if the equations of motion are taken from (6.9) or (6.11), but (6.12) is the geodesic equation. All test masses move along a geodesic. Nordström's theory does not consider electromagnetism, but a photon can be treated as a limit case of a test mass when the rest mass goes to zero. A photon has moving mass and therefore passes as a test mass and it moves aong a geodesic. Light does beng in a gravitational field also in the scalar theory.

Lastly, let me give a simple example where the General Relativity Theory fails and Nordström's theory gives a correct result. Let us ask if the gravitational field in a satellite on the earth's orbit is well approximated by the gravitational potential (6.3). I assume that it is very well approxilated by the gravitational potential (6.3) because that is the form engineers use for calculating the potential and if there would be an experiment on a satellite orbit where the General Relativity gives different and more accurate predictions that the gravitational potential of (6.3), then this experiment would be mentioned in all textbooks. Thus, (6.3) is very good as an approximation. It corresponds to a scalar potential field and the metric is of the form $g_{ab} = \eta_{ab}\phi^2$ exactly as Einstein calculated. The Ricci entries $R_{ab} = 0$ when

$a \neq b$ and

$$R_{00} = \frac{1}{r^2} \quad R_{jj} = -\frac{1}{r^2} + 8\frac{1}{r^2}\left(\frac{\partial r}{\partial x_j}\right)^2 - 2\frac{1}{r}\frac{\partial^2 r}{\partial x_j^2} \quad \text{for } j = 1, 2, 3. \tag{6.26}$$

The Ricci curvature is zero:

$$R = g^{ab}R_{ab} = \eta^{00}\phi^{-2}\left(R_{00} - \sum_{j=1}^{3} R_{jj}\right) = G^{-2}\rho^{-2}r^2\left(R_{00} - \sum_{j=1}^{3} R_{jj}\right)$$

$$= G^{-2}\rho^{-2}r^2\left(\frac{1}{r^2} + \frac{3}{r^2} - 8\frac{1}{r^2}\sum_{j=1}^{3}\left(\frac{\partial r}{\partial x_j}\right)^2 + 2\frac{1}{r}\sum_{j=1}^{3}\frac{\partial^2 r}{\partial x_j^2}\right)$$

$$= G^{-2}\rho^{-2}r^2\left(\frac{1}{r^2} + \frac{3}{r^2} - 8\frac{1}{r^2} + 2\frac{1}{r}\frac{2}{r}\right) = G^{-2}\rho^{-2}(1+3-8+4) = 0. \tag{6.27}$$

However, the Ricci entries are not zero. It should be possible to measure the gravitation potential in the space on a satellite orbit and check if (6.3) is a good approximation to the potential. I assume this has been done and (6.3), the expression used by engineers and physicists in many practical applications is indeed quite good. It follows that the Ricci entries R_{aa} are not zero but R is zero. Nordström's gravitation theory agrees with this. Einstein's gravitation theory claims that in a vacuum, like our space orbit, every R_{ab} must be zero. Then the solution is the Schwarzschild solution. When Einstein's equations

$$R_{ab} - \frac{1}{2}Rg_{ab} = k_0 g^{ab}T_{ab} + \lambda g^{ab}g_{ab} \tag{6.28}$$

are summed over the diagonal entries, the equation they give for the metric $g_{ab} = \eta_{ab}\phi^2$ is exactly

$$\phi^{-3}\Box\phi = -\frac{1}{6}R = \frac{1}{3}k_0 T + \frac{\lambda}{3} \tag{6.29}$$

That is, the identity in the left side is just a mathematical consequence of the metric $g_{ab} = \eta_{ab}\phi^2$ and the right side equation comes from (6.28). The off diagonal Ricci elements R_{ab}, $a \neq b$, are zero for this metric. As we agree that on the space orbit the potential is close approximated by (6.3), the metric is closely approximated by the scalar metric and R_{ab} are closly approximate by what the scalar metric gives. The only conclusion is that Einstein is wrong in his claim that in the vacuum outside a masspoint every Ricci entry should be zero, as it is in the Schwarzschild solution.

The equation (6.29) is just a form of (6.21)

$$\phi\Box\phi = -4\pi G T_{matter} \quad \text{i.e.,} \quad \phi^{-3}\Box\phi = -4\pi T \quad \text{as } T = G\phi^{-4}T_{matter}$$

Both theories have ten equations, which for this metric reduce to four equations. In both theories test masses move along geodesics. If the gravitational potential is as in (6.3) in the part of the universe we can measure, there cannot possible be any difference between the Ricci entries in one theory or the other. The only difference between these theories if the potential is (6.3) is that Einstein does not admit that the Ricci entries R_{aa} are not zero. But they are not zero. For the values of Ricci entries, see my calculations in [2]. Einstein believed that this is not so. All Ricci entries should vanish in a satellite orbit and there is no scalar gravitational potential (6.3).

Instead, the geometry is deformed, small balls are not round, time goes with a different speed to different directions, and the solution to gravity in a satellite orbit is the Schwarzschild solution.

I made a simple (simplistic) evaluation in [2] (following the calculations that supposedly show that the scalar theory contradicts measurements) and did not find tests where the scalar theory fails. Instead, the Schwarzschild solution for General Relativity in vacuum failed the Shapiro delay test. This is so because the ball gets increasingly deformed in this solution and the speed of light is not constant. A small ball (or a cube) must keep its form if the speed of light is to be constant. This is so because the speed of light is the relation of dx^i and dx^0 for $i > 0$. Ricci elements for the Schwarzschild solution are given in [2] and one can compare (6.26)-(6.27) to the way Ricci elements get zero values in the Schwarzschild solution.

On the Earth Newtonian physics, with small corrections such as the redshift, works well. The space is largely empty and is zero in the space, or close to it. It is possible that somewhere in the large universe there is a black hole where small balls of the geometry are not round and the solution is the one calculated by Schwarzschild. If so, then there is some place in the universe where the scalar theory does not hold and we must use Einstein's equations and allow for a more general metric. But this cannot be the case in the experiments of redshift, Shapiro delay and the movement of planets in our solar system.

I do not think there are tests that can be made and that can demonstrate that the metric in the space within a measurable distance from the earth is something else than $g_{ab} = \eta_{ab}\phi^2$ and that the potential (6.3) is a poor approximation in some point in the space. Possibly inside a black hole there is such a place where the metric is different, but it is not easy to make direct measurements inside black holes.

Evaluations that have been made in order to show the scalar theory incorrect are based on the use of the wrong equations of motion in (6.9) and (6.11), the wrong Lagrangian in (6.13), or a wrong general solution to the field equations.

There are other reasons why the General Relativity is not correct. The General Relativity Theory cannot be formulated as a renormalizable quantum field theory, while the scalar theory is renormalizable (as mentioned e.g. in [3]). It seems that all quantum field theories should be gauge field theories. Dietmar Ebert in [13] p. 53-57 there is one proposal how to turn GR into a gauge field theory, but the result cannot be renormalized. There is a considerable similarity between the concepts of GR and gauge field theories. For a scalar field these similarities may be realized into new insight to the problem.

The book [5] by Julius Wess and Jonathan Bagger starts ([5] p.4) by referring to the Coleman-Mandula theorem, which gives rather restrictive conditions to symmetries of the S-matrix. The authors release one condition and get the most general supersymmetric algebra on p. 8. They manage to formulate supergravity models, but in Kähler, i.e., complex, manifolds. The real dimension of a 4-dimensional Kähler manifold is eight and for that reason supergravity is not a proper quantization of Einstein's General Relativity. Their supergravity model includes scalar (chiral) fields, Kähler potentials.

Quantum gauge field theory, even without supersymmetry, is closer to the scalar theory of gravitation than to the General Relativity. In gauge field theory masses are created by spontaneous symmetry breaking induced by a scalar field. As the theory of gravitation is closely associated with masses, we should expect to find the mass creating scalar field having something to do with gravitation.

Measurements do not so far support the assumption that our universe has supersymmetry. A similar negative conclusion seems to be true with the superstrings theory. I have personally not studied superstrings (apart from briefly scanning a popular science book [5]) and cannot make my own judgment, but research in superstrings is already some 30 years old and has not led to a solution to quantum gravity. As for the scalar gravitation theory, it is well known (mentioned e.g. in [6]) that the theory is renormalizable and satisfies the strong equivalence principle. I found a further argument against the General Relativity in a most unexpected source: Martin Gardner [7] Chap. 17 tells how a number of physicists headed by J. A. Wheeler wrote a letter of ESP mentioning that there are only three or four experiments that give support to General Relativity.

We can add the recent finding of graviton to this short list of experiments, but it is not a decisive proof of GR. What actually was found (unless the results are noise, as has been suggested) is an elementary particle with zero mass, zero charge, spin 2 and weak coupling with hadrons. In Einstein's theory such a particle is a graviton. However, it may be a gluon. A gluon coming from a distant star can have very high momentum and for that reason have a small coupling constant with fermions. The property of the color force, called asymptotic freedom - color confinement (see [8] p. 147) does not actually say that gluons are confined to a small distance. It is a relation with renormalization mass and the coupling constant and it says that gluons with high momentum interact weakly with fermions.

6.3. Cosmological implications

The main implication of the scalar theory of gravitation to cosmology is that the field equation is relatively simple, especially simple it is for vacuum, but small amount of mass can be treated as a perturbation. It should be possible to derive results much easier in the scalar theory than in General Relativity.

One slightly negative implication of accepting Nordstrm's scalar theory is that the Schwarzschild solution is not the correct one as a vacuum solution in our space-time. The Schwarzschild solution has been widely studied in the context of black holes, see [10]. If the geometry of our part of the universe is $g_{ab} = \eta_{ab}\phi^2$ it does not necessarily exclude the possibility that there may be somewhere black holes behaving as the Schwarzschild solution. Even if they do not exist, the study of such constructions remains interesting in the mathematical sense. The authors of [10] were not interested in massless scalar fields in gravitation. Indeed, p. 324 in [10] (equation 8.61) states that a massless scalar field ϕ obeys $\Box\phi = 0$ and that there are no massless scalar fields in the nature. Neither claim is correct: a massless scalar field can have a $\lambda\phi^3$ term in the field equation (in the Lagrangian the term is $\lambda\phi^4$), and such massless scalar fields appear in spontaneous symmetry breaking as Goldstone boson fields.

Another implication, coming directly from the conformal nature of scalar fields, is that the problem of inflation in cosmology can be easily explained. In the early universe all matter was in a small space and the gravitational field potential ϕ was very high. Consequently, the space element was very large (each dx^i gets multiplies by $\phi(x)$ in the conformal mapping). It follows that even if we let the speed of light be the same constant, as the space element is very large, the expansion of the universe in the earliest picoseconds seems for us much faster than light. This is because the unit element for the time was also very large. We do not need to change the speed of light, as is done in [11]. The conformal mapping also solves the eternal question: what was before the universe was born? If at the beginning all matter was concentrated in a very small space, even to a singularity, the time unit element was huge, even infinite. There was no time before the universe started expanding. In the original singularity time did not tick.

Some very simple conclusions can be made based on ρ and ρ_1 in (6.6). The scalar theory gives the same solution as the Newtonian theory if ϕ is time-independent. We can conclude that ρ_1 cannot decrease because the gravitational potential of an empty space with mass in the center is (6.3). If there is mass in the space, then the potential falls slower than (6.3). Let us have ρ_1 behave as r^α. Then the mass distribution ρ behaves as $r^{\alpha-2}$ by (6.7). Integrating ρ over a ball of radius r gives

$$\int dr 4\pi r^2 r^{\alpha-2} \sim r^{\alpha?1}$$

assuming that the volume grows as in the flat space. Consequently, the mass goes to infinity if r can grow to infinity. The simplest explanation for this behavior is that r cannot grow to infinity because all mass is in a limited area: the universe has a finite size.

A manifold can have a finite size without having a border, in which case it is closed. In a stationary case the manifold can be taken as a closed, orientable 3-manifold. If the metric is given by a scalar potential, the mapping is conformal outside isolated singular points. We can generalize the mapping to the extent that the unit ball can deform to a limited extent (the speed of light can vary within some limits, just to mention, the Schwarzschild solution does not fill this condition, it is not quasiregular). In that case the mapping is quasiregular. I classified in 1988 closed, orientable 3-manifolds that admit a quasiregular map [12]. There is a number of possibilities. These spaces would then have (positive) Ricci curvature. However, the most natural choice is that our universe does have a border: if the space-time is almost flat, the universe is a ball with border. Topological oddities may only appear in areas of high gravitation potential.

6.4. Quantum gravity

After these easy preliminary comments we can move to quantization of gravity. One of the physicists signing the letter by Wheeler about ESP mentioned by Martin Gardner in [7] was Richard Mattuck. His excellent book [8] can serve as an introduction to the quantum theoretical setting, but it does not treat gauge fields. I am somewhat familiar with only two books that treat quantum gauge field theories, Wess-Bagger [4] and Bailin-Love [9], and neither of them is ideal as a study text. Wess-Bagger is difficult to apply in the standard model as the setting is supersymmetry while Bailin-Love is difficult to read due to errors, starting from the first page. (On p.1 there is written $\det A = Tr \ln A$. The symmetric matrix A must be diagonalized: $\ln \det A = Tr \ln U^T AU$ and in order to get the integral converge all eigenvalues must be nonnegative. The book has many other errors, especially in the beginning.) Nevertheless, these books do contain all necessary background for making certain observations of scalar fields. In quantum field theories the Lagrangian (density) (6.20) would be written (see [9]) as

$$L = (\partial_\nu \phi)(\partial^\nu \phi) - \frac{R}{12}\phi^4 \tag{6.30}$$

and it is of the form of a Lagrangian of a renormalizable quantum scalar field

$$\mathcal{L} = (\partial_\nu \phi)(\partial^\nu \phi) - \mu^2\phi^2 + \lambda\phi^4 \tag{6.31}$$

The field (6.31) has found an application as the Higgs field that is creating masses for elementary particles by spontaneous symmetry breaking in the electroweak gauge field theory.

The Higgs mechanism and spontaneous symmetry breaking with a special gauge for hiding the Goldstone boson field are all explained in [9] Chapter 13, so I will write ery few formulae in this text, and will make no new mathematics. I will only illustrate a possible approach.

There always appears at least one massless (complex) scalar field when gauge invariance is imposed to a Lagrangian density. Especially such a scalar field, a Goldstone boson field, appeared in the electroweak theory when the

Higgs mechanism was used to break the chiral (actually parity according to [9]) symmetry. Massless scalar field was seen as undesirable in the theory. In a special gauge, the unitary gauge, the Goldstone scalar field is eliminated from the equations. In this gauge a massless scalar field does not appear in renormalization calculations and evaluation of Feynman propagators.

But the Goldstone boson field has not disappeared because of this mathematical trick, and it is not an artifact like Faddeev-Popov ghosts. The elimination of the Goldstone boson is only possible in the range where the massive Higgs scalar field is present. As mentioned in [13], the range of massive fields should be (rouhly) inversely proportional to the mass. Therefore the Higgs massive scalar field cannot cancel the Goldstone massless scalar field in large distances. The massless scalar field must be visible in large distances. There are only two massless long range fields: the electromagnetic vector potential A_μ and the gravitation field. It is logical to expect that the massless Goldstone scalar field is the scalar gravitation field in Nordstrm's theory. As there are only two long range massless boson fields, we cannot expect more than one Higgs field. It follows that the approach of SU(N) GUT is hopeless: those theories have more gauge symmetries that need to be broken, and consequently they have more Goldstone boson fields.

There are some issues in interpreting the Goldstone boson field as the scalar field of gravitation. Firstly, we have to make the scalar field complex. This does not seem to be a problem. The phase ϕ is cancelled in Cristoffel symbols, and in the metric we can write $\phi^*\phi$ instead of ϕ^2. The Lagrangian density in (6.30) or (6.31) is easily written with a complex field.

The mass term $\mu^2\phi^2$ is missing from (6.30). The Higgs field has mass, but there is a Goldstone boson field that is massless (and so must have $\mu^2\phi^2 = 0$). Thus, (6.30) can describe a Goldstone boson field. But there seems to be one problem. In [9] λ in (6.31) is positive, while R in (6.30) is also positive. The fact that R is positive is clear from (22). From (6.29) we see that R is of the order of $G^{-2}\rho^{-2}$. It seems to me that [9] has an error in this place. The choice of the coordinates in [9] is such that $\square = \partial_0^2 - \Delta$. Therefore the field ϕ is negative and it does not have a local minimum at the minimum energy place. It has a local maximum and λ must be negative. This also explains why the mass term $-\mu^2\phi^2$ in [9] is negative. The number $-\mu^2$ should be positive as λ is negative and the Higgs field does not have an imaginary mass.

The issue with mass generation is briefly explained as follows. Consider the Lagrangian density of the Dirac field (describing a relativistic electron):

$$L = \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi \tag{6.32}$$

where $\bar{\psi} = (\psi^*)^T$ is the conjugate field. This Lagrangian is gauge invariant under the global gauge transformation

$$\psi \to e^{-q\lambda}\psi \quad , \quad \bar{\psi} \to e^{q\lambda}\bar{\psi} \tag{6.33}$$

where λ is a real number. This fermion field is described by a spinor field and it has mass m. It is clearly an electron. The Lagrangian (6.32) is not invariant under a local gauge transformation where $\lambda = \lambda(x)$ depends on the coordinates. Local gauge invariance can be achieved by changing the Lagrangian to

$$L = \bar{\psi}\left(i\gamma^{\mu}D_{\mu} - m\right)\psi \tag{6.34}$$

where $D_{\mu} = \partial_{\mu} + iqA_{\mu}$ and

$$A_{mu} \to A_{\mu} + \partial_{\mu}\lambda. \tag{6.35}$$

The vector field A_{μ} is the vector potential of electrodynamics and it follows the Maxwell equations, which come as E-L equations from the Lagrangian

$$F^{\mu\nu}F_{\mu\nu} \quad \text{where} \quad F_{\mu\nu} = \partial_{mu}A - \partial_{nu}A. \tag{6.36}$$

It is easy to check that the Lagrangian term (6.36) is invariant in the local gauge transformation (6.35). Electrons have mass and the Dirac field has a mass term $m\bar{\psi}\psi$ for the electron. This mass term is invariant in the transformation (6.33)

$$\bar{\psi}\psi \to e^{iq\lambda}\bar{\psi}e^{-iq\lambda}\psi = \bar{\psi}\psi. \tag{6.37}$$

If the vector field A_μ had mass it should have a corresponding mass term of the type $mA^*_\mu A^\mu$ (here the conjugate is simply a complex conjugate, not a Hermitian conjugate of matrices). However, the mass term for the vector field is not invariant:

$$A^*_\mu A^\mu \to \left(A^*_\mu + \partial_\mu \lambda\right)\left(A^\mu + \partial^\mu \lambda\right) \neq A^*_\mu A^{mu}. \tag{38}$$

As a result of this, the electromagnetic field and its field boson, photon, is massless. In a similar way the theory of strong interactions, QCD, is a gauge field theory with a local gauge symmetry. The field boson, gluon, is massless. There is a difference with the fermions of that theory. The fermions are quarks, which do have mass. The Lagrangian for QCD with the quark mass term can be found e.g. in [13] p. 144. However, Bailin and Love inform that there is a complication. The gauge group of QCD is non-Abelian and the mass terms of fermions in QCD transform as

$$\bar{\psi}\psi \to e^{iq\lambda_a T_a}\bar{\psi}e^{-iq\lambda_a T_a}\psi. \tag{6.39}$$

The difference to (6.37) is that instead of a real number λ there is a more complicated exponent. The terms in (6.29) do not commute and we cannot move the exponents next to each other and cancel them. The result is that the fermion mass terms break the gauge symmetry. It follows that mass terms cannot appear in the Lagrangian: for the purposes of the color force of strong interaction, quarks are massless (even though the mass terms do appear in the Lagrangian and are positive masses in the flavor symmetry and in electroweak interactions). This in any case is what [9] says.

We see that QED, the theory of electromagnetism, has no problem with fermion masses because the gauge group U(1) is Abelian. The theory of strong interactions, QCD, would have a problem with fermion masses as the gauge group is non-Abelian, but the color force can ignore the masses. In both of these theories boson fields (the interaction particle fields) are massless.

There remains the electroweak theorym which solved the problem of chirality violation in weak interactions by introducing the Higgs mechanism of spontaneous symmetry breaking. (The mechanism is well explained in [4],[9] and [13] and there is no need to describe it here.) This mechanism creates masses to the interaction particles. The mechanism also creates masses to fermions by spontaneus symmetry breaking. What happens with bosons in the electroweak theory is that the original massless field gets mixed with the massive Higgs scalar field. The original massless field has only two transversal modes. The mixed field has three modes: two transversal and one longitudinal. In a sense the mixed field has three dimensions and the original massless field has two dimensions. We get to a philosophical conjecture: maybe a massive field must have volume and a field that does not bound volume is massless. It would support the basic view of Einstein that mass is in some sense geometry.

References for Chapter 6:

[1] The Wikipedia page on Nordstrm's gravitation theory (as 23.2.2020)

[2] J. Jormakka, "Einstein's Field Theory is Wrong and Nordstrm's Correct" available in the vixra archive as number 1812.0067 Included in this book as chapter 5.

[3] Nathalie Deruelle, "Nordstrm's scalar gravitation theory and the equivalence principle," arXiv: 1104.4608, 2011.

[4] Julius Wess, Jonathan Bagger, *Supersymmetry and Super-gravity*, Princeton Univ. Press, New Jersey, 1992.

[5] Brian Greene, *The Elegant Universe, Superstrings, Hidden Dimensions and the Quest for the Ultimate Theory*, W. W. Norton, 1999.

[7] Martin Gardner, *Science good, bad and bogus*, Polish ed. Pandora, 1994 (original 1981).

[8] Richard D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, 2. ed. Dover, New York, 1992 (original 1967).

[9] David Bailin, Alexander Love, *Introduction to Gauge Field Theory*, IOP Publishing, 1986.

[10] Kip. S. Thorne, Richard H. Price, Douglas A. Macdonald, *Black Holes the Membrane Paradigm*, Yale University, 1986.

[11] João Magueijo, *Faster than the speed of light*, Penguin books, 2003.

[12] Jorma Jormakka, *The existence of quasiregular mappings from to closed orientable 3-manifolds*, Ann. Acad. Sci. Fenn. Series A, 69, Finnish Academy of Science, 1988, p. 1-44.

[13] Dietmar Ebert, *Eichentheorien, Grundlage der Elementarteilchenphysics*, Akademie-Verlag, Berlin, 1989.

7. Solutions to Yang-Mills equations

In the year 2000 the Clay Mathematics Institute (CMI) posed the following problem [1]:

Yang-Mills Existence and Mass Gap. Prove that for any compact simple group G, a nontrivial Yang-Mills theory exists on $\mathbb{R}^4$ *and has a mass gap* $\Delta > 0$*. Existence includes establishing axiomatic properties at least as strong as those cited in R. Streamer and A. Wightman (1964) or K. Osterwalden and R. Seiler (1973).*

Thus, the existence of a non-trivial Yang-Mills theory involves showing that the theory fills axioms of axiomatic quantum field theory, while the existence of a mass gap seems to be another question that may be shown also in some other way. The problem concerns a pure Yang-Mills Lagrangian, i.e., only the gauge field without spinor fields, Higgs fields, or other fields. The mass gap is expected to arise from self-intersections of the Yang-Mills gauge field. The issue how the mass gap could appear is unclear but [2] has proposed one mechanism. The state of research to this problem up to 2004 is summarized in [3]. After that there have been some efforts to prove the existence of a mass gap, e.g. [4], [5], but the problem is still considered open.

This article presents explicit solutions to the Yang-Mills Euler-Lagrange equations. The solutions give arbitrarily small positive values for energy. This shows that the Hamiltonian has arbitrarily small eigenvalues indicating that there is no mass gap. The solutions can be given on $\mathbb{R}^4$ with Minkowski's or Euclidean metric and they are simple, natural solutions that should be accepted as gauge fields in any non-trivial quantum field theory for the pure Yang-Mills Lagrangian.

The reason why there is no mass gap in these solutions is not any mystery. They are solutions to classical Yang-Mills fields and as there are no spinor fields, there is no reason why they should have a mass gap. However, they are also the largest set of solutions to the pure Yang-Mills equations, and there also is no good reason to exclude them from a nontrivial quantum Yang-Mills theory. If we exclude the largest set of solutions simply because they do not have a mass gap in order to show that there is a mass gap, is it not a trivial way to show such a thing? The quantization we do is admittedly naive but the CMI problem is posed to mathematicians, and the quantization that is done follows accepted guidelines. Quantization should be in some way compatible with path integral quantization, and that is the starting point. For making the claim that there is no natural way to excude the classical solutions it is not necessary to give an acceptable way to quantizise the field, only to show that in order to get a mass gap one should do something what the CMI problem does not require, such as modifications to the Yang-Mills equations, e.g., changing the coupling constant to a dimensioned parameter, or other things. Whether this answer to the millennium problem satisfies CMI is to be seen. Probably it does not, but at least the problem should be better formulated, i.e., stating that it is allowed to make changes in the Yang-Mills fields. Naturally, if one allows any changes to be made, then we lose the problem. Thus, it may be that no clearly stated millennium prize problem can be formulated from the quantum Yang-Mills fields.

Definitions and notations

We will first describe the problem setting as it can be presented in physics in tensor calculus, and at the end look at the more mathematical formulation with differential forms and the Hodge star operator. Unless otherwise stated, or the sum is written explicitely, there is summation over indices that are repeated on one side of an equation. For notations we refer to [6].

$$\mathcal{L} = -\frac{1}{2}Tr(F_{\mu\nu}F^{\mu\nu}) = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \tag{7.1}$$

where

$$F^{\mu\nu} = F_a^{\mu\nu}t_a \tag{7.2}$$

and t_a are the generators of the Lie group satisfying

$$Tr(t_a t_b) = \frac{1}{2}\delta_{ab} \quad [t_b, t_c] = if_{abc}t_a \tag{7.3}$$

The structure constants $f_{abc} = f^{abc}$ are selected antisymmetric in all indices. The gauge field

$$A^\mu = A^\mu_a t_a \tag{7.4}$$

defines the curvature $F^{\mu\nu}$ by

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu + ig[A^\mu, A^\nu] \tag{7.5}$$

In component form this gives

$$F^{\mu\nu}_a = \partial^\mu A^\nu_a - \partial^\nu A^\mu_a - g f_{abc} A^\mu_b A^\nu_c \tag{7.6}$$

Curvature is antisymmetric

$$F^{\mu\nu} = -F^{\nu\mu} \tag{7.7}$$

The number g is called coupling constant, and

$$\partial^\mu = \frac{\partial}{\partial x_\mu} \quad \partial_\mu = \frac{\partial}{\partial x^\mu} \tag{7.8}$$

are partial derivatives with respect to the contravariant coordinates x^μ and contravariant coordinates $x_\mu = g_{\mu\nu} x^\nu$. $x^0 = ct$ and x^j, $1 \leq j \leq 3$, are the space coordinates. The metric $g_{\mu\nu} = g^{\mu\nu}$ is Minkowski's metric

$$(g_{\mu\nu})_{\mu,\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{7.9}$$

Thus $x_0 = x^0$, $x_j = -x^j$ for $1 \leq j \leq 3$. For real vectors and tensors lowering and raising indices is made by

$$A^a_\mu = g_{\mu\nu} A^\nu_a \quad F^a_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} F^{\alpha\beta}_a \quad \partial_\mu = g_{\mu\nu} \partial^\nu \tag{7.10}$$

Therefore (7.6) can also be expressed as

$$g_{\mu\alpha} g_{\nu\beta} F^{\alpha\beta}_a = g_{\mu\alpha} g_{\nu\beta} \left(\partial^\alpha A^\beta_a - \partial^\alpha A^\beta_a - g f_{abc} A^\alpha_b A^\beta_c \right)$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{abc} A^b_\mu A^c_\nu \tag{7.11}$$

where we have written f^{abc} instead of f_{abc} to follow the summation convention for the index a. The natural setting of quantum field theories is that the component functions of the fields take complex values. Then raisingand lowering indices involves taking complex conjugates but we will only do calculations with real fields. Complex fields are better threated by the algebraic geometric formulation described briefly at the end of Section 7.2.

Let us notice that there is a summation over b and c in (7.6) and (7.11). We formulate this simple observation as a lemma since it is needed in the sequence.

Lemma 7.1 Let $c > b$. The last term in (7.2) can be expressed as

$$f_{abc} A^\mu_b A^\nu_c = \sum_{c>b} f_{abc} \left(A^\mu_b A^\nu_c - A^\mu_c A^\nu_b \right) \tag{7.12}$$

Proof. Expanding the commutator $[A^\mu, A^\nu]$

$$\left(\sum_b A^\mu_b t_b \right) \left(\sum_c A^\nu_c t_c \right) - \left(\sum_c A^\nu_c t_c \right) \left(\sum_b A^\mu_b t_b \right)$$

$$= \sum_{b,c} \left(A^\mu_b A^\nu_c t_b t_c - A^\nu_c A^\mu_b t_c t_b \right) \tag{7.13}$$

Since A^μ_a are scalars $A^\nu_c A^\mu_b = A^\mu_b A^\nu_c$. Thus we get

$$[A^\mu, A^\nu] = \sum_{b,c} A^\mu_b A^\nu_c [t_b, t_c] = \sum_{b,c} A^\mu_b A^\nu_c i f_{abc} t_a$$

$$= \sum_{a,b,c} i \left(f_{abc} A^\mu_b A^\nu_c + f_{acb} A^\mu_c A^\nu_b \right) t_a \tag{7.14}$$

$$= \sum_a \sum_{c>b} i f_{abc} \left(A^\mu_b A^\nu_c - A^\mu_c A^\nu_b \right) t_a$$

EOP

As an example, let the group be SU(2). It has three generators

$$t_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad t_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad t_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Then

$$A^\mu = \sum_{a=1}^{3} A_a^\mu t_a$$

$$[A^2, A^3] = i\left(f_{123} A_2^2 A_3^3 + f_{132} A_2^2 A_3^3\right) t_1$$

$$+i\left(f_{231} A_3^2 A_1^3 + f_{213} A_3^2 A_1^3\right) t_2 + i\left(f_{312} A_1^2 A_2^3 + f_{321} A_2^2 A_1^3\right) t_3$$

showing that Lemma 7.1 holds in this example. The proposed solutions make use of the following lemma.

Lemma 7.2 Let the gauge field have the form

$$A_a^\mu = s_a E^\mu \tag{7.15}$$

Then $F_a^{\mu\nu}$ has the form

$$F_a^{\mu\nu} = s_a G^{\mu\nu} \tag{7.16}$$

and (7.6) and (7.11) reduce to

$$F_a^{\mu\nu} = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu$$

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a \tag{7.17}$$

Proof. Because of (7.11) it suffices to show (7.16) and the first equation in (7.17). From Lemma 7.1

$$f_{abc} A_b^\mu A_c^\nu = \sum_{c>b} f_{abc}\left(s_b E^\mu s_c E^\nu - s_c E^\mu s_b E^\nu\right) = 0 \tag{7.18}$$

since s_a and E^μ are scalars and commutate. EOP

The Euler-Lagrange equations for $\mathcal{L} = \mathcal{L}(A^\mu, \partial^\nu A^\mu)$ are

$$\partial^\nu\left(\frac{\mathcal{L}}{\partial\left(\partial^\nu A_a^\mu\right)}\right) = \frac{\partial \mathcal{L}}{\partial A_a^\mu} \tag{7.19}$$

Lemma 7.3 Let
$$\mathcal{L} = -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a \tag{7.20}$$
and A_a^μ be real functions. Then

$$\frac{\partial F_d^{\mu\nu}}{\partial A_a^\mu} = -g f_{dac} A_c^\nu$$

$$\frac{\partial F_{\mu\nu}^d}{\partial A_a^\mu} = -g f_{dac} A_c^\nu g_{\mu\mu} g_{\nu\nu}$$

$$\frac{\partial \mathcal{L}}{\partial A_a^\mu} = \frac{1}{2} g f_{abc} A_b^\nu F_{\mu\nu}^c \tag{7.21}$$

$$\frac{\partial F_d^{\mu\nu}}{\partial \left(\partial^\nu A_a^\mu\right)} = -\delta_{ad}$$

$$\frac{\partial F_{\mu\nu}^d}{\partial \left(\partial^\nu A_a^\mu\right)} = \delta_{ad}(g_{\mu\nu} g_{\nu\mu} - g_{\mu\mu} g_{\nu\nu})$$

$$\frac{\partial \mathcal{L}}{\partial \left(\partial^\nu A_a^\mu\right)} = \frac{1}{2} F_{\mu\nu}^a$$

and the Euler-Lagrange equations are

$$\partial^\mu F_{\mu\nu}^a - g f_{abc} A_b^\mu F_{\mu\nu}^c = 0 \tag{7.22}$$

Proof. Directly computing

$$\frac{\partial F_d^{\mu\nu}}{\partial A_a^\mu} = \frac{\partial}{\partial A_a^\mu} \left(-g f_{dbc} A_b^\mu A_c^\nu\right) = -g \delta_{ab} f_{dbc} A_c^\nu = -g f_{dac} A_c^\nu \tag{7.23}$$

$$\begin{aligned} \frac{\partial F_{\mu\nu}^d}{\partial A_a^\mu} &= \frac{\partial}{\partial A_a^\mu} \left(-g f_{dbc} A_\mu^b A_\nu^c\right) = \frac{\partial}{\partial A_a^\mu} \left(-g f_{dbc} g_{\mu\alpha} A_b^\alpha A_\nu^c\right) \\ &= -g f_{dac} g_{\mu\alpha} \delta_{\alpha\mu} \delta_{ab} A_\nu^c = -g f_{dac} A_\nu^c g_{\mu\mu} \\ &= -g f_{dac} g_{\nu\alpha} A_c^\alpha g_{\mu\mu} = -g f_{dac} A_c^\nu g_{\mu\mu} g_{\nu\nu} \end{aligned} \tag{7.24}$$

$$\frac{\partial \mathcal{L}}{\partial A_a^\mu} = -\frac{1}{4} \left(\left(\frac{\partial F_d^{\mu\nu}}{\partial A_a^\mu} \right) F_{\mu\nu}^d + F_d^{\mu\nu} \left(\frac{\partial F_{\mu\nu}^d}{\partial A_a^\mu} \right) \right) \tag{7.25}$$

$$= \frac{1}{4} g f_{dac} \left(A_c^\nu F_{\mu\nu}^d + F_d^{\mu\nu} A_c^\nu g_{\mu\mu} g_{\nu\nu} \right)$$

$$= \frac{1}{4} g f_{dac} \left(A_c^\nu F_{\mu\nu}^d + A_c^\nu g_{\mu\alpha} g_{\nu\beta} F_d^{\alpha\beta} \right)$$

$$= \frac{1}{4} g f_{dac} \left(A_c^\nu F_{\mu\nu}^d + A_c^\nu F_{\mu\nu}^d \right)$$

$$= \frac{1}{2} g f_{dac} A_c^\nu F_{\mu\nu}^d = \frac{1}{2} g f_{acd} A_c^\nu F_{\mu\nu}^d$$

$$= \frac{1}{2} g f_{abc} A_b^\nu F_{\mu\nu}^c$$

$$\frac{\partial F_d^{\mu\nu}}{\partial \left(\partial^\nu A_a^\mu \right)} = \frac{\partial}{\partial \left(\partial^\nu A_a^\mu \right)} \left(\partial^\mu A_d^\nu - \partial^\nu A_d^\mu \right) = -\delta_{ad}$$

$$\frac{\partial F_{\mu\nu}^d}{\partial \left(\partial^\nu A_a^\mu \right)} = \frac{\partial}{\partial \left(\partial^\nu A_a^\mu \right)} \left(\partial_\mu A_\nu^d - \partial_\nu A_\mu^d \right)$$

$$= \frac{\partial}{\partial \left(\partial^\nu A_a^\mu \right)} \left(g_{\mu\alpha} g_{\nu\beta} \left(\partial^\alpha A_d^\beta - \partial^\beta A_d^\alpha \right) \right)$$

$$= g_{\mu\alpha} g_{\nu\beta} \delta_{ad} \delta_{\mu\beta} \delta_{\nu\alpha} - g_{\mu\alpha} g_{\nu\beta} \delta_{ad} \delta_{\nu\beta} \delta_{\mu\alpha}$$

$$= -\delta_{ad} \left(g_{\mu\nu} g_{\nu\mu} - g_{\mu\mu} g_{\nu\nu} \right)$$

$$\frac{\partial \mathcal{L}}{\partial \left(\partial^\nu A_a^\mu \right)} = -\frac{1}{4} \left(-\delta_{ad} F_{\mu\nu}^d - g_{\mu\mu} g_{\nu\nu} \delta_{ad} F_d^{\mu\nu} \right)$$

$$= \frac{1}{4} \left(F_{\mu\nu}^a + g_{\mu\alpha} g_{\nu\beta} F_a^{\alpha\beta} \right)$$

$$= \frac{1}{4} \left(F_{\mu\nu}^a + g_{\mu\alpha} g_{\nu\beta} F_a^{\alpha\beta} \right)$$

$$= \frac{1}{4} \left(F_{\mu\nu}^a + F_{\mu\nu}^a \right) = \frac{1}{2} F_{\mu\nu}^a$$

Inserting (7.23) and (7.24) to the Euler-Langange equations (7.19) gives

$$\partial^\nu F_{\mu\nu}^a - g f_{abc} A_b^\nu F_{\mu\nu}^c = 0$$

As $F^c_{\mu\nu} = -F^c_{\nu\mu}$ we can also write

$$\partial^\nu F^a_{\nu\mu} - g f_{abc} A^\nu_b F^c_{\nu\mu} = 0$$

and changing ν and μ yields (7.22)

$$\partial^\mu F^a_{\mu\nu} - g f_{abc} A^\mu_b F^c_{\mu\nu} = 0$$

EOP

The Lagrangian can be expressed as

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu}_a F^a_{\mu\nu} = -\frac{1}{2} F^{\mu\nu} F_{\mu\nu}|_{\nu>\mu} \tag{7.26}$$

$$= -\frac{1}{2}\left(F^{01}_a F^a_{01} + F^{02}_a F^a_{02} + F^{03}_a F^a_{03} + F^{12}_a F^a_{12} + F^{13}_a F^a_{13} + F^{23}_a F^a_{23}\right)$$

For Minkowski's metric

$$\mathcal{L} = -\frac{1}{2}\left(-(F^a_{01})^2 - (F^a_{02})^2 - (F^a_{03})^2 + (F^a_{12})^2 + (F^a_{13})^2 + (F^a_{23})^2\right) \tag{7.27}$$

since $F^a_{0j} = -F^{0j}_a$ and $F^a_{kj} = F^{kj}_a$ for $1 \leq j, k \leq 3$.

The Hamiltonian density of a scalar field φ is defined as

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \varphi\right)}\left(\partial_0 \varphi\right) - \mathcal{L} = \pi \partial_0 \varphi - \mathcal{L} \tag{7.28}$$

where

$$\pi = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \varphi\right)}$$

The energy of the field is a conserved property

$$P_0 = \int d^3x \mathcal{H} \tag{7.29}$$

In the case of a gauge field A^μ_a we define the Hamiltonian density as

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 A^\mu_a\right)}\left(\partial^0 A^\mu_a\right) - \mathcal{L} \tag{7.30}$$

where summation over a and μ is implied. For the Yang-Mills Lagrangian we have calculated in Lemma 7.2

$$\frac{\partial \mathcal{L}}{\partial \left(\partial_0 A_a^\mu\right)} \left(\partial^0 A_a^\mu\right) = \frac{1}{2} F_{\mu 0}^a \tag{7.31}$$

Thus

$$\mathcal{H} = \frac{1}{2} F_{\mu 0}^a \partial^0 A_a^\mu - \mathcal{L} \tag{7.32}$$

The energy of the field is a conserved property also in this case

$$P^0 = \int d^3x \mathcal{H}$$

As Minkowski's metric is indefinite, it is sometimes better to move to either positive or negative definite metric. A convinient choice for computations is the following metric

$$(g_{\mu\nu})_{\mu,\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{7.33}$$

We will call it negative definite Euclidean metric, though in $\mathbb{R}^4$ a proper metric should be positive definite. This negative definite metric has the advantage that if we do not raise or lower the indices for the x_0 coordinate, all formulas remain valid. When we do lower x_0 indices, as in (2.27), there is a change of sign. Additionally, the x_0 coordinate must be replaced by ix_0. This creates an additional i when derivating with respect to x_0.

The problem setting of CMI uses the more modern algebraic geometric formulation where the Yang-Mills action is

$$\mathcal{S} = \frac{1}{4g^2} \int TrF \wedge *F \tag{7.34}$$

Actually [1] calls this action the Lagrangian but the Langrangian is the property that is integrated over the space in action. This terminology is corrected in [3]. The Yang-Mills equations (2.22) can be expressed with the Hodge star operator as

$$0 = d_A F = d_a * F \quad F = dA + A \wedge A \tag{7.35}$$

where d_A is the gauge-covariant extension of the exterior derivative. This is described in a clearer way in [2]. The gauge field A is a one-form

$$A(x) = A^a_\mu(x) t^a dx^\mu \tag{7.36}$$

with the values on the Lie algebra of a compact simple Lie group G. The curvature is a two-form

$$\begin{aligned} F &= dA + A \wedge A \\ F &= F^a_{\mu\nu} t^a dx^\mu \wedge dx^\nu \\ F &= \partial_\mu A_\nu - \partial_\nu A_\mu + f^{abc} A^b_\mu A^c_\nu \end{aligned} \tag{7.37}$$

Instead of the Lagrangian (7.1) we define a four-form

$$\mathcal{A} = TrF \wedge *F = F^a_{\mu\nu} F^a_{\mu\nu} d^4 x \tag{7.38}$$

and the action is

$$\mathcal{S} = \frac{1}{4g^2} \int \mathcal{A} \tag{7.39}$$

There are differences in the normalization $-\frac{1}{2}$ in (7.1) and in the placement of the coupling constant g in (7.11) and (7.22). There is also a more essential difference in $\mathcal{A}$ compared to (7.1). The summation is $F^a_{\mu\nu} F^a_{\mu\nu}$ and not $F^{\mu\nu}_a F^a_{\mu\nu}$ as in (7.1). This causes a difference in (7.27) and it seems that CMI has wanted to pose the problem in Euclidean metric instead of Minkowski's metric. This is not essential, we get the same result, apart from a multiplicative constant, for both of the metrics (7.9) and (7.33).

Lemmas and theorems

Lemma 7.4 Let the gauge field satisfy $A^3_a = 0$ for every a. The Euler-Lagrange equations can be expressed as ($0 \le l, k \le 2$)

$$\partial^3 A^a_l = F^a_{l3} \tag{7.40}$$

$$\partial^3 \partial^3 A_k^a = \partial^l F_{lk}^a - g f_{abc} A_b^l F_{l3}^a \tag{7.41}$$

$$\partial^3 \partial^l A_l^a - g f_{abc} A_b^l F_{l3}^a = 0 \tag{7.42}$$

Proof. Let $l \in \{0, 1, 2\}$. Rewriting (7.22) and inserting the gauge $A_b^3 = 0$ yields

$$\partial^3 F_{3\nu}^a + \partial^l F_{l\nu}^a - g f_{abc} A_b^l F_{l\nu}^c = 0 \tag{7.43}$$

As $F_{33}^a = 0$ by (7.7) the case $\nu = 3$ yields

$$\partial^l F_{l3}^a - g f_{abc} A_b^l F_{l3}^c = 0 \tag{7.44}$$

Inserting $A_3^a = 0$ to (7.11) yields

$$F_{3l}^a = \partial_3 A_l^a \tag{7.45}$$

and inserting $\partial^3 = -\partial_3$ and $F_{3l}^a = -F_{l3}^a$ gives (7.41). Inserting (7.40) to (7.44) yields (7.42). The other values $k \in \{0, 1, 2\}$ in (7.43) give

$$-\partial^3 F_{3k}^a = \partial^l F_{lk}^a - g f_{abc} A_b^l F_{lk}^c \tag{7.46}$$

and inserting (7.45) yields

$$-\partial^3 \partial_3 A_k^a = \partial^l F_{lk}^a - g f_{abc} A_b^l F_{lk}^c$$

Changing $\partial_3 = -\partial^3$ gives (7.41). EOP

Lemma 7.5 Let the gauge field A_a^μ be of the form

$$A_a^\mu = s_a E^\mu \quad , \quad E^3 = 0 \tag{7.47}$$

Then the Euler-Lagrange equations are

$$\partial^3 A_l^a = F_{l3}^a \quad 0 \le l, k \le 2$$

$$\partial^3 \partial^3 A_k^a = \partial^l F_{lk}^a \tag{7.48}$$

$$\partial^3 \partial^l A_l^a = 0$$

Proof. Since $E^3 = 0$ Lemma 7.4 applies. By Lemma 7.2 $F^a_{\mu\nu}$ is of the form

$$F^a_{\mu\nu} = s_a G_{\mu\nu} \tag{7.49}$$

As in Lemma 7.2

$$f_{abc} A^\mu_b F^c_{\mu\nu} = \sum_{c>b} f_{abc} \left(s_b E^\mu s_c G_{\mu\nu} - s_c E^\mu s_b G_{\mu\nu} \right) = 0$$

since s_a and E^μ are scalars and commutate. EOP

Lemma 7.6 Let the gauge field $A^\mu_{a,m}$ be of the form

$$A^\mu_{a,m} = s_a E^\mu_m \quad , \quad E^3_m = 0$$

for some finite set of indices $m \in B$ and let us assume that each $A^\mu_{a,m}$ is a gauge fiend such that $A^\mu_{a,m}$ and the corresponding $F^{a,m}_{\mu\nu}$ satisfy the Euler-Lagrange equations (7.22). Then

$$A^\mu_a = \sum_m A^\mu_{a,m} \tag{7.50}$$

defines the curvature

$$F^a_{\mu\nu} = \sum_m F^{a,m}_{\mu\nu} \tag{7.51}$$

such that A^μ_a and $F^a_{\mu\nu}$ satisfy the Euler-Lagrange equations (7.22).

Proof. In this case (7.22) reduces to the linear equations (7.48). Thus, the sums (7.50), (7.51) also satisfy (7.48). The equations (7.17) show that $F^a_{\mu\nu}$ is the sum (7.51). EOP

Lemma 7.7 Let the gauge field A^μ_a of the form

$$A^\mu_a = s_a E^\mu \quad , \quad E^3 = 0$$

be a complex gauge field satisfying (7.22). Let the real and imaginary parts be

$$A^\mu_{a,R} = Re A^\mu_a \tag{7.52}$$

and

$$A^{\mu}_{a,I} = ImA^{\mu}_{a}$$

and the corresponding curvatures be

$$F^{a,R}_{\mu\nu} = ReF^{a}_{\mu\nu} \tag{7.53}$$

and

$$F^{a,I}_{\mu\nu} = ImF^{a}_{\mu\nu}$$

are real functions satisfying (7.48).

ıProof. The equations (7.22) reduce to (7.48) in this case. The equations (7.48) are linear and the coordinates x_μ, x^μ and partial derivatives $\partial^\mu, \partial_\mu$ are all real. Thus the real and imaginary parts of A^μ_a and $F^a_{\mu\nu}$ satisfy (7.48) separately. EOP

Lemma 7.8 Let $\alpha_{ij} \in \mathbb{C}$, $0 \leq i \leq 3$, $j = 1, 2, \ldots,$ and $d_k \neq 0$, $0 \leq k \leq 2$, satisfy for every j

$$\alpha^2_{3j} = \sum_{l=0}^{2} \alpha^2_{lj} \tag{7.54}$$

$$\sum_{l=0}^{2} d_l \alpha_{lj} = 0 \tag{7.55}$$

The condition

$$\alpha_{3j}\alpha_{3k} = \sum_{l=0}^{2} \alpha_{lj}\alpha_{lk} \tag{7.56}$$

for any k, j with $k > j$ holds if either

$$\sum_{l=0}^{2} d^2_l = 0 \tag{7.57}$$

or there exists a constant c that for every j

$$\frac{\alpha_{1j}}{\alpha_{2j}} = c \tag{7.58}$$

The inverse is also true: if (7.56) holds then either (7.57) or (7.58) must hold.

Proof. If every $d_l = 0$ then (7.57) holds, thus we assume that at least one $d_l \neq 0$. By symmetry we may assume $d_0 \neq 0$. Squaring (7.56) and inserting (7.54) yields

$$\sum_{l=0}^{2}\sum_{m=0}^{2}\alpha_{lj}^2\alpha_{mk}^2 = \sum_{l=0}^{2}\sum_{m=0}^{2}\alpha_{lj}\alpha_{lk}\alpha_{mj}\alpha_{mk}$$

Separating α_{0j} terms gives

$$\alpha_{0j}^2(\alpha_{1k}^2+\alpha_{2k}^2)+\alpha_{0k}^2(\alpha_{1k}^2+\alpha_{2k}^2) \tag{7.59}$$

$$-2(\alpha_{1j}\alpha_{1k}+\alpha_{2j}\alpha_{2k})\alpha_{0j}\alpha_{0k}+(\alpha_{ij}\alpha_{2k}-\alpha_{1k}\alpha_{2j})^2 = 0$$

Inserting (7.55) in the form

$$\alpha_{0m} = -\frac{d_1}{d_0}\alpha_{1m} - \frac{d_2}{d_0}\alpha_{2m}$$

for $m \in \{j,k\}$ into (7.59) gives after some calculation

$$\left(\sum_{l=0}^{2} d_l^2\right)(\alpha_{1j}\alpha_{2k}-\alpha_{1k}\alpha_{2j})^2 = 0 \tag{7.60}$$

proving the lemma. EOP

Lemma 7.9 Let $d_l, \alpha_{lj} \in \mathbb{C}$, $1 \leq j \leq 3$, $0 \leq l \leq 2$, satisfy

$$\sum_{l=0}^{2} d_l\alpha_{lj} = 0$$

for every j. The vectors

$$\rho_j = \sum_{l=0}^{2}\alpha_{lj}x_l \tag{7.61}$$

are linearly dependent.

Proof. The determinant of this linear transform

$$\begin{vmatrix} \alpha_{01} & \alpha_{11} & \alpha_{21} \\ \alpha_{02} & \alpha_{12} & \alpha_{22} \\ \alpha_{03} & \alpha_{13} & \alpha_{23} \end{vmatrix} \tag{7.62}$$

gives zero when the condition

$$d_2\alpha_{2j} = -d_0\alpha_{0j} - d_1\alpha_{1j} \tag{7.63}$$

is inserted. EOP

Lemma 7.10 Let $\alpha_{ij} \in \mathbb{C}$, $0 \le i \le 3$, $j \ge 0$, $d_i \in \mathbb{C}$, $0 \le i \le 3$, satisfy

$$d_3 = 0 \tag{7.64}$$

$$\sum_{l=0}^{2} d_l^2 = 0 \tag{7.65}$$

$$\sum_{l=0}^{2} d_l\alpha_{lj} = 0 \tag{7.66}$$

for every j, and

$$\alpha_{3j}^2 = \sum_{l=0}^{2} \alpha_{lj}^2 \tag{7.67}$$

for every j. Let $h : \mathbb{C} \to \mathbb{C}$ be holomorphic in some open set U and

$$r_j = \sum_{\mu=0}^{3} \alpha_{\mu j}x_\mu \tag{7.68}$$

Then the gauge field

$$A_\mu^a = s_a d_\mu e^{\sum_j h(r_j)} \quad , \quad d_3 = 0 \tag{7.69}$$

defines $F_{\mu\nu}^a$ which satisfies the Euler-Lagrange equations (7.22).

Proof. We have expressed A_μ^a in contravariant coordinates x_ν instead of covariant coordinates x^ν since the derivations in (7.48) are all $\frac{\partial}{\partial x_\nu}$. From (7.69) follows that $A^3 = 0$ and the gauge field is of the form

$$A_a^\mu = s_a E^\mu$$

By Lemma 7.5 the Euler-Lagrange equations (7.22) reduce to (7.48). Inserting (7.69) to (7.17) yields

$$F^a_{\mu\nu} = s_a e^{\sum_j h(r_j)} \sum_j (d_\nu \alpha_{\mu j} - d_\mu \alpha_{\nu j}) h'(r_j) \tag{7.70}$$

Then

$$\partial^l F^a_{lk} = s_a e^{\sum_j h(r_j)} d_k \left(\sum_j \left(\sum_{l=0}^{2} \alpha^2_{lj} \right) h''(r_j) + \sum_{l=0}^{2} \left(\sum_j \alpha_{lj} h'(r_j) \right)^2 \right)$$

$$-s_a e^{\sum_j h(r_j)} \sum_j \alpha_{kj} \sum_{l=0}^{2} d_l \alpha_{lj} (h''(r_j)$$

$$-s_a e^{\sum_j h(r_j)} \sum_j \alpha_{kj} h'(r_j) \sum_m h'(r_m) \sum_{l=0}^{2} d_l \alpha_{lm}$$

Simplifying the expression by (3.27) and (3.28)

$$\partial^l F^a_{lk} = s_a e^{\sum_j h(r_j)} d_k \left(\sum_j \alpha^2_{3j} h''(r_j) + \sum_{l=0}^{2} \left(\sum_j \alpha_{lj} h'(r_j) \right)^2 \right)$$

Using Lemma 7.8 we can express

$$\sum_{l=0}^{2} \left(\sum_j \alpha_{lj} h'(r_j) \right)^2 = \sum_{l=0}^{2} \sum_j \alpha^2_{lj} \left(h'(r_j)\right)^2 + 2 \sum_{l=0}^{2} \sum_j \sum_{k>j} \alpha_{lj} \alpha_{lk} h'(r_j) h'(r_k)$$

$$= \sum_j \alpha^2_{3j} \left(h'(r_j)\right)^2 + 2 \sum_j \sum_{k>j} \alpha_{3j} \alpha_{3k} h'(r_j) h'(r_k)$$

Thus

$$\partial^l F^a_{lk} = s_a e^{\sum_j h(r_j)} d_k \left(\sum_j \alpha^2_{3j} h''(r_j) + \left(\sum_j \alpha_{3j} h'(r_j) \right)^2 \right)$$

$$= \frac{\partial^2}{\partial x_3^2} A_k^a = \partial^3 \partial^3 A_k^a$$

The first condition in (7.48) is obvious from (7.11) since $A_3^a = 0$, and the last condition in (7.48) hods since by (7.66)

$$\partial^3 \partial^l A_l^a = \partial^3 \left(s_a \sum_j \left(\sum_{l=0}^{2} d_l \alpha_{lj} \right) h'(r_j) e^{\sum_j h(r_j)} \right) = 0.$$

EOP

Let us select three linearly independent vectors $r_j = \sum_{\mu=0}^{3} \alpha_{\mu j} x_\mu$ and set the numbers d_μ as

$$d_0 = \sqrt{2}(1-i) \quad d_1 = d_2 = 1+i \quad d_3 = 0 \tag{7.71}$$

$$\begin{aligned} r_1 &= x_1 - x_2 + \sqrt{2} x_3 \\ r_2 &= x_1 - x_2 - \sqrt{2} x_3 \\ r_3 &= i \frac{1}{\sqrt{2}} x_0 - x_1 + \frac{1}{\sqrt{2}} x_3 \end{aligned} \tag{7.72}$$

Then

$$\begin{aligned} x_1 &= \frac{1}{4} r_1 - \frac{1}{4} r_2 - r_3 + i \frac{1}{\sqrt{2}} x_0 \\ x_2 &= -\frac{1}{4} r_1 - \frac{3}{4} r_2 - r_3 + i \frac{1}{\sqrt{2}} x_0 \\ x_3 &= \frac{1}{2\sqrt{2}} r_1 - \frac{1}{2\sqrt{2}} r_2 \end{aligned} \tag{7.73}$$

These numbers fill the conditions (7.64)-(7.67). We cannot get more than three linearly independent vectors. From Lemma 7.9 it follows that there are only two linearly independent linear combinations of $\{x_0, x_1, x_2\}$, and the third vector is obtained from the gauged coordinate x_3: the condition (7.67) allows two values for α_{3j}. Let us express r_j and $h(r_j)$ as sums of real and imaginary parts.

$$r_j = \rho_j + i\sigma_j \quad , \quad \rho_j, \sigma_j \in \mathbb{R}$$

$$\begin{aligned}\rho_1 &= x_1 - x_2 + \sqrt{2}x_3 \\ \rho_2 &= x_1 - x_2 - \sqrt{2}x_3 \\ \rho_3 &= -x_1 + \frac{1}{\sqrt{2}}x_3 \\ \sigma_1 = \sigma_2 = 0 &\quad \sigma_3 = \frac{1}{\sqrt{2}}x_0 \\ h(r_j) &= u(\rho_j, \sigma_j) + iv(\rho_j, \sigma_j)\end{aligned} \tag{7.74}$$

Thus, if $x_0 = 0$ then $\sigma_j = 0$ for every j. As h is holomorphic, u and v are harmonic functions on $\mathbb{R}^2$. Thus, u and v cannot be bounded on the whole $\mathbb{R}^2$, but they can be bounded on a strip $|x_0| \leq M$ for a finite M. Assuming that $h(r_j)$ goes sufficiently fast to zero if $|\rho_j|$ grows, then for any fixed value of x_0 the integral of the Euclidean norm of the gauge potential (7.69) over the space coordinates x_1, x_2, x_3 is finite. Also the path integral from finite time t' to another finite time t'' is finite. We have much freedom in selecting $u(\rho, 0)$. We can choose a real analytic function $f : \mathbb{R} \to \mathbb{R}$ that vanishes when $|\rho|$ grows, set $u(\rho, 0) = f(\rho)$ and extend u to a holomorphic function h. We should expect the solution to behave in the way (7.74) describes. It is a localized gauge field, gauge boson, which moves in the x_1, x_2 direction with the speed of light as a function of x_0. We select a concrete case that gives easy calculations. Let

$$f(\rho_j) = -\beta^2 \rho_j^2 \tag{7.75}$$

and extend it to

$$h(r_j) = -\beta^2 r_j^2 \tag{7.76}$$

The real and imaginary parts of $d_\mu = c_\mu + ie_\mu$ are

$$\begin{aligned} c_0 = \sqrt{2} \quad c_1 = c_2 = 1 \quad c_3 = 0 \\ e_0 = -\sqrt{2} \quad e_1 = e_2 = 1 \quad e_3 = 0 \end{aligned} \tag{7.77}$$

We evaluate the gauge potential at $x_0 = 0$ and take the real part.

Lemma 7.11 Let the gauge field be

$$A_\mu^a = s_a d_\mu e^{-\beta^2 \sum_{j=1}^3 r_j^2} \tag{7.78}$$

where r_j and d_μ are as in (7.71)-(7.72) and $\beta, s_a \in \mathbb{R}$. Then

$$A_\mu^{a,R}(0, x_1, x_2, x_3) = Re A_\mu^a(0, x_1, x_2, x_3) = s_a c_\mu e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \tag{7.79}$$

$$F_{\mu\nu}^{a,R}(0, x_1, x_2, x_3) = Re F_{\mu\nu}^a(0, x_1, x_2, x_3)$$

$$= -s_a 2\beta^2 e^{-\beta^2 \sum_j \rho_j^2} \sum_j Re(d_\nu \alpha_{\mu j} - d_\mu \alpha_{\nu j})\rho_j \tag{7.80}$$

Proof. Inserting $x_0 = 0$ to (7.78) and (7.70) yields the result. EOP

We need the Gaussian integrals

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2} = \sqrt{2\pi} a^{-\frac{1}{2}}$$

$$\int_{-\infty}^{\infty} x e^{-\frac{1}{2}ax^2} = 0 \tag{7.81}$$

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}ax^2} = \sqrt{2\pi} a^{-\frac{3}{2}}$$

Lemma 7.12 Let the gauge field satisfy

$$A_\mu^{a,R}(0, x_1, x_2, x_3) = s_a c_\mu e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \tag{7.82}$$

where ρ_j and c_μ are as in (3.35),(3.38) and $\beta, s_a \in \mathbb{R}$. Then

$$\int d^3x (A_k^a(0, x_1, x_2, x_3))^2 = s_a^2 c_k^2 (\frac{\pi}{2})^{\frac{3}{2}} \frac{1}{\beta^3} \tag{7.83}$$

Proof. We change the variables to y_1, y_2, y_3

$$y_1 = \sqrt{3}x_1 - \frac{2}{\sqrt{3}}x_2 - \frac{1}{\sqrt{6}}x_3$$

$$y_2 = \sqrt{\frac{2}{3}}x_2 - \frac{1}{\sqrt{3}}x_3 \tag{7.84}$$

$$y_3 = 2x_3$$

Then

$$\sum_{j=1}^{3} \rho_j^2 = y_1^2 + y_2^2 + y_3^2$$

As y_2 and y_3 are not functions of x_1 we can change the order of integration

$$\int d^3x e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} = \int d^3x e^{-\beta^2 \sum_{j=1}^3 y_j^2}$$

$$= \int d^2x e^{-\beta^2(y_2^2+y_3^2)} \int dx_1 e^{-\frac{1}{2}(\sqrt{2}\beta)^2 y_1^2}$$

$$= \int d^2x e^{-\beta^2(y_2^2+y_3^2)} \frac{1}{\sqrt{3}} \int dy_1 e^{-\frac{1}{2}(\sqrt{2}\beta)^2 y_1^2}$$

$$= \int d^2x e^{-\beta^2(y_2^2+y_3^2)} \frac{1}{\sqrt{3}} \sqrt{2\pi}(\sqrt{2}\beta)^{-1}$$

As y_3 is not a function of x_2 we can change the order of integration

$$= \frac{1}{\sqrt{3}} \sqrt{2\pi}(\sqrt{2}\beta)^{-1} \int dx_3 e^{-\beta^2 y_3^2} \int dx_2 e^{-\frac{1}{2}(\sqrt{2}\beta)^2 y_2^2}$$

$$= \frac{1}{\sqrt{3}} \sqrt{\frac{3}{2}} \sqrt{2\pi}(\sqrt{2}\beta)^{-1} \int dx_3 e^{-\beta^2 y_3^2} \int dy_2 e^{-\frac{1}{2}(\sqrt{2}\beta)^2 y_2^2}$$

$$= \frac{1}{\sqrt{3}} \sqrt{\frac{3}{2}} (2\pi)(\sqrt{2}\beta)^{-2} \int dx_3 e^{-\beta^2 y_3^2}$$

$$= \frac{1}{\sqrt{3}} \sqrt{\frac{3}{2}} \frac{1}{2} (2\pi)(\sqrt{2}\beta)^{-2} \int dy_3 e^{-\beta^2 y_3^2}$$

$$= \frac{1}{\sqrt{3}} \sqrt{\frac{3}{2}} \frac{1}{2} (2\pi)^{\frac{3}{2}} (\sqrt{2}\beta)^{-3}$$

Thus

$$\int d^3x e^{-2\beta^2 \sum \rho_j^2} = \left(\frac{\pi}{2}\right)^{\frac{3}{2}} \frac{1}{\beta^3}$$

EOP

Lemma 7.13 Let the gauge field satisfy

$$A_\mu^{a,R}(0, x_1, x_2, x_3) = s_a c_\mu e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \tag{7.85}$$

where ρ_j and c_μ are as in (7.74), (7.77) and $\beta, s_a \in \mathbb{R}$, and

$$\mathcal{L}_R = -\frac{1}{4} F_{a,R}^{\mu\nu} F_{\mu\nu}^{a,R}$$

Then in Minkowski's metric (7.9) at $x_0 = 0$

$$\mathcal{L}_R = 0 \tag{7.86}$$

while in the negative definite metric (7.33) at $x_0 = 0$

$$\mathcal{L}_R = -\frac{1}{2} \sum_a (2\beta^2 s_a)^2 e^{-2\beta^2 \sum_j \rho_j^2} 4(4\rho_1^2 + 4\rho_2^2 + \rho_3^2 - 4\rho_2\rho_3) \tag{7.87}$$

Proof. In Minkowski's metric $\mathcal{L}$ is given by (7.27). Thus

$$\mathcal{L}_R = -\frac{1}{2} \left(-(F_{01}^{a,R})^2 - (F_{02}^{a,R})^2 - (F_{03}^{a,R})^2 + (F_{12}^{a,R})^2 + (F_{13}^{a,R})^2 + (F_{23}^{a,R})^2 \right) \tag{7.88}$$

From (7.80) we see that

$$F_{\mu\nu}^{a,R}(0, x_1, x_2, x_3) = -s_a 2\beta^2 e^{-\beta^2 \sum_j \rho_j^2} \sum_j Re(d_\nu \alpha_{\mu j} - d_\mu \alpha_{\nu j}) \rho_j \tag{7.89}$$

The parameters selected in (7.71)-(7.72) are

$$\alpha_{01} = 0 \quad \alpha_{02} = 0 \quad \alpha_{03} = i\frac{1}{\sqrt{2}}$$

$$\alpha_{11} = 1 \quad \alpha_{12} = 1 \quad \alpha_{13} = -1$$

$$\alpha_{21} = -1 \quad \alpha_{22} = -1 \quad \alpha_{23} = 0$$

$$\alpha_{31} = \sqrt{2} \quad \alpha_{32} = -\sqrt{2} \quad \alpha_{33} = \frac{1}{\sqrt{2}}$$

$$c_0 = \sqrt{2} \quad c_1 = c_2 = 1 \quad c_3 = 0$$

$$e_0 = -\sqrt{2} \quad e_1 = e_2 = 1 \quad e_3 = 0$$

Let us compute the needed components

$$\sum_{j=1}^{3} \alpha_{0j}\rho_j = i\frac{1}{\sqrt{2}}\rho_3 \quad \sum_{j=1}^{3} \alpha_{1j}\rho_j = \rho_1 + \rho_2 - \rho_3$$

$$\sum_{j=1}^{3} \alpha_{2j}\rho_j = -\rho_1 - \rho_2 \quad \sum_{j=1}^{3} \alpha_{3j}\rho_j = \sqrt{2}\rho_1 - \sqrt{2}\rho_2 + \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} Re(d_1\alpha_{0j} - d_0\alpha_{1j})\rho_j = -\frac{1}{\sqrt{2}}\rho_3 - c_0\sum_{j=1}^{3} \alpha_{1j}\rho_j \tag{7.90}$$

$$= -\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_1 + \rho_2 - \rho_3)$$

$$\sum_{j=1}^{3} Re(d_2\alpha_{0j} - d_0\alpha_{2j})\rho_j = -\frac{1}{\sqrt{2}}\rho_3 - c_0\sum_{j=1}^{3} \alpha_{2j}\rho_j \tag{7.91}$$

$$= -\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(-\rho_1 - \rho_2)$$

$$\sum_{j=1}^{3} Re(d_3\alpha_{0j} - d_0\alpha_{3j})\rho_j = -c_0\sum_{j=1}^{3} \alpha_{3j}\rho_j \tag{7.92}$$

$$= -2\rho_1 + 2\rho_2 - \rho_3$$

$$\sum_{j=1}^{3} Re(d_2\alpha_{1j} - d_1\alpha_{2j})\rho_j = c_2\sum_{j=1}^{3} \alpha_{1j}\rho_j - c_1\sum_{j=1}^{3} \alpha_{2j}\rho_j$$

$$= 2\rho_1 + 2\rho_2 - \rho_3$$

$$\sum_{j=1}^{3} Re(d_3\alpha_{1j} - d_1\alpha_{3j})\rho_j = c_3\sum_{j=1}^{3} \alpha_{1j}\rho_j - c_1\sum_{j=1}^{3} \alpha_{3j}\rho_j$$

$$= -\sqrt{2}\rho_1 + \sqrt{2}\rho_2 - \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} Re(d_3\alpha_{2j} - d_2\alpha_{3j})\rho_j = c_3 \sum_{j=1}^{3} \alpha_{2j}\rho_j - c_2 \sum_{j=1}^{3} \alpha_{3j}\rho_j$$

$$= -\sqrt{2}\rho_1 + \sqrt{2}\rho_2 - \frac{1}{\sqrt{2}}\rho_3$$

The sum of the squares with the signs as in (7.88) is

$$-(-\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_1 + \rho_2 - \rho_3))^2 - (-\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(-\rho_1 - \rho_2))^2$$

$$-(-2\rho_1 + 2\rho_2 - \rho_3)^2 + (2\rho_1 + 2\rho_2 - \rho_3)^2$$

$$+(-\sqrt{2}\rho_1 + \sqrt{2}\rho_2 - \frac{1}{\sqrt{2}}\rho_3)^2 + (-\sqrt{2}\rho_1 + \sqrt{2}\rho_2 - \frac{1}{\sqrt{2}}\rho_3)^2$$

$$= 0$$

Inserting the sum to (7.88) and calculating (7.87) yields

$$\mathcal{L}_R = \frac{1}{2}\sum_a (2\beta^2 s_a)^2 e^{-2\beta^2 \sum_j \rho_j^2} 0 = 0$$

In the negative definite metric (7.33) holds $g_{\mu\mu} = -1$ for all μ, so

$$\mathcal{L}_R = -\frac{1}{2}\left((F_{01}^{a,R})^2 + (F_{02}^{a,R})^2 + (F_{03}^{a,R})^2 + (F_{12}^{a,R})^2 + (F_{13}^{a,R})^2 + (F_{23}^{a,R})^2\right) \tag{7.93}$$

Then the sum of the terms is

$$\mathcal{L}_R = -\frac{1}{2}\sum_a (2\beta^2 s_a)^2 e^{-2\beta^2 \sum_j \rho_j^2} 4(4\rho_1^2 + 4\rho_2^2 + \rho_3^2 - 4\rho_2\rho_3)$$

EOP

Lemma 7.14 Let the gauge field satisfy

$$A_\mu^{a,R}(0, x_1, x_2, x_3) = s_a c_\mu e^{-\beta^2 \sum_{j=1}^{3} \rho_j^2} \tag{7.94}$$

where ρ_j and c_μ are as in (7.74), (7.77) and $\beta, s_a \in \mathbb{R}$. Then

$$\int d^3\mathcal{L}_R = -\frac{1}{\beta}\frac{\pi^{\frac{3}{2}}}{16}\sum_a s_a^2 B \tag{7.95}$$

where in Minkowski's metric at $x_0 = 0$

$$B = 0$$

In the negative definite metric of (7.33)

$$B = \frac{13}{3} + \frac{2}{3} + 4$$

Proof. From (7.74) and (7.84) follows that

$$\rho_1 = \frac{1}{\sqrt{3}}y_1 - \frac{1}{\sqrt{6}}y_2 + \frac{1}{\sqrt{2}}y_3$$

$$\rho_2 = \frac{1}{\sqrt{3}}y_1 - \frac{1}{\sqrt{6}}y_2 - \frac{1}{\sqrt{2}}y_3$$

$$\rho_3 = -\frac{1}{\sqrt{3}}y_1 - \sqrt{\frac{2}{3}}y_2$$

For Minskowski's metric

$$P(\rho) = 0 = B_1y_1^2 + B_2y_2^2 + B_3y_3^2 + B_4y_1y_2 + B_5y_1y_3 + B_6y_2y_3$$

where $B_k = 0$ for all k. For the metric in (7.33)

$$P(\rho) = 4\rho_1^2 + 4\rho_2^2 + \rho_3^2 - 4\rho_2\rho_3$$

$$= B_1y_1^2 + B_2y_2^2 + B_3y_3^2 + B_4y_1y_2 + B_5y_1y_3 + B_6y_2y_3$$

where

$$B_1 = \frac{13}{3} \quad B_2 = \frac{2}{3} \quad B_3 = 4$$

$$B_4 = -\frac{4}{3}\sqrt{2} \quad B_5 = -\frac{4}{\sqrt{6}} \quad B_6 = -\frac{4}{\sqrt{3}}$$

We do the integration with generic parameters B_j. Then

$$\int d^3xe^{-\frac{1}{2}(2\beta)^2(\rho_1^2+\rho_2^2+\rho_3^2)}P(\rho)$$

$$= \int d^3xe^{-\frac{1}{2}(2\beta)^2(y_1^2+y_2^2+y_3^2)}(B_1y_1^2+B_2y_2^2+B_3y_3^2+B_4y_1y_2+B_5y_1y_3+B_6y_2y_3)$$

As y_2 and y_3 are not functions of x_1 we can change the order of integration and change the integration parameter x_1 to y_1.

$$= \int d^2xe^{-\frac{1}{2}(2\beta)^2(y_2^2+y_3^2)}\left(B_1\int dx_1y_1^2e^{-\frac{1}{2}(2\beta)^2y_1^2}\right.$$

$$+(B_2y_2^2+B_3y_3^2+B_6y_2y_3)\int dx_1e^{-\frac{1}{2}(2\beta)^2y_1^2}$$

$$\left.+(B_4y_2+B_5y_3)\int dx_1y_1e^{-\frac{1}{2}(2\beta)^2y_1^2}\right)$$

$$= \frac{1}{\sqrt{3}}\int d^2xe^{-\frac{1}{2}(2\beta)^2(y_2^2+y_3^2)}\left(B_1\int dy_1y_1^2e^{-\frac{1}{2}(2\beta)^2y_1^2}\right.$$

$$+(B_2y_2^2+B_3y_3^2+B_6y_2y_3)\int dy_1e^{-\frac{1}{2}(2\beta)^2y_1^2}$$

$$\left.+(B_4y_2+B_5y_3)\int dy_1y_1e^{-\frac{1}{2}(2\beta)^2y_1^2}\right)$$

$$= \frac{1}{\sqrt{3}}\int d^2xe^{-\frac{1}{2}(2\beta)^2(y_2^2+y_3^2)}\left(B_1\sqrt{2\pi}\frac{1}{(2\beta)^3}\right.$$

$$\left.+(B_2y_2^2+B_3y_3^2+B_6y_2y_3)\sqrt{2\pi}\frac{1}{2\beta}\right)$$

As y_3 is not a function of x_2 we can change the order of integration and change the integration parameter x_2 to y_2.

$$= \frac{1}{\sqrt{3}}\sqrt{\frac{3}{2}}\sqrt{2\pi}\int dx_3 e^{-\frac{1}{2}(2\beta)^2 y_3^2}\left(B_1\frac{1}{(2\beta)^3}\int dy_2 e^{-\frac{1}{2}(2\beta)^2 y_2^2}\right.$$
$$+B_2\frac{1}{2\beta}\int dy_2 y_2^2 e^{-\frac{1}{2}(2\beta)^2 y_2^2}$$
$$+B_3 y_3^2\frac{1}{2\beta}\int dy_2 e^{-\frac{1}{2}(2\beta)^2 y_2^2}$$
$$\left.+B_6 y_3\frac{1}{2\beta}\int dy_2 y_2 e^{-\frac{1}{2}(2\beta)^2 y_2^2}\right)$$
$$= \frac{1}{\sqrt{3}}\sqrt{\frac{3}{2}}\sqrt{2\pi}\int dx_3 e^{-\frac{1}{2}(2\beta)^2 y_3^2}\left(B_1\frac{1}{(2\beta)^3}\sqrt{2\pi}\frac{1}{2\beta}\right.$$
$$\left.+B_2\frac{1}{2\beta}\sqrt{2\pi}\frac{1}{(2\beta)^3}\ +B_3 y_3^2\frac{1}{2\beta}\sqrt{2\pi}\frac{1}{2\beta}\right)$$
$$= \frac{1}{\sqrt{3}}\sqrt{\frac{3}{2}\frac{1}{2}}2\pi\int dy_3 e^{-\frac{1}{2}(2\beta)^2 y_3^2}\left(B_1\frac{1}{(2\beta)^4}\right.$$
$$\left.+B_2\frac{1}{(2\beta)^4}+B_3 y_3^2\frac{1}{(2\beta)^2}\right)$$
$$= \frac{1}{\sqrt{3}}\sqrt{\frac{3}{2}\frac{1}{2}}(2\pi)^{\frac{3}{2}}(B_1+B_2+B_3)\frac{1}{(2\beta)^5}$$
$$= \pi^{\frac{3}{2}}\frac{1}{(2\beta)^5}(B_1+B_2+B_3)$$

EOP

Lemma 7.15 Let the gauge field satisfy

$$A_\mu^{a,R}(0,x_1,x_2,x_3) = s_a c_\mu e^{-\beta^2\sum_{j=1}^3 \rho_j^2} \tag{7.96}$$

where ρ_j and c_μ are as in (7.74), (7.77) and $\beta, s_a \in \mathbb{R}$. Then at $x_0 = 0$

$$\int d^3\mathcal{H}_R = \frac{1}{\beta}\frac{\pi^{\frac{3}{2}}}{16}\sum_a s_a^2 B \tag{7.97}$$

where in Minkowski's metric

$$B = 0$$

and in the metric (7.33) we get

$$B = \frac{13}{3} + \frac{2}{3} + 4$$

Proof. From (7.32) for the real gauge field $A_\mu^{a,R}$

$$\mathcal{H}_R = \frac{1}{2} F_{\mu 0}^{a,R} \partial^0 A_{a,R}^\mu - \mathcal{L}_R \tag{7.98}$$

From Lemma 7.11

$$A_\mu^{a,R} = s_a c_\mu e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \cos\left(\frac{\beta^2}{\sqrt{2}} x_0\right) + s_a e_\mu e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \sin\left(\frac{\beta^2}{\sqrt{2}} x_0\right)$$

Thus

$$\partial^0 A_\mu^{a,R} = s_a e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \frac{\beta^2}{\sqrt{2}} \left(-c_\mu \sin\left(\frac{\beta^2}{\sqrt{2}} x_0\right) + e_\mu \cos\left(\frac{\beta^2}{\sqrt{2}} x_0\right)\right)$$

as $\partial^0 \sum_j \rho_j^2 = 0$. At $x_0 = 0$

$$\partial^0 A_\mu^{a,R} = s_a e_\mu \frac{\beta^2}{\sqrt{2}} e^{-\beta^2 \sum_{j=1}^3 \rho_j^2}$$

Thus

$$\begin{aligned} \partial^0 A_0^{a,R}(0, x_1, x_2, x_3) &= -s_a \beta^2 e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \\ \partial^0 A_1^{a,R}(0, x_1, x_2, x_3) &= s_a \frac{\beta^2}{\sqrt{2}} e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \\ \partial^0 A_2^{a,R}(0, x_1, x_2, x_3) &= s_a \frac{\beta^2}{\sqrt{2}} e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \\ \partial^0 A_3^{a,R}(0, x_1, x_2, x_3) &= 0 \end{aligned} \tag{7.99}$$

From (7.89)-(7.92)

$$\frac{1}{2} F_{00}^{a,R}(0, x_1, x_2, x_3) = 0$$

$$\frac{1}{2}F_{10}^{a,R}(0,x_1,x_2,x_3) = -\frac{1}{2}s_a 2\beta^2 e^{-\beta^2\sum_j \rho_j^2}(\frac{1}{\sqrt{2}}\rho_3 + \sqrt{2}(\rho_1+\rho_2-\rho_3))$$

$$\frac{1}{2}F_{20}^{a,R}(0,x_1,x_2,x_3) = -\frac{1}{2}s_a 2\beta^2 e^{-\beta^2\sum_j \rho_j^2}(\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_1+\rho_2))$$

$$\frac{1}{2}F_{30}^{a,R}(0,x_1,x_2,x_3) = -\frac{1}{2}s_a 2\beta^2 e^{-\beta^2\sum_j \rho_j^2}(2\rho_1 - 2\rho_2 - \rho_3)$$

Since

$$A^\mu_{a,R} = g^{\mu\nu}A^{a,R}_\nu$$

and in the metric (2.9) $A^a_0 = A^0_a,\ A^a_j = -A^j_a,\ j > 0,$

$$\partial^0 A^0_{a,R}(0,x_1,x_2,x_3) = \partial^0 A^{a,R}_0(0,x_1,x_2,x_3) = -s_a\beta^2 e^{-\beta^2\sum_{j=1}^3 \rho_j^2}$$

$$\partial^0 A^1_{a,R}(0,x_1,x_2,x_3) = -\partial^0 A^{a,R}_1(0,x_1,x_2,x_3) = -s_a\frac{\beta^2}{\sqrt{2}}e^{-\beta^2\sum_{j=1}^3 \rho_j^2}$$

$$\partial^0 A^2_{a,R}(0,x_1,x_2,x_3) = -\partial^0 A^{a,R}_2(0,x_1,x_2,x_3) = -s_a\frac{\beta^2}{\sqrt{2}}e^{-\beta^2\sum_{j=1}^3 \rho_j^2}$$

$$\partial^0 A^3_{a,R}(0,x_1,x_2,x_3) = 0$$

Thus

$$\frac{1}{2}F_{\mu 0}^{a,R}\partial^0 A^\mu_{a,R} = \sum_a\sum_{\mu=0}^3 \frac{1}{2}F_{\mu 0}^{a,R}\partial^0 A^\mu_{a,R}$$

$$= \sum_a\left(\frac{1}{2}F_{10}^{a,R}\partial^0 A^1_{a,R} + \frac{1}{2}F_{20}^{a,R}\partial^0 A^2_{a,R}\right)$$

$$= \beta^4\frac{1}{\sqrt{2}}\left(\sum_a s_a^2\right)e^{-2\beta^2\sum_{j=1}^3 \rho_j^2}(\frac{1}{\sqrt{2}}\rho_3 + \sqrt{2}(\rho_1+\rho_2-\rho_3))$$

$$+\beta^4\frac{1}{\sqrt{2}}\left(\sum_a s_a^2\right)e^{-2\beta^2\sum_j \rho_j^2}(\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_1+\rho_2))$$

$$= \beta^4\frac{1}{\sqrt{2}}\left(\sum_a s_a^2\right)e^{-2\beta^2\sum_{j=1}^3 \rho_j^2}(2\frac{1}{\sqrt{2}} - 1)\rho_3$$

Inserting y_1, y_2, y_3 from (7.84) allows us to perform the integration

$$\int d^3x \frac{1}{2} F^{a,R}_{\mu 0} \partial^0 A^{\mu}_{a,R} \tag{7.100}$$

$$= \int d^3x \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) e^{-2\beta^2 \sum_{j=1}^3 y_j^2} (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} y_3$$

$$= \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int d^2x y_3 e^{-2\beta^2(y_2^2+y_3^2)} \int dx_1 e^{-2\beta^2 y_1^2}$$

$$= \frac{1}{\sqrt{3}} \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int d^2x y_3 e^{-2\beta^2(y_2^2+y_3^2)} \int dy_1 e^{-2\beta^2 y_1^2}$$

$$= \frac{\sqrt{2\pi}}{2\beta} \frac{1}{\sqrt{3}} \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int d^2x y_3 e^{-2\beta^2(y_2^2+y_3^2)}$$

$$= \frac{\sqrt{2\pi}}{2\beta} \frac{1}{\sqrt{3}} \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int d^x_3 y_3 e^{-2\beta^2 y_3^2} \int dx_2 e^{-2\beta^2 y_2^2}$$

$$= \sqrt{\frac{3}{2}} \frac{\sqrt{2\pi}}{2\beta} \frac{1}{\sqrt{3}} \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int dx_3 y_3 e^{-2\beta^2 y_3^2} \int dy_2 e^{-2\beta^2 y_2^2}$$

$$= \sqrt{\frac{3}{2}} \left(\frac{\sqrt{2\pi}}{2\beta}\right)^2 \frac{1}{\sqrt{3}} \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int dx_3 y_3 e^{-2\beta^2 y_3^2}$$

$$= \frac{1}{2}\sqrt{\frac{3}{2}} \left(\frac{\sqrt{2\pi}}{2\beta}\right)^2 \frac{1}{\sqrt{3}} \beta^4 \frac{1}{\sqrt{2}} \left(\sum_a s_a^2\right) (2\frac{1}{\sqrt{2}} - 1)\frac{1}{2} \int dy_3 y_3 e^{-2\beta^2 y_3^2} = 0$$

Thus

$$\int d^3\mathcal{H}_R = -\int d^3\mathcal{L}_R$$

and (7.97) follows from Lemma 7.14. If the metric is as in (7.33) then $A_a^0 = -A_o^a$ but this term disappears and the integral in (7.100) still yields zero. If in addition to changing the metric there has been a replacement $x_0 \to ix_0$ as is often done in order to move from Minkowski's metric to Euclidean metric, derivation with respect to x_0 gives an additional i. This changes the coefficients c_j to e_i in some places but (7.100) still holds because the integral disappears because it is of first order in ρ_j, and that is also true for the metric (7.33) and a change $x_0 \to ix_0$. Thus, for the metric (7.33) we get another parameters than for Minkowski's metric but the form is the same. EOP

Theorem 7.16 Let the gauge field be

$$A^a_\mu = s_a d_\mu e^{-\beta^2 \sum_{j=1}^3 r_j^2} \tag{7.101}$$

where r_j and d_μ are as in (7.72),(7.71) and $\beta, s_a \in \mathbb{R}$. The real part is

$$A^{a,R}_\mu(0, x_1, x_2, x_3) = s_a c_\mu e^{-\beta^2 \sum_{j=1}^3 \rho_j^2} \tag{7.102}$$

where ρ_j and c_μ are as in (7.74),(7.77). The following statements hold

$$\int d^3x (A^a_k(0, x_1, x_2, x_3))^2 = s_a^2 c_k^2 \left(\frac{\pi}{2}\right)^{\frac{3}{2}} \frac{1}{\beta^3} \tag{7.103}$$

$$\int d^3x A^{a\,*}_\mu A^a_{mu} = \sum_a \sum_{k=0}^2 \int d^3x (A^a_k(0, x_1, x_2, x_3))^2$$

$$= \sum_a s_a^2 \sum_{k=0}^2 c_k^2 \left(\frac{\pi}{2}\right)^{\frac{3}{2}} \frac{1}{\beta^3} \tag{7.104}$$

where A^* denotes the complex conjugate of A. The real part of the gauge field $A^{a,R}_\mu$ and the real part of the curvature $F^{a,R}_{\mu\nu}$ satisfy the Lagrange-Euler equations

$$\mathcal{L}_R = -\frac{1}{4} F^{\mu\nu}_{a,R} F^{a,R}_{\mu\nu} \tag{7.105}$$

and the energy is

$$P^{0,R} = \int d^3x \mathcal{H}_R = \frac{1}{\beta} \frac{\pi^{\frac{3}{2}}}{16} \sum_a s_a^2 B \tag{7.106}$$

where in Minkowski's metric

$$B = 0$$

and in the metric (2.33)

$$B = \frac{13}{3} + \frac{2}{3} + 4$$

Proof. From Lemma 7.11 follows that (7.102) is the real part of (7.101). The claim (7.103) is shown in Lemma 7.12. The imaginary part in A^a_μ is a phase $e^{-i\sqrt{2}x_0}$ which cancels in $A^a_\mu * A^a_{mu}$, thus (7.104) holds. The real part of the gauge field and the curvature satisfy the Euler-Lagrange equations by Lemma 7.7, thus (7.105) holds. In Lemma 7.15 we showed that at $x_0 = 0$ equation (7.106) holds. As P^0 is a conserved property, see (7.29), (7.106) holds for all values of x_0. EOP

Theorem 7.17 Let $A = (A_{mu})_\mu$, $A_\mu = A^a_\mu t_a$ be a complex gauge field defined by

$$A^a_\mu = s_a d_\mu e^{-\beta^2 \sum_{j=1}^3 r_j^2} \tag{7.107}$$

The numbers r_j and d_μ are as in (7.73),(7.71) and $\beta, s_a \in \mathbb{R}$. The norm is

$$||A|| = \int d^3x A^{a\,*}_\mu A^a_\mu = \sum_a s_a^2 \sqrt{2}\pi^{\frac{3}{2}} \frac{1}{\beta^3} \tag{7.108}$$

where A^* denotes the complex conjugate of A. The gauge field and the corresponding curvature satisfy Euler-Lagrange equations for

$$\mathcal{L} = -\frac{1}{4} F_a^{\mu\nu} F^a_{\mu\nu} \tag{7.109}$$

In both metrics (7.9) and (7.33)

$$E_\beta = \frac{P^0}{||A||} = \beta^2 C \tag{7.110}$$

where

$$P^0 = \int d^3x \mathcal{H} = \sum_a s_a^2 \pi^{\frac{3}{2}} \sqrt{2} C \frac{1}{\beta} \tag{7.111}$$

and C is a nonegative constant.

Proof. In the case of a complex field, the Lagrangian has two parts, the real and the imaginary. If the field defines a solution to the Euler-Langange equations, the energy (7.29) is conserved. Thus, also the imaginary part is conserved though we only computed the real part. We get the same dependence of β for the imaginary part. For the real part of the Lagrangian we get from Theorem 7.16

$$\frac{P_{0,R}}{||A||} = \beta^2 \frac{1}{16\sqrt{2}} B \tag{7.112}$$

where we have inserted $\sum_{k=0}^{2} c_k^2 = 4$. Including the imaginary part changes the constant, but it is nonnegative. EOP

We can find a gauge field that gives positive energy for Minkowski's metric as as sum.

Lemma 7.18 Let the gauge field be

$$A_\mu^a = s_a d_{\mu 1} e^{-\beta^2 \sum_{j=1}^{3} r_{j2}^2} + s_a d_{\mu 2} e^{-\beta^2 \sum_{j=1}^{3} r_{j2}^2} \tag{7.113}$$

where $\beta, s_a \in \mathbb{R}$ and

$$r_{jk} = \rho_{j,k} + i\sigma_{j,k} = \sum_{\mu=0}^{3} \alpha_{\mu,j,k} x_\mu$$

$$d_{lk} = c_{lk} + ie_{lk}$$

$$\alpha_{011} = 0 \quad \alpha_{021} = 0 \quad \alpha_{031} = i\frac{1}{\sqrt{2}}$$

$$\alpha_{111} = 1 \quad \alpha_{121} = 1 \quad \alpha_{131} = -1$$

$$\alpha_{211} = -1 \quad \alpha_{221} = -1 \quad \alpha_{231} = 0$$

$$\alpha_{311} = \sqrt{2} \quad \alpha_{321} = -\sqrt{2} \quad \alpha_{331} = \frac{1}{\sqrt{2}}$$

$$c_{01} = \sqrt{2} \quad c_{11} = 1 \quad c_{21} = 1 \quad c_{31} = 0$$

$$e_{01} = -\sqrt{2} \quad e_{11} = 1 \quad e_{22} = 1 \quad e_{31} = 0$$

$$\alpha_{012} = 0 \quad \alpha_{022} = 0 \quad \alpha_{032} = i\frac{1}{\sqrt{2}}$$

$$\alpha_{112} = 1 \quad \alpha_{122} = 1 \quad \alpha_{132} = -1$$

$$\alpha_{212} = 1 \quad \alpha_{222} = 1 \quad \alpha_{232} = 0$$

$$\alpha_{312} = \sqrt{2} \quad \alpha_{322} = -\sqrt{2} \quad \alpha_{332} = \frac{1}{\sqrt{2}}$$

$$c_{02} = \sqrt{2} \quad c_{12} = 1 \quad c_{22} = -1 \quad c_{32} = 0$$

$$e_{02} = -\sqrt{2} \quad e_{12} = 1 \quad e_{22} = -1 \quad e_{32} = 0$$

Then

$$A_\mu^{a,R}(0, x_1, x_2, x_3) = s_a c_{\mu 1} e^{-\beta^2 \sum_{j=1}^3 \rho_{j2}^2} + s_a c_{\mu 2} e^{-\beta^2 \sum_{j=1}^3 \rho_{j2}^2} \tag{7.114}$$

and

$$\mathcal{L}_R = -\frac{1}{4} F_{a,R}^{\mu\nu} F_{\mu\nu}^{a,R}$$

In Minkowski's metric (2.9) at $x_0 = 0$

$$\mathcal{L}_R = -2\frac{1}{2} \sum_a (2\beta^2 s_a)^2 e^{-2\beta^2 \sum_j y_j^2}$$

$$\bullet \left(-\frac{13}{3} y_1^2 + 8y_2^2 - \frac{170}{21} y_3^2 + \frac{8}{21}\sqrt{14} y_1 y_3 \right) \tag{7.115}$$

where

$$y_1 = \sqrt{6}x_1 - \frac{1}{\sqrt{3}}x_3 \quad y_2 = 2x_2 \quad y_3 = \sqrt{\frac{14}{3}}x_3 \tag{7.116}$$

Proof. By Lemma 7.6 the sum of solutions satisfying (7.48) is also a solution satisfying (7.48). Most of the proof is as in Lemma 7.13. We are interested in the cross term that comes from squaring $F_{\mu\nu}^{a,R}$. As it has two components from the two fields in the sum, the squares of each field give two squares and a cross term (twice the product of the terms). Both of the squares disappear as in Lemma 7.13 but the cross term gives the term in (7.105) and it does not disappear. We compute only this term in detail. Let us notice that $r_{31} = r_{32}$ and for simplicity we will write

$$\rho_3 = \rho_{31} = \rho_{32} = -x_1 + \frac{1}{\sqrt{2}}x_3$$

We notice that

$$\sum_{j=1}^{3} \rho_{j1}^2 + \sum_{j=1}^{3} \rho_{j2}^2 = y_1^2 + y_2^2 + y_3^2 \tag{7.117}$$

Let us compute the needed components

$$\sum_{j=1}^{3} \alpha_{0j1}\rho_{j1} = i\frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} \alpha_{1j1}\rho_{j1} = \rho_{11} + \rho_{21} - \rho_3 = 2x_1 - 2x_2 - \rho_3$$

$$\sum_{j=1}^{3} \alpha_{2j1}\rho_{j1} = -\rho_{11} - \rho_{21} = -2x_1 + 2x_2$$

$$\sum_{j=1}^{3} \alpha_{3j1}\rho_{j1} = \sqrt{2}\rho_{11} - \sqrt{2}\rho_{21} + \frac{1}{\sqrt{2}}\rho_3 = 4x_3 + \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} \alpha_{0j2}\rho_{j2} = i\frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} \alpha_{1j2}\rho_{j2} = \rho_{12} + \rho_{22} - \rho_3 = 2x_1 + 2x_2 - \rho_3$$

$$\sum_{j=1}^{3} \alpha_{2j2}\rho_{j2} = \rho_{12} + \rho_{22} = 2x_1 + 2x_2$$

$$\sum_{j=1}^{3} \alpha_{3j2}\rho_{j2} = \sqrt{2}\rho_{12} - \sqrt{2}\rho_{22} + \frac{1}{\sqrt{2}}\rho_3 = 4x_3 + \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} Re(d_{11}\alpha_{0j1} - d_{01}\alpha_{1j1})\rho_{j1} = -\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_{11} + \rho_{21} - \rho_3) \tag{7.118}$$

$$= \frac{1}{\sqrt{2}}x_3 - 2\sqrt{2}x_1 + 2\sqrt{2}x_2$$

$$\sum_{j=1}^{3} Re(d_{12}\alpha_{0j2} - d_{02}\alpha_{1j2})\rho_{j2} = -\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_{12} + \rho_{22} - \rho_3) \tag{7.119}$$

$$= \frac{1}{\sqrt{2}}x_3 - 2\sqrt{2}x_1 - 2\sqrt{2}x_2$$

$$\sum_{j=1}^{3} Re(d_{21}\alpha_{0j1} - d_{01}\alpha_{2j1})\rho_{j1} = -\frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(-\rho_{11} - \rho_{21}) \tag{7.120}$$

$$= -\frac{1}{\sqrt{2}}x_3 + 2\sqrt{2}x_1 - 2\sqrt{2}x_2$$

$$\sum_{j=1}^{3} Re(d_{22}\alpha_{0j2} - d_{02}\alpha_{2j2})\rho_{j2} = \frac{1}{\sqrt{2}}\rho_3 - \sqrt{2}(\rho_{11} + \rho_{21}) \tag{7.121}$$

$$= \frac{1}{\sqrt{2}}x_3 - 2\sqrt{2}x_1 - 2\sqrt{2}x_2$$

$$\sum_{j=1}^{3} Re(d_{31}\alpha_{0j1} - d_{01}\alpha_{3j1})\rho_{j1} = -2\rho_{11} + 2\rho_{21} - \rho_3 \tag{7.122}$$

$$= -4\sqrt{2}x_3 - \rho_3$$

$$\sum_{j=1}^{3} Re(d_{32}\alpha_{0j2} - d_{02}\alpha_{3j2})\rho_{j2} = -2\rho_{12} + 2\rho_{22} - \rho_3 \tag{7.123}$$

$$= -4\sqrt{2}x_3 - \rho_3$$

$$\sum_{j=1}^{3} Re(d_{21}\alpha_{1j1} - d_{11}\alpha_{2j1})\rho_{j1} = 2\rho_{11} + 2\rho_{21} - \rho_3$$

$$= 4x_1 - 4x_2 - \rho_3$$

$$\sum_{j=1}^{3} Re(d_{22}\alpha_{1j2} - d_{12}\alpha_{2j2})\rho_{j2} = -2\rho_{12} - 2\rho_{22} + \rho_3$$

$$= -4x_1 - 4x_2 + \rho_3$$

$$\sum_{j=1}^{3} Re(d_{31}\alpha_{1j1} - d_{11}\alpha_{3j1})\rho_{j1} = -\sqrt{2}\rho_{11} + \sqrt{2}\rho_{21} - \frac{1}{\sqrt{2}}\rho_3 = -4x_3 - \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} Re(d_{32}\alpha_{1j2} - d_{12}\alpha_{3j2})\rho_{j2} = \sqrt{2}\rho_{12} + \sqrt{2}\rho_{22} + \frac{1}{\sqrt{2}}\rho_3 = 4x_3 + \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} Re(d_{31}\alpha_{2j1} - d_{21}\alpha_{3j1})\rho_{j1} = -\sqrt{2}\rho_{11} + \sqrt{2}\rho_{21} - \frac{1}{\sqrt{2}}\rho_3 = -4x_3 - \frac{1}{\sqrt{2}}\rho_3$$

$$\sum_{j=1}^{3} Re(d_{32}\alpha_{2j2} - d_{22}\alpha_{3j2})\rho_{j2} = -\sqrt{2}\rho_{12} + \sqrt{2}\rho_{22} - \frac{1}{\sqrt{2}}\rho_3 = -4x_3 - \frac{1}{\sqrt{2}}\rho_3$$

The cross term in Minkowski's metric is

$$-(\frac{1}{\sqrt{2}}\rho_3 - 2\sqrt{2}x_1 + 2\sqrt{2}x_2)(\frac{1}{\sqrt{2}}\rho_3 - 2\sqrt{2}x_1 - 2\sqrt{2}x_2)$$

$$-(-\frac{1}{\sqrt{2}}\rho_3 + 2\sqrt{2}x_1 - 2\sqrt{2}x_2)(\frac{1}{\sqrt{2}}\rho_3 - 2\sqrt{2}x_1 - 2\sqrt{2}x_2)$$

$$-(-4\sqrt{2}x_3 - \rho_3)^2$$

$$+(4x_1 - 4x_2 - \rho_3)(-4x_1 - 4x_2 + \rho_3)$$

$$+(-4x_3 - \frac{1}{\sqrt{2}}\rho_3)(4x_3 + \frac{1}{\sqrt{2}}\rho_3)$$

$$+(-4x_3 - \frac{1}{\sqrt{2}}\rho_3)^2$$

$$= -16x_1^2 + 16x_2^2 - 32x_3^2 - 2\rho_3^2 + \rho_3(8x_1 - 8\sqrt{2}x_3)$$

$$= -26x_1^2 + 16x_2^2 - 41x_3^2 + 14\sqrt{2}x_1x_3$$

$$= -\frac{13}{3}y_1^2 + 8y_2^2 - \frac{170}{21}y_3^2 + \frac{8}{21}\sqrt{14}y_1y_3$$

Inserting this result as in Lemma 7.13 yields the claim. EOP

Theorem 7.19 Let $A = (A_{mu})_\mu$, $A_\mu = A_\mu^a t_a$ be a complex gauge field as in Lemma 7.18. The gauge field and the corresponding curvature satisfy Euler-Lagrange equations for

$$\mathcal{L} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \tag{7.124}$$

In Minkowski's metric

$$E_\beta = \frac{P^0}{||A||} = \beta^2 C \tag{7.125}$$

where C is a positive constant.

Proof. The Lagrangian is computed in Lemma 7.18. As in Lemma 7.12 the norm $||A||$ is not zero and depends on β as β^{-3}. As in Lemma 7.14 the Lagrangian in (7.54) when integrated over the space coordinates is proportional to $B = B_1 + B_2 + B_3 = -\frac{13}{3} + 8 - \frac{170}{21}$ which is nonzero and the integral over space coordinates does not vanish. As in Lemma 7.15 the first part of the Hamiltonian density (7.30) does not contribute to the integral:

$$\int d^3\mathcal{H}_R = -\int d^3\mathcal{L}_R$$

The rest is as in Theorem 7.17. EOP

Mass Gap and Quantization of Yang-Mills fields

The first question is what is mass gap. L. Faddeev explains the issue in [2] but let us proceed in a similar way as in [6] from quantum mechanics and scalar quantum field theory to quantum Yang-Mills theory. We take a simple scalar wave function of one variable

$$\varphi(x_1) = e^{-\frac{1}{2}ax_1^2} \tag{7.126}$$

Then

$$\left(\frac{1}{a^2}\frac{\partial^2}{\partial x_1^2} + \frac{1}{a}\right)\varphi(x_1) = x_1^2\varphi(x_1) \tag{7.127}$$

It follows that

$$\frac{1}{a}\int_{-\infty}^{\infty} dx_1\varphi(x_1) = \sqrt{2\pi}a^{\frac{3}{2}} \tag{7.128}$$

while also

$$\int_{-\infty}^{\infty} dx_1 x_1^2\varphi(x_1) = \sqrt{2\pi}a^{-\frac{3}{2}} \tag{7.129}$$

The function $\varphi(x_1)$ is time-independent as it does not depend on x_0. We can consider it as a state in the Schrödinger picture

$$|q >= |q >_S \tag{7.130}$$

We can consider

$$\hat{A} = A(x_1) = \frac{1}{a^2}\frac{\partial^2}{\partial x_1^2} + \frac{1}{a} \tag{7.131}$$

as an operator acting on the state $|q>$. In order to take an inner product of $\hat{A}|q>$ with another state $|q'>$ corresponing to the field $\varphi'(x_1)$ it is more convenient to define the operator as

$$\hat{H} = H(x_1', x_1) = \left(\frac{1}{a^2}\frac{\partial}{\partial x_1'}\frac{\partial}{\partial x_1} + \frac{1}{a}\right)\delta(x_1 - x_1') \tag{7.132}$$

Then

$$< q'|\hat{H}|q >= \int dx_1' \int dx_1 \varphi'(x_1')H(x_1', x_1)\varphi(x_1) \tag{7.133}$$

Especially

$$< q|\hat{H}|q >= \int dx_1' \int dx_1 \varphi(x_1')H(x_1', x_1)\varphi(x_1) \tag{7.134}$$

$$= \int dx_1 x_1^2 e^{-2\frac{1}{2}ax_1^2} = \sqrt{2\pi}(\sqrt{2}a)^{-\frac{3}{2}}$$

while

$$< q|q >= \int dx_1' \int dx_1 \varphi(x_1')\delta(x_1 - x_1')\varphi(x_1) \quad (7.135)$$

$$= \int dx_1 \phi(x_1)^* \phi(x_1) = \int dx_1 e^{-2\frac{1}{2}ax_1^2} = \sqrt{2\pi}(\sqrt{2}a)^{-\frac{1}{2}}$$

Thus

$$< q|\hat{H}|q >= E < q|q > \qquad E = \frac{1}{\sqrt{2}a} \quad (7.136)$$

Thus, E is the expectation value of the operator $\hat{H}$ at the state $|q >$. Let us assume that the state $|q >$ is expressed as a linear combination of the eigenstates of the Hamiltonian operator $\hat{H}$. If E can be selected arbitrarily small then we can select a sequence of states $|q_n >$ where E_n goes to zero. This means that either the sequence of the states $|q_n >$ converges to the vacuum state, or that there is no minimal positive eigenstate for $\hat{H}$. The state where $|q_n >$ converges if $a \to \infty$ is zero, which is not a vacuum state. We conclude that there is no minimal positive eighenvalue for $\hat{H}$, i.e., there is no mass gap. We can also write the equation with the Hamiltonian density $\mathcal{H}$

$$\int dx_1' \int dx_1 \varphi(x_1')H(x_1', x_1)\varphi(x_1) = \int dx_1 \mathcal{H} \quad (7.137)$$

The set of eigenvalues of the Hamiltonian operator forms the energy-mass spectrum of the field. The zero function $\varphi(x_1) = 0$ always satisfies the eigenvalue equation but it is not an acceptable eigenstate since it has zero form. There is assumed to exist an eigenstate with eigenvalue zero, the vacuum. The vacuum is not unique in all theories, but it must be unique in a theory filling Wightman's axioms. If there is a minimum positive value E in the energy-mass spectrum, we say that there is a mass gap. The eigenstates are closely related to a parameter called mass because the physical interpretation of the parameter m in an equation

$$\left(\partial_\mu \partial^\mu + m^2\right)\varphi = 0 \quad (7.138)$$

is mass.

Let us now proceed to find the Hamiltonian operator for the Hamiltonian density $\mathcal{H}_R$ in Lemma 7.15. We notice that in Lemma 7.15

$$\int d^3\mathcal{H}_R = -\int d^3\mathcal{L}_R$$

From Lemmas 7.13 and 7.14 we see that the Lagrangian can be expressed in variables y_1, y_2, y_3 as

$$\mathcal{L}_R = \frac{1}{2}\sum_a (2\beta^2 s_a)^2 e^{-2\beta^2 \sum_j \rho_j^2} P(\rho)$$

$$= \frac{1}{2}\sum_a (2\beta^2 s_a)^2 e^{-2\beta^2 \sum_j y_j^2} 4(B_1y_1^2+B_2y_2^2+B_3y_3^2+B_4y_1y_2+B_5y_1y_3+B_6y_2y_3)$$

We can ignore the terms $(B_4y_1y_2+B_5y_1y_3+B_6y_2y_3)$ since they disappear in the integration in Lemma 7.14 and conclude that the Hamiltonian density in the case of this field takes the form

$$\mathcal{H}_R = Ce^{-\frac{1}{2}(2\beta)^2 \sum_j y_j^2}(B_1y_1^2 + B_2y_2^2 + B_3y_3^2) \quad C = 16\sqrt{2}\sum_a s_a^2\beta^4 \quad (7.139)$$

Comparing this expression with (7.129) we can write the Hamiltonian operator as

$$\hat{H} = H(y', y) = C\sum_{j=1}^{3} B_j \left(\frac{1}{(2\beta)^4}\frac{\partial}{\partial y_j'}\frac{\partial}{\partial y_j} + \frac{1}{(2\beta)^2}\right)\prod_{j=1}^{3}\delta(y_j - y_j') \quad (7.140)$$

where $y' = (y_1', y_2', y_3')$, $y = (y_1, y_2, y_3)$. Let us mention that the Hamiltonian takes this simple form only for the field (7.101), not for every field. Then

$$< A|\hat{H}|A >= E < A|A > \quad (7.141)$$

takes the form

$$\int d^3y' \int d^3y A(y')H(y', y)A(y) = E\int d^3y' \int d^3y A(y')\delta(y - y')A(y) \quad (7.142)$$

which is the same as

$$\int d^3y\mathcal{H}_R = E\int d^3yA(y)^*A(y) \tag{7.143}$$

I refer to formulae (2.19), (4.31) and (17.50) in Bailin and Love [6] for the connection between the Hamiltonian operator and (7.132) and (7.140). There are of course many approaches but following the approach in [6] the operators (7.132) and (7.140) can be understood to describe the Hamiltonian operator for a field theory.

We see that as β in Theorem 7.17 can be freely selected, either there is no mass gap or vacuum is not unique, provided that the gauge fields A^a_μ in (7.107) and (7.113) are acceptable. We obtained $B = 0$ for the energy of the real part in Minkowski's metric in Lemma 7.15 for the gauge field in (7.107). While we gave another gauge field with positive energy in (7.113), let us notice that the result is negative for the CMI problem also if the constant $C = 0$. If $C = 0$ it implies that the vacuum is not unique and contradicts Wightman's Axiom II that demands that with the exception of vacuum all states have positive energy.

Let us now continue to the question if A^a_μ in (7.107) and (7.113) can be excluded in a non-trivial quantum field theory for the Yang-Mills Lagrangian (7.1).

Quantization of the Yang-Mills theory can be made by two methods; by the path integral method, or by axiomatic quantum field theory. Canonical quantization is also possible but considered difficult for non-Abelian qauge fields. Let us first look at the path integral method. Basically quantization of a Yang-Mills field is made by writing the ground-state-to-ground-state amplitude $W[J]$ as a path integral

$$W[J^\mu_a] \sim \int \mathcal{D}A^\mu exp\left\{-i\hbar^{-1}\int d^4x\left(\mathcal{L}_{YM} + J^\mu_a A^a_\mu\right)\right\} \tag{7.144}$$

However, there are problems in the path integral and the form (7.144) is not to be followed precisely. The path integral may become infinite for a number of reasons and a proper quantization should avoid these pitfalls. The character of such arguments is either mathematical or physical. For instance, the reason why the field should disappear when the space coordinates grow is physical. Mere integrability of a function does not require that it vanishes in infinity as positive and negative parts can cancel.

The discussion in [6] on page 117 mentions the need for fixing the gauge in a case where there are infinitely many A^{μ} related by a gauge transform, and mentions divergences even in the case that the coupling constant $g = 0$ in (7.6).

However, there are more problems in (7.126) when considering A^a_{μ} in (7.107). As can be seen in (7.73), the field A^a_{μ} is a localized wave packet, gauge boson, that moves with the speed of light (x_0) in the (x_1, x_2) plane to the direction $e_1 + e_2$ where e_j is the unit vector of the jth coordinate. This is very natural behavior for a localized wave packet. It cannot stay in a limited box, and it cannot be bounded in the time dimension x_0 because it stays localized and the energy is conserved as A^a_{μ} is a solution to the Euler-Lagrange equations. Thus, the groud-state-to-ground state amplitude

$$W[J] \sim \lim_{t'' \to \infty, t' \to -\infty} < q'', t'' | q', t' >^J \tag{7.145}$$

is not a proper quantity for this field. We can calculate transitions between any finite times t' to t'', and then the path integral is finite.

$$< q'', t'' | q', t' >^J \sim \int \mathcal{D}A^{\mu} exp \left\{ -i\hbar^{-1} \int_{t'}^{t''} dt \int d^3x \left(\mathcal{L}_{YM} + J^{\mu}_a A^a_{\mu} \right) \right\} \tag{7.146}$$

As essential problem is that as $h : \mathbb{C} \to \mathbb{C}$ in (7.69) must be holomorphic so that differentiation can be made, its real and imaginary parts cannot be bounded. We give a physicality argument

It is reasonable to require that the field vanishes when the space coordinates go to $\pm\infty$. However, the time coordinate is different. The future cannot effect the past and therefore possible divergences in the future are not an appropriate boundary condition for a physical problem setting. Likewise, there may well be a finite beginning instance of the time and therefore extension of x_0 to $-\infty$ is highly speculative. Thus, the integration over x_0 in

$W[J]$ is physically motivated only between two finite time instances t' and t''. Accepting that this argument for avoiding infinities in the path integral is as reasonable as other tricks that have been used to the same goal in the semi-mathematical path integral method, such as cutoffs, renormalization, gauge fixing, etc., the gauge field A^a_μ in (7.107) is acceptable in a non-trivial quantum Yang-Mills theory created through the path integral method.

There are other possible mechanisms to render (7.146) finite. In perturbation theory the path integral cannot include solutions to the linear Lagrange equations. Thus, (7.107) could be excluded. As there is no other motivation for exclusion than obtaining a finite integral, one should consider the physicality argument above as a more acceptable way to get a finite (7.146). In any case, there are various ad hoc methods used in the path integral method that have the aim of removing infinities from (7.146).

The gauge fields of the type (7.69) admit a non-trivial quantum field theory for (7.1). We can divide the path integral to two (and later more) parts, where the first part only has fields of the type (7.69). They are easy to handle, sums and real parts also satisfy the Euler-Lagrange equations, as is shown in Lemmas 7.6 and 7.7. Sums of these type of fields with different s_a yield equations that involve the structure coefficients f_{abc} and may have some special solutions. We can briefly look at a sum of solutions of the type (3.30) with different s_a.

Lemma 7.17 Let the gauge field

$$A^\mu_a = (s_{a,1}d_{\mu,1} + s_{a,2}d_{\mu,2})e^{\sum_j h(r_j)} \quad A^3_a = 0 \tag{7.147}$$

be a solution to (7.22). Then h satisfies an equation of the type

$$\sum_j C_{akj}h''(r_j) + \sum_{j,m} D_{akjm}h'(r_j)h'(r_m)$$

$$+\sum_j E_{akj}h'(r_j)e^{\sum_j h(r_j)} + F_{ak}e^{\sum 2h(r_j)} \tag{7.148}$$

where $C_{akj}, D_{akjm}, E_{akj}, F_{ak}$ are constants.

Proof. Calculating $F^a_{\mu\nu}$ yields

$$F^a_{\mu\nu} = \sum_j \alpha_{\mu j} h'(r_j)(s_{a,1}d_{\nu,1} + s_{a,2}d_{\nu,2})e^{\sum_j h(r_j)}$$

$$-\sum_j \alpha_{\nu j} h'(r_j)(s_{a,1}d_{\mu,1} + s_{a,2}d_{\mu,2})e^{\sum_j h(r_j)}$$

$$-g\sum_{c>b} f_{abc}(s_{b,1}s_{c,2} - s_{b,2}s_{c,1})(d_{\mu,1}d_{\nu,2} - (d_{\mu,2}d_{\nu,1})e^{\sum_j 2h(r_j)}$$

The last term does not disappear, thus $F^a_{l,k}$ is of the form

$$F^a_{lk} = \sum_j a_{lkj} h'(r_j)e^{\sum h(r_j)} + b_{lk}e^{\sum_j 2h(r_j)}$$

As $A^3_a = 0$ the equations (7.22) reduce to (7.48). The terms in (7.48) are of the following form, the constants are complex numbers

$$\partial^l F^a_{lk} = \sum_{j,l} a_{lkj}\alpha_{lj} h''(r_j)e^{\sum_j h(r_j)} + \sum_{j,l,m} a_{lkj}\alpha_{lm} h'(r_j)h'(r_m)e^{\sum_j h(r_j)}$$

$$+\sum_{j,l} b_{lk} 2h(r_j)\alpha_{lj} e^{\sum_j 2h(r_j)}$$

$$\partial^3\partial^3 A^a_k = \sum_j c_{akj} h''(r_j)e^{\sum_j h(r_j)} + \sum_j d_{akjm} h'(r_j)h'(r_m)e^{\sum_j h(r_j)}$$

$$-gf_{abc}A^l_a F^a_{lk} = \sum_j e_{lkj} h'(r_j)e^{\sum_j 2h(r_j)} + f_{lk}e^{\sum_j 3h(r_j)}$$

EOP

Lemma 7.18 Let $h(r_j) = \beta^2 r_j^2$ in Lemma 7.17. There are no solutions with $F_{ak} \neq 0$ of the type in Lemma 7.17.

Proof. The term $e^{-2\beta^2 \sum r_j^2}$ in (7.148) is not cancelled by anything. EOP

Lemma 7.18 shows that there are no interactions for solutions of the type (7.88) but there could be solutions of as in Lemma 7.17 for some other

$h(r_j)$. In [4] a type of function is proposed as a solution for (7.48) but it is not explicitly shown that such a solution exists. Even simple solutions of the type (7.69) are not trivial and may give solutions that do not appear in the free field case, i.e., when the coupling constant g is zero.

In any case, the largest group of solutions is surely (7.69) since there we have a free function h, while if the structure constants appear in the equations, we get a nonlinear partial differential equation, at least as difficult or worse as in Lemma 7.17, which typically have fewer solutions. If more solution families are found, correction terms can be calculated from the remaining parts of the path integral. Thus, we can make a non-trivial theory for a pure Yang-Mills Lagrangian and compute first order approximations. Let us mention that Quantum Electrodynamics (QED) and Quantum Chromodynamics (QCD) are not non-trivial quantum field theories for the pure Yang-Mills Langangian but there the spinor fields interacting with gauge fields create the interesting results. As a conclusion, there is no good reason to exlude A^a_μ in (7.107) or (7.113) in the path integral approach.

The other approach is axiomatic quantum field theory where Wightman's axioms, or something as strong, has especially been mentioned in the CMI problem. We do not need to construct a theory filling axioms similar or stronger that Wightman's but only to investigate if a theory filling such conditions should include A^a_μ in (7.107), properly normalized, as a state. This involves showing two things. Firstly, that they can be included, and secondly that a theory that does not include them should be called trivial. Let us briefly go through Wightman's axioms. Wightman does not consider gauge fields at all and so we have to modify the axioms.

Axiom I. *The states of a quantum field theory are normalised vectors in a separable Hilbers space, $\mathcal{H}$, two such that they differ by a complex phase giving raise to the same state.* If we normalize A^a_μ it is a normalized vector in a separable Hilbert space. Any states that differ by a complex phase give rise to the same state. Thus, A^a_μ and $A^{a,R}_\mu$ give the same state as these differ by a complex phase. Apparently we can compute the real Lagrangian $\mathcal{L}$ as this is what Axiom I seems to imply. Fortunately, $A^{a,R}_\mu$ is a solution to the real Euler-Lagrange equations.

Axiom II. *The space $\mathcal{H}$ carries a continuous unitary representation $(a,\Lambda)\mapsto U(a,\Lambda)$ of the restricted orthochronous Poincare group. In $\mathcal{H}$ there exists a vector, unique up to a phase, (called the vacuum state) that is invariant under all $U(a,\Lambda)$ and for all other vectors $\Psi \in \mathcal{H}$ the energy is positive.* The only issue of concern here is that the energy of A^a_μ, and of $A^{a,R}_\mu$, is positive, which is shown in Theorem 7.17 for the metric of (7.33). For Minkowski's metric it was shown the energy of the real Hamiltonian is zero. If also the imaginary part is zero this means that the vacuum is not unique. We may want to discard (7.107) in Minkowski's metric but (7.113) gives positive energy and there is no reason to discard that field.

Axiom IIIa. Deals only with the vacuum state and is of no concern to A^a_μ being an acceptable state or not.

Axiom IIIb. *For any pair of vectors Φ and Ψ, the map*

$$f \mapsto < \Phi, \phi(f)\Psi >$$

is continuous. Here $\phi(f)=\int d^4x\phi(x)f(x)$ is the smeared field. The function f is tempered, i.e., belongs to S, the set of infinitely differentiable functions on $\mathbb{R}^4$ which vanish faster than any power of Euclidean distance. It guarantees that the integral converges. The inner product is given by an integral over $\mathbb{R}^4$. If one of the vectors Φ and Ψ is A^a_μ and another one is not, then fulfilment of the axiom depends on the other vector. If both vectors are of the type A^a_μ then the map $f \mapsto < \Phi, \Psi(f) >$ is continuous.

Axiom IV. *Suppose that $f,g \in S$ are such that* supp$f$ *is space-like to* supp$g$*; then $\phi(f)\phi(g) = \phi(g)\phi(f)$.* This holds for the free field. $\phi = A^a_\mu$ is a solution to free field equations as the part with the structure constants cancels.

These axioms do not have requirements that exclude A^a_μ. Thus, A^a_μ can be included in a theory filling the axioms at if the energy is positive. If the energy is zero, there is a problem in the theory. In that case we may exclude A^a_μ in order to resolve the problem and fill Wightman's axioms, but it is a bit artificial way. The second part is to show that they must be included in a non-trivial theory. The solutions A^a_μ are natural solutions to the Euler-Langange equations and especially if the coupling constant $g = 0$ or the group is $U(1)$ they are among the possible solutions. They give arbitrarily small eigenvalues to the Hamiltonian. While it may be possible to create a theory which does not include these solutions and is still valid for $U(1)$ and $g = 0$ cases, such a theory is trivial since it can be made by the following trivial procedure. Take any theory filling the axioms. If it includes the states A^a_μ, then exlude all states that have these states as minimal solutions for the Lagrangian. The resulting theory does not have these eigenstates for the Hamiltonian. Indeed, we can make a theory with two states only, vacuum and an eigenstate of the Hamiltonian with a non-zero eigenvalue. Then all axioms are easily filled. A trick of this type can always be made and it avoids the essential problem of showing that there is a mass gap and has no physical relevance. Thus, we should call trivial any quantum field theory that does not include the solutions of the type A^a_μ if (more accurately, as) they can be included.

There are two manuscripts [4],[5] arguing that a mass gap exists. Both start by imposing the temporal, or Weyl, gauge $A^0 = 0$. If we impose this gauge and then look at the boundary conditions, the solutions (7.107) cannot be found. This is because when we localize the gauge field we need three linearly independent vectors r_j in (7.72). As can be seen in the selected space gauge $A^3 = 0$, we only get two vectors for the non-gauged coordinates as is shown in Lemma 7.9. The third vector must be obtained from the gauged coordinate. Had we gauged time, then the equations in (7.73) would show that x_0 is limited, as now is x_3, while x_1 and x_2 would be linearly dependent on x_3. Then the field would not be integrable over the space coordinates, while it would be limited in time. Instead of fixing the gauge first, we must first look at the boundary conditions. This shows that the temporal gauge is not the correct choice, while a space gauge can work.

Final comments of the CMI Millennium Prize problem

The CMI problem setting called for mathematical clarity to the area of gauge fields. Much of this lack of clarity has traditionally been caused by mathematical unclarities in the path integral method. Everything is formulated in simple lemmas which are given proofs. This does not imply that the lemmas are considered new, it is only for clarity. The presentation of Yang-Mills fields follows the approach in [6].

There are some final words about the clarity of the CMI problem statement itself.

The problem statement does not specify whether the gauge group should be local or global, and [5] understands that it is global gauge group. It probably must be local gauge group since the relevant issues arise from local gauge invariance. However, this should have been stated.

The metric in the CMI problem setting is unclear. Minkowski's metric (7.9) is the correct choice for quantum field theory but the CMI problem setting only mentions $\mathbb{R}^4$ and the expert's explanation in [2], referred to in (7.77), seems to point to the Euclidean metric and to real curvature. We can present the results in $\mathbb{R}^4$ with the Euclidean metric also. The convenient way to do it is to use the negative definite metric (7.33). In Section 7.3 we have used (7.11). It is still valid for the metric (7.33), as (7.5) is the definition and in (7.11) we have simply multiplied (7.6) by $g_{\mu\beta}g_{\nu\beta}$. We have also used (7.22). In the derivation of (7.22) we have kept the metric explicitely and not used the values of $g_{\mu\nu}$ from (7.9). Thus, (7.22) is also valid for (7.33). There are no raising or lowering x_0 indices in Lemmas 7.4-7.12, thus they stay valid. In Lemma 7.13 we use (7.27) but give the result also for the metric in (7.33). Lemma 7.14 has no changes. There is derivation with respect to x_0 in Lemma 7.15 but the conclusions remain since they are caused by the disappearance of the integral (7.100) as is mentioned in to proof. It follows that Theorem 7.17 holds also for the metric (7.33) for some other constant C. It is assumed that the fields can be complex as it is the situation in the physical problem and the Hodge star operation is defined for differential forms in complex manifolds. But as it is unclear in (7.38) and in the problem setting the calculations were done for the real part of the curvature covering the possibility that the problem statement implies real fields. It would have been much clearer if the CMI problem statement had stated if Minkowski's metric is assumed, and if the fields are complex or real.

Referring to axiomatic field theory by mentioning axioms that do not as such apply to gauge fields, use of words such as non-trivial, etc. would make any positive solutions to the CMI problem difficult to argue. This would not be an issue if proposed solutions to the CMI problems would be positively received and carefully reviewed. It would be an issue if the opposite were the case.

The results of this article are easier to verify:

It seems that the CMI problem refers to Euclidean metric and real curvature. As there is no minus sign in (7.34) and (7.38) while there is one in (7.1) it seems that the metric is as in (7.33). In this case there is no mass gap since we can by selection of β in (7.112) make the eigenvalue of the Hamiltonian as small as desired.

If the problem means Minkowski's metric and real fields, then the gauge field in (7.113) shows that there is no mass gap. However, the field (7.107) gives zero energy and indicates that vacuum is not unique and Wightman's axioms cannot be filled. We may want to exclude the field (7.107) in this case but there is no good reason for excluding it.

If the problem means complex fields in either metric, the conclusions are the same.

The results presented here should not be called a trivial free field theory. The coupling constant g is not set to zero. The solutions that have been found are of such a type that the part with structure constants cancel. As Lemmas 7.17 and 7.18 indicate, nontrivial results can be found starting from the solutions in (7.69) and (7.78). Localization of the field in space is not trivial and in general this word should be avoided if clarity is desired because clarity is best achieved by writing down all steps. This article may be correctly called elementary and easy, but not trivial.

References to Chapter 7:

[1] A. Jaffe and E. Witten: Quantum Yang-Mills Theory. available online at www.claymath.org.

[2] L. D. Faddeev: Mass in Quantum Yang-Mills Theory (comment on a Clay Millennium Problem). arxiv:0911.1013v1 5. Nov 2009.

[3] M. R. Douglas: Report on the Status of the Yang-Mills Millennium Proze Problem, April 2004. available on-line at www.claymath.org.

[4] M. Frasca: Mass gap in a Yang-Mills theory in the strong coupling limit. arxiv:0511173v6 26. Jan 2007.

[5] A. Dynin: Energy-mass spectrum of Yang-Mills bosons is infinite and discrete. arxiv:0903.4727v2 20. May 2009.

[6] D. Bailin and A. Love: *Introduction to Gauge Field Theory. Adam Hilger, Bristol and Boston, IOP Publishing Limited 1986.*

8. Annex: Ricci numbers in Schwarzschild's solution

This is just an example for people who do not know or or not remember how to calculate the Ricci tensor and the Ricci scalar. The example is Schwarzschild's solution to the Einstein equations.

Schwatzschield's solution was the first exact solution of Einstein's equations. Even the smallest steps on the derivation of the Ricci tensor R_{ab} entires for Swarzschild's solution are given. No knowledge of physics is needed in following the text.

Schwarzschild's metric is defined by a line element ds^2 of the form

$$ds^2 = B(r)dt^2 + A(r)dr^2 - r^2 d\theta^2 - r^2 \sin^2(\theta) d\phi^2$$

where in the actual solution we get

$$A(r) = -\left(1 - \frac{2GM}{rc^2}\right)^{-1} \quad , \quad B(r) = 1 - \frac{2GM}{rc^2}$$

but in the calculation we only need to know that they are functions of r alone. I will for simplicity set the speed of the light c to 1, thus, the coordinates are polar $(x^0, x^1, x^2, x^3) = (t, r, \theta, \phi)$.

The metric can be written as

$$ds^2 = g_{ab} dx^a dx^b$$

and the metric tensor g_{ab} is:

$$g_{00} = B(r) \quad , \quad g_{11} = -A(r) \quad , \quad g_{2,2} = -r^2 \quad , \quad g_{33} = -r^2 \sin^2\theta$$

$$g_{ab} = 0 \quad \text{if} \quad a \neq b$$

First we have to calculate the Christoffel symbols Γ^a_{bc} for this metric. By definition

$$\Gamma^a_{bc} = \frac{1}{2} g^{ad} \left(g_{bd,c} + g_{dc,b} - g_{bc,d}\right).$$

The terms g^{ab} are

$$g^{00} = B(r)^{-1} \quad , \quad g^{11} = -A(r)^{-1} \quad , \quad g^{2,2} = -r^{-2}$$

$$g^{33} = -r^{-2}\sin^{-2}\theta \quad , \quad g^{ab} = 0 \quad \text{if} \quad a \neq b$$

i.e., if the tensor g_{ab} is understood as a matrix, g^{ab} is the inverse matrix.

In the definition of Γ^a_{bc} there is a summation over d. This is the Einstein summation convention used throughout these formulas: there is a summation over each index that appears in one term both as an upper and a lower index. If summation is not intended, it is mentioned unless obvious from the context. As $g^{ab} = 0$ if $a \neq b$, the summation over d reduces to one term, $d = a$, thus

$$\Gamma^a_{bc} = \frac{1}{2} g^{aa} (g_{ba,c} + g_{ac,b} - g_{ac,d})$$

with no summation convention.

There is also a comma as in $g_{ba,c}$. This comma is a shorthand for a partial derivate: for any $f(x^0, x^1, x^2, x^3)$ and a base $\{x^0, x^1, x^2, x^3\}$ the notation means $f_{,a} = \frac{\partial f}{\partial x^a}$, thus $g_{ba,c} = \frac{\partial}{\partial x^a} g_{ba}$.

Because $g_{ba} = 0$ if $b \neq a$, most of the terms Γ^a_{bc} vanish. There remains only

$$\Gamma^a_{aa} = \frac{1}{2} g^{aa} g_{aa,a}$$

$$\Gamma^a_{ba} = \frac{1}{2} g^{aa} g_{aa,b}$$

and

$$\Gamma^a_{bb} = \frac{1}{2} g^{aa} g_{bb,a}.$$

With these g_{ab} holds $\Gamma^a_{bc} = \Gamma^a_{cb}$. This equality holds for all spaces that do not have torque and Riemann manifolds do not have torque, but we can verify it here directly without knowing the theory. It follows that we only need to calculate Γ^a_{bc} where $c \geq b$ in those three cases. There are four cases of Γ^a_{aa}:

$$\Gamma^0_{00} = \frac{1}{2} B^{-1} \frac{\partial f}{\partial t} B = 0$$

where $B' = B'(r)$ *is the derivate of* $B(r)$.

$$\Gamma^1_{11} = \frac{1}{2}(-A^{-1})\frac{\partial f}{\partial r}(-A) = \frac{1}{2}A'A^{-1}$$

$$\Gamma^2_{22} = \frac{1}{2}(-r^{-2})\frac{\partial f}{\partial \theta}(-r^2) = 0$$

$$\Gamma^3_{33} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta)\frac{\partial f}{\partial \phi}(-r^2\sin\theta) = 0.$$

There are twelve cases of Γ^a_{ba}:

$$\Gamma^0_{01} = \frac{1}{2}B^{-1}\frac{\partial f}{\partial r}B = \frac{1}{2}B'B^{-1}$$

$$\Gamma^0_{02} = \frac{1}{2}B^{-1}\frac{\partial f}{\partial \theta}B = 0$$

$$\Gamma^0_{03} = \frac{1}{2}B^{-1}\frac{\partial f}{\partial \phi}B = 0$$

$$\Gamma^1_{01} = \frac{1}{2}(-A^{-1})\frac{\partial f}{\partial t}(-A) = 0$$

$$\Gamma^1_{12} = \frac{1}{2}(-A^{-1})\frac{\partial f}{\partial \theta}(-A) = 0$$

$$\Gamma^1_{13} = \frac{1}{2}(-A^{-1})\frac{\partial f}{\partial \phi}(-A) = 0$$

$$\Gamma^2_{20} = \frac{1}{2}(-r^{-2})\frac{\partial f}{\partial t}(-r^2) = 0$$

$$\Gamma^2_{21} = \frac{1}{2}(-r^{-2})\frac{\partial f}{\partial r}(-r^2) = \frac{1}{r}$$

$$\Gamma^2_{23} = \frac{1}{2}(-r^{-2})\frac{\partial f}{\partial \phi}(-r^2) = 0$$

$$\Gamma^3_{30} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta)\frac{\partial f}{\partial t}(-r^2\sin^2\theta) = 0$$

$$\Gamma^3_{31} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta)\frac{\partial f}{\partial r}(-r^2)\sin^2\theta) = \frac{1}{r}$$

$$\Gamma^3_{32} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta)\frac{\partial f}{\partial \phi}(-r^2)\sin^2\theta) = \cot\theta$$

and additionally there are twelve cases by the symmetry $\Gamma^a_{bc} = \Gamma^a_{cb}$.

There are twelve cases of Γ^a_{bb}*:*

$$\Gamma^0_{11} = \frac{1}{2} B^{-1} \frac{\partial f}{\partial t}(-A) = 0$$

$$\Gamma^0_{22} = \frac{1}{2} B^{-1} \frac{\partial f}{\partial t}(-r^2) = 0$$

$$\Gamma^0_{22} = \frac{1}{2} B^{-1} \frac{\partial f}{\partial t}(-r^2 \sin^2\theta) = 0$$

$$\Gamma^1_{00} = \frac{1}{2}(-A^{-1}) \frac{\partial f}{\partial r} B = \frac{1}{2} B' A^{-1}$$

$$\Gamma^1_{22} = \frac{1}{2}(-A^{-1}) \frac{\partial f}{\partial r}(-r^2) = -r A^{-1}$$

$$\Gamma^1_{33} = \frac{1}{2}(-A^{-1}) \frac{\partial f}{\partial r}(-r^2 \sin^2\theta) = -\frac{1}{2} r \sin^2\theta A^{-1}$$

$$\Gamma^2_{00} = \frac{1}{2}(-r^{-2}) \frac{\partial f}{\partial \theta} B = 0$$

$$\Gamma^2_{11} = \frac{1}{2}(-r^{-2}) \frac{\partial f}{\partial \theta}(-A) = 0$$

$$\Gamma^2_{33} = \frac{1}{2}(-r^{-2}) \frac{\partial f}{\partial \theta}(-r^2 \sin^2\theta) = -\sin\theta\cos\theta$$

$$\Gamma^2_{00} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta) \frac{\partial f}{\partial \phi} B = 0$$

$$\Gamma^2_{11} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta) \frac{\partial f}{\partial \phi}(-A) = 0$$

$$\Gamma^2_{22} = \frac{1}{2}(-r^{-2}\sin^{-2}\theta) \frac{\partial f}{\partial \phi}(-r^2) = 0$$

and additionally there are twelve cases by the symmetry $\Gamma^a_{bc} = \Gamma^a_{cb}$.

The nonzero Cristoffel symbols are

$$\Gamma^0_{01} = \Gamma^0_{10} = \frac{1}{2} B' B^{-1}$$

$$\Gamma^1_{00} = \frac{1}{2} B' A^{-1}$$

$$\Gamma^1_{11} = \frac{1}{2}A'A^{-1}$$

$$\Gamma^1_{22} = -rA^{-1}$$

$$\Gamma^1_{33} = -r\sin^2\theta A^{-1}$$

$$\Gamma^2_{33} = -\sin\theta\cos\theta$$

$$\Gamma^2_{12} = \Gamma^2_{21} = \frac{1}{r}$$

$$\Gamma^3_{13} = \Gamma^3_{31} = \frac{1}{r}$$

$$\Gamma^3_{23} = \Gamma^3_{32} = \cot\theta$$

Occasionally you see

$$\Gamma^1_{00} = -\frac{1}{2}B'A^{-1}$$

This can come only if $g_{00} = -B(r)$. *This is not the case in Schwarzschild's solution where* $B(r) = 1 - \frac{2GM}{rc^2}$. *The change of the sign of* g_{00} *affects only* Γ^1_{00} *and changes the sign of* R_{00}. *No other* R_{ab} *is affected.*

Next we calculate components of the Ricci curvature tensor R_{bd}. *The Ricci curvature tensor is defined as* $R_{bd} = R^a_{bad}$ *where*

$$R^a_{bcd} = \Gamma^a_{bd,c} - \Gamma^a_{bc,d} + \Gamma^e_{bd}\Gamma^a_{ec} - \Gamma^e_{bc}\Gamma^a_{ed}$$

is the Riemann curvature tensor. The Ricci scalar is then $R = g^{ab}R_{ab}$ *and Einstein's field equations are*

$$R_{ab} - \frac{1}{2}R = k_0T_{ab} + \lambda g_{ab}$$

where $k = 0 = 8\pi G/c^4$, T_{ab} *is the stress tensor and* λ *is the cosmological constant, often assumed zero.*

In order to calculate R_{bd} *we only need terms* R^a_{bcd} *where* $c = a$. *There is summation over* e *and* a. *Let us calculate the term*

$$\Gamma^e_{bd}\Gamma^a_{ec} - \Gamma^e_{bc}\Gamma^a_{ed}$$

for different b *and* d. *Expanding the summations gives*

$$\Gamma^0_{bd}(\Gamma^0_{00}+\Gamma^1_{01}+\Gamma^2_{02}+\Gamma^3_{03})$$
$$-(\Gamma^0_{b0}\Gamma^0_{0d}+\Gamma^0_{b1}\Gamma^1_{0d}+\Gamma^0_{b2}\Gamma^2_{0d}+\Gamma^0_{b3}\Gamma^3_{0d})$$

$$\Gamma^1_{bd}(\Gamma^0_{10}+\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13})$$
$$-(\Gamma^1_{b0}\Gamma^0_{1d}+\Gamma^1_{b1}\Gamma^1_{1d}+\Gamma^1_{b2}\Gamma^2_{1d}+\Gamma^1_{b3}\Gamma^3_{1d})$$

$$\Gamma^2_{bd}(\Gamma^0_{20}+\Gamma^1_{21}+\Gamma^2_{22}+\Gamma^3_{23})$$
$$-(\Gamma^2_{b0}\Gamma^0_{2d}+\Gamma^2_{b1}\Gamma^1_{2d}+\Gamma^2_{b2}\Gamma^2_{2d}+\Gamma^2_{b3}\Gamma^3_{2d})$$

$$\Gamma^3_{bd}(\Gamma^0_{30}+\Gamma^1_{31}+\Gamma^2_{32}+\Gamma^3_{33})$$
$$-(\Gamma^3_{b0}\Gamma^0_{3d}+\Gamma^3_{b1}\Gamma^1_{3d}+\Gamma^3_{b2}\Gamma^2_{3d}+\Gamma^3_{b3}\Gamma^3_{3d})$$

The Ricci curvature tensor is symmetric in the lower indices $R_{ab}=R_{ba}$, *thus we need to calculate ten terms. I mark the nonzero Cristoffel symbols in bold unless they are multiplied by zero. Look for the list of thirteen nonzero symbols:* $\Gamma^0_{01}=\Gamma^0_{10}$, Γ^1_{00}, Γ^1_{11}, Γ^1_{22}, Γ^1_{33}, Γ^2_{33}, $\Gamma^2_{12}=\Gamma^2_{21}$, $\Gamma^3_{13}=\Gamma^3_{31}$, $\Gamma^3_{23}=\Gamma^3_{32}$.

For $b=d=0$ *we get*

$$\Gamma^0_{00}(\Gamma^0_{00}+\Gamma^1_{01}+\Gamma^2_{02}+\Gamma^3_{03})$$
$$-(\Gamma^0_{00}\Gamma^0_{00}+\mathbf{\Gamma}^0_{01}\mathbf{\Gamma}^1_{00}+\Gamma^0_{02}\Gamma^2_{00}+\Gamma^0_{03}\Gamma^3_{00})$$

$$\mathbf{\Gamma}^1_{00}(\mathbf{\Gamma}^0_{10}+\mathbf{\Gamma}^1_{11}+\mathbf{\Gamma}^2_{12}+\mathbf{\Gamma}^3_{13})$$
$$-(\mathbf{\Gamma}^1_{00}\mathbf{\Gamma}^0_{10}+\Gamma^1_{01}\Gamma^1_{10}+\Gamma^1_{02}\Gamma^2_{10}+\Gamma^1_{03}\Gamma^3_{10})$$

$$\Gamma^2_{00}(\Gamma^0_{20}+\Gamma^1_{21}+\Gamma^2_{22}+\Gamma^3_{23})$$

$$-(\Gamma^2_{00}\Gamma^0_{20}+\Gamma^2_{01}\Gamma^1_{20}+\Gamma^2_{02}\Gamma^2_{20}+\Gamma^2_{03}\Gamma^3_{20})$$

$$\Gamma^3_{00}(\Gamma^0_{30}+\Gamma^1_{31}+\Gamma^2_{32}+\Gamma^3_{33})$$

$$-(\Gamma^3_{00}\Gamma^0_{30}+\Gamma^3_{01}\Gamma^1_{30}+\Gamma^3_{02}\Gamma^2_{30}+\Gamma^3_{03}\Gamma^3_{30}).$$

The term simplifies to

$$=\Gamma^1_{00}\left(\Gamma^0_{01}+\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13}-\Gamma^0_{01}-\Gamma^0_{01}\right)$$

$$=\Gamma^1_{00}\left(\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13}-\Gamma^0_{01}\right)$$

$$=\frac{1}{2}B'A^{-1}\left(\frac{1}{2}A'A^{-1}+\frac{1}{r}+\frac{1}{r}-\frac{1}{2}B'B^{-1}\right)$$

$$=\frac{1}{4}A'B'A^{-2}-\frac{1}{4}(B')^2A^{-1}B^{-1}+\frac{1}{r}B'A^{-1}$$

The part

$$\Gamma^a_{bd,c}-\Gamma^a_{bc,d}$$

with $c=a$ *expands as*

$$\Gamma^0_{bd,0}+\Gamma^1_{bd,1}+\Gamma^2_{bd,2}+\Gamma^3_{bd,3}$$

$$-\Gamma^0_{b0,d}-\Gamma^1_{b1,d}-\Gamma^2_{b2,d}-\Gamma^3_{b3,d}$$

Thus for $b=d=0$ *we get*

$$\Gamma^0_{00,0}+\boldsymbol{\Gamma}^1_{00,1}+\Gamma^2_{00,2}+\Gamma^3_{00,3}$$

$$-\Gamma^0_{00,0}-\Gamma^1_{01,0}-\Gamma^2_{02,0}-\Gamma^3_{03,0}$$

where boldface indicates nonzero and noncancelled terms. We get

$$\frac{\partial}{\partial r}\left(\frac{1}{2}B'A^{-1}\right)=-\frac{1}{2}A'B'A^{-2}+\frac{1}{2}B''A^{-2}$$

The Ricci element R_{00} *is the sum of these two terms*

$$R_{00} = -\frac{1}{2}A'B'A^{-2} + \frac{1}{2}B''A^{-2}$$

$$+\frac{1}{4}A'B'A^{-2} - \frac{1}{4}(B')^2A^{-1}B^{-1} + \frac{1}{r}B'A^{-1}$$

$$R_{00} = \frac{1}{2}B''A^{-2} - \frac{1}{4}A'B'A^{-2} - \frac{1}{4}(B')^2A^{-1}B^{-1} + \frac{1}{r}B'A^{-1}$$

Notice that if the sign of Γ^1_{00} *is changed, the sign of* R_{00} *is changed.*

We continue to $b = d = 1$. *Again, look for the list of thirteen nonzero symbols:* $\Gamma^0_{01} = \Gamma^0_{10}$, Γ^1_{00}, Γ^1_{11}, Γ^1_{22}, Γ^1_{33}, Γ^2_{33}, $\Gamma^2_{12} = \Gamma^2_{21}$, $\Gamma^3_{13} = \Gamma^3_{31}$, $\Gamma^3_{23} = \Gamma^3_{32}$.

The first term with nonzero elements marked in boldface (unless multiplied by zero) is

$$\Gamma^0_{11}(\Gamma^0_{00} + \Gamma^1_{01} + \Gamma^2_{02} + \Gamma^3_{03})$$

$$-(\mathbf{\Gamma}^0_{10}\mathbf{\Gamma}^0_{01} + \Gamma^0_{11}\Gamma^1_{01} + \Gamma^0_{12}\Gamma^2_{01} + \Gamma^0_{13}\Gamma^3_{01})$$

$$\mathbf{\Gamma}^1_{11}(\mathbf{\Gamma}^0_{10} + \mathbf{\Gamma}^1_{11} + \mathbf{\Gamma}^2_{12} + \mathbf{\Gamma}^3_{13})$$

$$-(\Gamma^1_{10}\Gamma^0_{11} + \mathbf{\Gamma}^1_{11}\mathbf{\Gamma}^1_{11} + \Gamma^1_{12}\Gamma^2_{11} + \Gamma^1_{13}\Gamma^3_{11})$$

$$\Gamma^2_{11}(\Gamma^0_{20} + \Gamma^1_{21} + \Gamma^2_{22} + \Gamma^3_{23})$$

$$-(\Gamma^2_{10}\Gamma^0_{21} + \Gamma^2_{11}\Gamma^1_{21} + \mathbf{\Gamma}^2_{12}\mathbf{\Gamma}^2_{21} + \Gamma^2_{13}\Gamma^3_{21})$$

$$\Gamma^3_{11}(\Gamma^0_{30} + \Gamma^1_{31} + \Gamma^2_{32} + \Gamma^3_{33})$$

$$-(\Gamma^3_{10}\Gamma^0_{31} + \Gamma^3_{11}\Gamma^1_{31} + \Gamma^3_{12}\Gamma^2_{31} + \mathbf{\Gamma}^3_{13}\mathbf{\Gamma}^3_{31})$$

We get

$$= -(\Gamma^0_{01})^2 + \Gamma^1_{11}\left(\Gamma^0_{01} + \Gamma^1_{11} + \Gamma^2_{12} + \Gamma^3_{13}\right)$$

$$-(\Gamma^1_{11})^2 - (\Gamma^2_{21})^2 - (\Gamma^3_{31})^2$$

$$= \Gamma^1_{11}\left(\Gamma^0_{01} + \Gamma^2_{12} + \Gamma^3_{13}\right) - (\Gamma^0_{01})^2 - (\Gamma^2_{21})^2 - (\Gamma^3_{31})^2$$

$$= \frac{1}{2}A'A^{-1}(\frac{1}{2}B'B^{-1} + \frac{1}{r} + \frac{1}{r}) - \frac{1}{4}(B')^2B^{-2} - \frac{1}{r^2} - \frac{1}{r^2}$$

The part

$$\Gamma^0_{11,0} + \Gamma^1_{11,1} + \Gamma^2_{11,2} + \Gamma^3_{11,3}$$

$$-\mathbf{\Gamma}^0_{10,1} - \Gamma^1_{11,1} - \mathbf{\Gamma}^2_{12,1} - \mathbf{\Gamma}^3_{13,1}$$

gives

$$= -\frac{\partial}{\partial r}(\frac{1}{2}B'B^{-1} + \frac{1}{r} + \frac{1}{r})$$

$$= \frac{1}{2}(B')^2B^{-2} - B''B^{-1} + \frac{2}{r^2}$$

The Ricci term R_{11} *is thus*

$$R_{11} = \frac{1}{2}(B')^2B^{-2} - \frac{1}{2}B''B^{-1} + \frac{2}{r^2}$$

$$+\frac{1}{2}A'A^{-1}(\frac{1}{2}B'B^{-1} + \frac{1}{r} + \frac{1}{r}) - \frac{1}{4}(B')^2B^{-2} - \frac{1}{r^2} - \frac{1}{r^2}$$

$$R_{11} = -\frac{1}{2}B''B^{-1} + \frac{1}{2}(B')^2B^{-2} + \frac{1}{4}A'B'A^{-1}B^{-1} + \frac{1}{r}A'A^{-1}$$

The next Ricci tensor element is for $b = d = 2$. *As always, look for the thirteen nonzero Christoffel symbols:* $\Gamma^0_{01} = \Gamma^0_{10}$, Γ^1_{00}, Γ^1_{11}, Γ^1_{22}, Γ^1_{33}, Γ^2_{33}, $\Gamma^2_{12} = \Gamma^2_{21}$, $\Gamma^3_{13} = \Gamma^3_{31}$, $\Gamma^3_{23} = \Gamma^3_{32}$.

$$\Gamma^0_{22}(\Gamma^0_{00} + \Gamma^1_{01} + \Gamma^2_{02} + \Gamma^3_{03})$$

$$-(\Gamma^0_{20}\Gamma^0_{02} + \Gamma^0_{21}\Gamma^1_{02} + \Gamma^0_{22}\Gamma^2_{02} + \Gamma^0_{23}\Gamma^3_{02})$$

$$\mathbf{\Gamma}^1_{22}(bf\Gamma^0_{10} + bf\Gamma^1_{11} + \mathbf{\Gamma}^2_{12} + \mathbf{\Gamma}^3_{13})$$

$$-(\Gamma^1_{20}\Gamma^0_{12} + \Gamma^1_{21}\Gamma^1_{12} + \boldsymbol{\Gamma}^1_{22}\boldsymbol{\Gamma}^2_{12} + \Gamma^1_{23}\Gamma^3_{12})$$

$$\Gamma^2_{22}(\Gamma^0_{20} + \Gamma^1_{21} + \Gamma^2_{22} + \Gamma^3_{23})$$

$$-(\Gamma^2_{20}\Gamma^0_{22} + \boldsymbol{\Gamma}^2_{21}\boldsymbol{\Gamma}^1_{22} + \Gamma^2_{22}\Gamma^2_{22} + \Gamma^2_{23}\Gamma^3_{22})$$

$$\Gamma^3_{22}(\Gamma^0_{30} + \Gamma^1_{31} + \Gamma^2_{32} + \Gamma^3_{33})$$

$$-(\Gamma^3_{20}\Gamma^0_{32} + \Gamma^3_{21}\Gamma^1_{32} + \Gamma^3_{22}\Gamma^2_{32} + \boldsymbol{\Gamma}^3_{23}\boldsymbol{\Gamma}^3_{32})$$

This yields

$$= \Gamma^1_{22}\left(\Gamma^0_{01} + \Gamma^1_{11} + \Gamma^2_{12} + \Gamma^3_{13}\right)$$

$$-\Gamma^1_{22}\Gamma^2_{12} - \Gamma^1_{22}\Gamma^2_{12} - \Gamma^3_{32}\Gamma^3_{32}$$

$$= \Gamma^1_{22}\left(\Gamma^0_{01} + \Gamma^1_{11} + \Gamma^3_{13} - \Gamma^2_{21}\right)$$

$$-(\Gamma^3_{32})^2$$

$$= -rA^{-1}\left(\frac{1}{2}B'B^{-1} + \frac{1}{2}A'A^{-1} + \frac{1}{r} - \frac{1}{r}\right) - \cot^2\theta$$

$$= -\frac{1}{2}rA'A^{-1} - \frac{1}{2}rB'A^{-1}B^{-1} - \cot^2\theta$$

The other part has two nonzero elements

$$\Gamma^0_{22,0} + \boldsymbol{\Gamma}^1_{22,1} + \Gamma^2_{22,2} + \Gamma^3_{22,3}$$

$$-\Gamma^0_{20,2} - \Gamma^1_{21,2} - \Gamma^2_{22,2} - \boldsymbol{\Gamma}^3_{23,2}$$

and gives

$$= \frac{\partial}{\partial r}(-rA^{-1}) - \frac{\partial}{\partial\theta}\cot\theta$$

$$= -A^{-1} + rA'A^{-2} + \sin^{-2}\theta$$

Thus

$$R_{22} = -A^{-1} + rA'A^{-2} + \frac{1}{\sin^2\theta}$$

$$-\frac{1}{2}rA'A^{-1} - \frac{1}{2}rB'A^{-1}B^{-1} - \frac{\cos^2\theta}{\sin^2\theta} - A^{-1}$$

$$R_{22} = \frac{1}{2}rA'A^{-2} - \frac{1}{2}rB'A^{-1}B^{-1} + 1 - A^{-1}$$

The next term has $b = d = 3$. The first part yields

$$\Gamma^0_{33}(\Gamma^0_{00} + \Gamma^1_{01} + \Gamma^2_{02} + \Gamma^3_{03})$$

$$-(\Gamma^0_{30}\Gamma^0_{03} + \Gamma^0_{31}\Gamma^1_{03} + \Gamma^0_{32}\Gamma^2_{03} + \Gamma^0_{33}\Gamma^3_{03})$$

$$\boldsymbol{\Gamma}^1_{33}(\boldsymbol{\Gamma}^0_{10} + \boldsymbol{\Gamma}^1_{11} + \boldsymbol{\Gamma}^2_{12} + \boldsymbol{\Gamma}^3_{13})$$

$$-(\Gamma^1_{30}\Gamma^0_{13} + \Gamma^1_{31}\Gamma^1_{13} + \Gamma^1_{32}\Gamma^2_{13} + \boldsymbol{\Gamma}^1_{33}\boldsymbol{\Gamma}^3_{13})$$

$$\boldsymbol{\Gamma}^2_{33}(\Gamma^0_{20} + \Gamma^1_{21} + \Gamma^2_{22} + \boldsymbol{\Gamma}^3_{23})$$

$$-(\Gamma^2_{30}\Gamma^0_{23} + \Gamma^2_{31}\Gamma^1_{23} + \Gamma^2_{32}\Gamma^2_{23} + \Gamma^2_{33}\Gamma^3_{23})$$

$$\Gamma^3_{33}(\Gamma^0_{30} + \Gamma^1_{31} + \Gamma^2_{32} + \Gamma^3_{33})$$

$$-(\Gamma^3_{30}\Gamma^0_{33} + \boldsymbol{\Gamma}^3_{31}\boldsymbol{\Gamma}^1_{33} + \boldsymbol{\Gamma}^3_{32}\boldsymbol{\Gamma}^2_{33} + \boldsymbol{\Gamma}^3_{33}\boldsymbol{\Gamma}^3_{33})$$

This time there are nine nonzero terms. The result is

$$= \Gamma^1_{33}\left(\Gamma^0_{01} + \Gamma^1_{11} + \Gamma^2_{12} + \Gamma^3_{13}\right)$$

$$-\Gamma^1_{33}\Gamma^3_{13} + \Gamma^2_{33}\Gamma^3_{23} - \Gamma^2_{33}\Gamma^3_{23} - \Gamma^1_{33}\Gamma^3_{13} - \Gamma^2_{33}\Gamma^3_{23}$$

$$= \Gamma^1_{33}\left(\Gamma^0_{01} + \Gamma^1_{11} + \Gamma^2_{12} - \Gamma^3_{13}\right) - \Gamma^2_{33}\Gamma^3_{23}$$

$$= -A^{-1} r\sin^2\theta\left(\frac{1}{2}B'B^{-1}+\frac{1}{2}A'A^{-1}+\frac{1}{r}-\frac{1}{r}\right)-\cot\theta(-\sin\theta\cos\theta)$$

$$= -\frac{1}{2}r\sin^2\theta\left(A'A^{-2}+B'A^{-1}B^{-1}\right)+\cos^2\theta)$$

From the other term we get two nonzero terms

$$\Gamma^0_{33,0}+\mathbf{\Gamma}^1_{33,1}+\mathbf{\Gamma}^2_{33,2}+\Gamma^3_{33,3}$$

$$-\Gamma^0_{30,3}-\Gamma^1_{31,3}-\Gamma^2_{32,3}-\Gamma^3_{33,3}$$

$$= \frac{\partial}{\partial r}(-A^{-1}r\sin^2\theta)+\frac{\partial}{\partial\theta}(-\sin\theta\cos\theta)$$
$$= A'A^{-2}r\sin^2\theta-\sin^2\theta A^{-1}-\cos^2\theta+\sin^2\theta$$

The Ricci element R_{33} *is thus*

$$R_{33} = A'A^{-2}r\sin^2\theta-\sin^2\theta A^{-1}-\cos^2\theta+\sin^2\theta$$
$$-\frac{1}{2}r\sin^2\theta\left(A'A^{-2}+B'A^{-1}B^{-1}\right)+\cos^2\theta)$$

$$R_{33} = \sin^2\theta\left(rA'A^{-2}-\frac{1}{2}rA'A^{-2}-\frac{1}{2}rB'A^{-1}B^{-1}-A^{-1}\right)+\sin^2\theta$$

$$R_{33} = \sin^2\theta\left(\frac{1}{2}rA'A^{-2}-\frac{1}{2}rB'A^{-1}B^{-1}+1-A^{-1}\right)$$

The next six Ricci tensor terms all vanish but they still have to be checked in a proper calculation. For $b=0$, $d=1$ *all nonzero parts in the following sum are multiplied by zero:*

$$\Gamma^0_{01}(\Gamma^0_{00}+\Gamma^1_{01}+\Gamma^2_{02}+\Gamma^3_{03})$$

$$-(\Gamma^0_{00}\Gamma^0_{01}+\Gamma^0_{01}\Gamma^1_{01}+\Gamma^0_{02}\Gamma^2_{01}+\Gamma^0_{03}\Gamma^3_{01})$$

$$\Gamma^1_{01}(\Gamma^0_{10}+\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13})$$

$$-(\Gamma^1_{00}\Gamma^0_{11}+\Gamma^1_{01}\Gamma^1_{11}+\Gamma^1_{02}\Gamma^2_{11}+\Gamma^1_{03}\Gamma^3_{11})$$

$$\Gamma^2_{01}(\Gamma^0_{20}+\Gamma^1_{21}+\Gamma^2_{22}+\Gamma^3_{23})$$

$$-(\Gamma^2_{00}\Gamma^0_{21}+\Gamma^2_{01}\Gamma^1_{21}+\Gamma^2_{02}\Gamma^2_{21}+\Gamma^2_{03}\Gamma^3_{21})$$

$$\Gamma^3_{01}(\Gamma^0_{30}+\Gamma^1_{31}+\Gamma^2_{32}+\Gamma^3_{33})$$

$$-(\Gamma^3_{00}\Gamma^0_{31}+\Gamma^3_{01}\Gamma^1_{31}+\Gamma^3_{02}\Gamma^2_{31}+\Gamma^3_{03}\Gamma^3_{31})$$

The second part

$$\boldsymbol{\Gamma}^0_{01,0}+\Gamma^1_{01,1}+\Gamma^2_{01,2}+\Gamma^3_{01,3}$$

$$-\Gamma^0_{00,1}-\Gamma^1_{01,1}-\Gamma^2_{02,1}-\Gamma^3_{03,1}$$

gives

$$\frac{\partial}{partialt}(\frac{1}{2}B'B^{-1})=0.$$

Thus, $R_{01}=0$. *For* $b=0$, $d=2$

$$\Gamma^0_{02}(\Gamma^0_{00}+\Gamma^1_{01}+\Gamma^2_{02}+\Gamma^3_{03})$$

$$-(\Gamma^0_{00}\Gamma^0_{02}+\Gamma^0_{01}\Gamma^1_{02}+\Gamma^0_{02}\Gamma^2_{02}+\Gamma^0_{03}\Gamma^3_{02})$$

$$\Gamma^1_{02}(\Gamma^0_{10}+\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13})$$

$$-(\Gamma^1_{00}\Gamma^0_{12}+\Gamma^1_{01}\Gamma^1_{12}+\Gamma^1_{02}\Gamma^2_{12}+\Gamma^1_{03}\Gamma^3_{12})$$

$$\Gamma^2_{02}(\Gamma^0_{20}+\Gamma^1_{21}+\Gamma^2_{22}+\Gamma^3_{23})$$

$$-(\Gamma^2_{00}\Gamma^0_{22}+\Gamma^2_{01}\Gamma^1_{22}+\Gamma^2_{02}\Gamma^2_{22}+\Gamma^2_{03}\Gamma^3_{22})$$

$$\Gamma^{3}_{02}(\Gamma^{0}_{30}+\Gamma^{1}_{31}+\Gamma^{2}_{32}+\Gamma^{3}_{33})$$

$$-(\Gamma^{3}_{00}\Gamma^{0}_{32}+\Gamma^{3}_{01}\Gamma^{1}_{32}+\Gamma^{3}_{02}\Gamma^{2}_{32}+\Gamma^{3}_{03}\Gamma^{3}_{32})$$

again has no nonzero elements. The other part

$$\mathbf{\Gamma}^{0}_{02,0}+\Gamma^{1}_{02,1}+\Gamma^{2}_{02,2}+\Gamma^{3}_{02,3}$$

$$-\Gamma^{0}_{00,2}-\Gamma^{1}_{01,2}-\Gamma^{2}_{02,2}-\Gamma^{3}_{03,2}$$

has only zero elements and thus $R_{02}=0$.

For $b=0$, $d=3$ *there are no nonzero elements in the sum:*

$$\Gamma^{0}_{03}(\Gamma^{0}_{00}+\Gamma^{1}_{01}+\Gamma^{2}_{02}+\Gamma^{3}_{03})$$

$$-(\Gamma^{0}_{00}\Gamma^{0}_{03}+\Gamma^{0}_{01}\Gamma^{1}_{03}+\Gamma^{0}_{02}\Gamma^{2}_{03}+\Gamma^{0}_{03}\Gamma^{3}_{03})$$

$$\Gamma^{1}_{03}(\Gamma^{0}_{10}+\Gamma^{1}_{11}+\Gamma^{2}_{12}+\Gamma^{3}_{13})$$

$$-(\Gamma^{1}_{00}\Gamma^{0}_{13}+\Gamma^{1}_{01}\Gamma^{1}_{13}+\Gamma^{1}_{02}\Gamma^{2}_{13}+\Gamma^{1}_{03}\Gamma^{3}_{13})$$

$$\Gamma^{2}_{03}(\Gamma^{0}_{20}+\Gamma^{1}_{21}+\Gamma^{2}_{22}+\Gamma^{3}_{23})$$

$$-(\Gamma^{2}_{00}\Gamma^{0}_{23}+\Gamma^{2}_{01}\Gamma^{1}_{23}+\Gamma^{2}_{02}\Gamma^{2}_{23}+\Gamma^{2}_{03}\Gamma^{3}_{23})$$

$$\Gamma^{3}_{03}(\Gamma^{0}_{30}+\Gamma^{1}_{31}+\Gamma^{2}_{32}+\Gamma^{3}_{33})$$

$$-(\Gamma^{3}_{00}\Gamma^{0}_{33}+\Gamma^{3}_{01}\Gamma^{1}_{33}+\Gamma^{3}_{02}\Gamma^{2}_{33}+\Gamma^{3}_{03}\Gamma^{3}_{33})$$

and also

$$\mathbf{\Gamma}^{0}_{03,0}+\Gamma^{1}_{03,1}+\Gamma^{2}_{03,2}+\Gamma^{3}_{03,3}$$

$$-\Gamma^{0}_{00,3}-\Gamma^{1}_{01,3}-\Gamma^{2}_{02,3}-\Gamma^{3}_{03,3}$$

has only zero elements. Thus, $R_{03}=0$.

For $b=1$, $d=2$ there are two nonzero elements in

$$\Gamma^0_{12}(\Gamma^0_{00}+\Gamma^1_{01}+\Gamma^2_{02}+\Gamma^3_{03})$$

$$-(\Gamma^0_{10}\Gamma^0_{02}+\Gamma^0_{11}\Gamma^1_{02}+\Gamma^0_{12}\Gamma^2_{02}+\Gamma^0_{13}\Gamma^3_{02})$$

$$\Gamma^1_{12}(\Gamma^0_{10}+\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13})$$

$$-(\Gamma^1_{10}\Gamma^0_{12}+\Gamma^1_{11}\Gamma^1_{12}+\Gamma^1_{12}\Gamma^2_{12}+\Gamma^1_{13}\Gamma^3_{12})$$

$$\boldsymbol{\Gamma}^2_{12}(\Gamma^0_{20}+\Gamma^1_{21}+\boldsymbol{\Gamma}^2_{22}+\Gamma^3_{23})$$

$$-(\Gamma^2_{10}\Gamma^0_{22}+\Gamma^2_{11}\Gamma^1_{22}+\boldsymbol{\Gamma}^2_{12}\boldsymbol{\Gamma}^2_{22}+\Gamma^2_{13}\Gamma^3_{22})$$

$$\Gamma^3_{12}(\Gamma^0_{30}+\Gamma^1_{31}+\Gamma^2_{32}+\Gamma^3_{33})$$

$$-(\Gamma^3_{10}\Gamma^0_{32}+\Gamma^3_{11}\Gamma^1_{32}+\Gamma^3_{12}\Gamma^2_{32}+\Gamma^3_{13}\Gamma^3_{32})$$

but these elements cancel. The second part

$$\boldsymbol{\Gamma}^0_{12,0}+\Gamma^1_{12,1}+\Gamma^2_{12,2}+\Gamma^3_{12,3}$$

$$-\Gamma^0_{10,2}-\Gamma^1_{11,2}-\Gamma^2_{12,2}-\Gamma^3_{13,2}$$

has three nonzero elements on the lower row, but they are functions of r and the partial derivative by θ is zero, so $R_{12}=0$.

For $b=1$, $d=3$ there are no nonzero elements in

$$\Gamma^0_{13}(\Gamma^0_{00}+\Gamma^1_{01}+\Gamma^2_{02}+\Gamma^3_{03})$$

$$-(\Gamma^0_{10}\Gamma^0_{03}+\Gamma^0_{11}\Gamma^1_{03}+\Gamma^0_{12}\Gamma^2_{03}+\Gamma^0_{13}\Gamma^3_{03})$$

$$\Gamma^1_{13}(\Gamma^0_{10}+\Gamma^1_{11}+\Gamma^2_{12}+\Gamma^3_{13})$$

$$-(\Gamma^1_{10}\Gamma^0_{13}+\Gamma^1_{11}\Gamma^1_{13}+\Gamma^1_{12}\Gamma^2_{13}+\Gamma^1_{13}\Gamma^3_{13})$$

$$\Gamma^2_{13}(\Gamma^0_{20} + \Gamma^1_{21} + \Gamma^2_{22} + \Gamma^3_{23})$$

$$-(\Gamma^2_{10}\Gamma^0_{23} + \Gamma^2_{11}\Gamma^1_{23} + \Gamma^2_{12}\Gamma^2_{23} + \Gamma^2_{13}\Gamma^3_{23})$$

$$\Gamma^3_{13}(\Gamma^0_{30} + \Gamma^1_{31} + \Gamma^2_{32} + \Gamma^3_{33})$$

$$-(\Gamma^3_{10}\Gamma^0_{33} + \Gamma^3_{11}\Gamma^1_{33} + \Gamma^3_{12}\Gamma^2_{33} + \Gamma^3_{13}\Gamma^3_{33})$$

The second part

$$\boldsymbol{\Gamma}^0_{13,0} + \Gamma^1_{13,1} + \Gamma^2_{13,2} + \Gamma^3_{13,3}$$

$$-\Gamma^0_{10,3} - \Gamma^1_{11,3} - \Gamma^2_{12,3} - \Gamma^3_{13,3}$$

has three nonzero elements on the lower row, but they are functions of r *and the partial derivative by* ϕ *is zero. Thus,* $R_{13} = 0$.

Finally for $b = 2$, $d = 3$ *we find that the sum*

$$\Gamma^0_{23}(\Gamma^0_{00} + \Gamma^1_{01} + \Gamma^2_{02} + \Gamma^3_{03})$$

$$-(\Gamma^0_{20}\Gamma^0_{03} + \Gamma^0_{21}\Gamma^1_{03} + \Gamma^0_{22}\Gamma^2_{03} + \Gamma^0_{23}\Gamma^3_{03})$$

$$\Gamma^1_{23}(\Gamma^0_{10} + \Gamma^1_{11} + \Gamma^2_{12} + \Gamma^3_{13})$$

$$-(\Gamma^1_{20}\Gamma^0_{13} + \Gamma^1_{21}\Gamma^1_{13} + \Gamma^1_{22}\Gamma^2_{13} + \Gamma^1_{23}\Gamma^3_{13})$$

$$\Gamma^2_{23}(\Gamma^0_{20} + \Gamma^1_{21} + \Gamma^2_{22} + \Gamma^3_{23})$$

$$-(\Gamma^2_{20}\Gamma^0_{23} + \Gamma^2_{21}\Gamma^1_{23} + \Gamma^2_{22}\Gamma^2_{23} + \Gamma^2_{23}\Gamma^3_{23})$$

$$\Gamma^3_{23}(\Gamma^0_{30} + \Gamma^1_{31} + \Gamma^2_{32} + \Gamma^3_{33})$$

$$-(\Gamma^3_{20}\Gamma^0_{33} + \Gamma^3_{21}\Gamma^1_{33} + \Gamma^3_{22}\Gamma^2_{33} + \Gamma^3_{23}\Gamma^3_{33})$$

has no nonzero elements and in

$$\mathbf{\Gamma}^0_{23,0} + \Gamma^1_{23,1} + \Gamma^2_{23,2} + \Gamma^3_{23,3}$$

$$-\Gamma^0_{20,3} - \Gamma^1_{21,3} - \Gamma^2_{22,3} - \Gamma^3_{23,3}$$

the only two nonzero elements, $\Gamma^2_{23,3}$, *cancel. Therefore also* $R_{23} = 0$.

The Ricci curvature tensor has the following nonzero elements

$$R_{00} = \frac{1}{2}B''A^{-2} - \frac{1}{4}A'B'A^{-2} - \frac{1}{4}(B')^2A^{-1}B^{-1} + \frac{1}{r}B'A^{-1}$$

$$R_{11} = -\frac{1}{2}B''B^{-1} + \frac{1}{2}(B')^2B^{-2} + \frac{1}{4}A'B'A^{-1}B^{-1} + \frac{1}{r}A'A^{-1}$$

$$R_{22} = \frac{1}{2}rA'A^{-2} - \frac{1}{2}rB'A^{-1}B^{-1} + 1 - A^{-1}$$

$$R_{33} = \sin^2\theta\left(\frac{1}{2}rA'A^{-2} - \frac{1}{2}rB'A^{-1}B^{-1} + 1 - A^{-1}\right)$$

The Ricci scalar is

$$R = g^{ab}R_{ab}$$

which in this case reduces to

$$R = \sum_{a=0}^{3}\frac{R_{aa}}{g_{aa}}$$

Writing

$$f(B,A) = \frac{1}{2r}B'A{-}1B^{-1}$$

$$g(B,A) = \frac{1}{2}B''A^{-2}B^{-1} - \frac{1}{4}A'B'A^{-2}B^{-1} - \frac{1}{4}(B')^2A^{-1}B^{-2}$$

$$h(A) = -\frac{1}{2r}A'A^{-2}$$

$$s(A) = -\frac{1}{r^2}(1 - A^{-1})$$

we have

$$\frac{R_{00}}{g_{00}} = 2f(B,A) + g(B,A)$$

$$\frac{R_{11}}{g_{11}} = g(B,A) + 2h(A)$$

$$\frac{R_{22}}{g_{22}} = f(B,A) + h(A) + s(A)$$

$$\frac{R_{33}}{g_{33}} = f(B,A) + h(A) + s(A)$$

The Ricci scalar is thus

$$R = 4f(B,A) + 2g(B,A) + 4h(A) + 2s(A)$$

A solution of Einstein's equations for the vacuum, i.e., $T_{ab} = 0$ *and* $\lambda = 0$ *in Einstein's field equation*

$$R_{ab} - \frac{1}{2}R = 0$$

requires that every R_{ab} *must be zero. From it forllows that* $R = 0$. *Thus, if* $A(r)$ *and* $B(r)$ *give an exact solution for Einstein's equation in the vacuum,* $R_{aa} = 0$ *for every* a.

If $R_{00} = R_{11} = 0$ *then*

$$\frac{R_{00}}{g_{00}} - \frac{R_{11}}{g_{11}} = 2f(B,A) - 2h(A) = 0$$

From $f(B,A) - h(A) = 0$ *follows*

$$2r(f(B,A) - h(A)) = B'A{-}1B^{-1} + A'A^{-2} = 0$$

This equation gives

$$A'A^{-1} + B'B^{-1} = 0$$

Thus $AB = K$ *for some constant* K.

Inserting $f(A,B) = h(A)$ *the Einstein's equations simplify to*

$$2h(A) + g(B,A) = 0$$

$$2h(A) + s(A) = 0$$

$$g(B, A) + 4h(A) + s(A) = 0$$

Eliminating $g(B, A) = -2h(A)$ *and* $s(A) = -2h(A)$ *yields*

$$-2h(A) + 4h(A) - 2h(A) = 0$$

showing that the Ricci scalar equation is satisfied.

In the derivation of Schwarzschild's solution $B'B^{-1} = -A'A^{-1}$ *is inserted to*

$$f(B, A) = \frac{1}{2r}B'A{-}1B^{-1} = -\frac{1}{2r}A'A{-}2$$

Expanding

$$f(B, A) + h(A) + s(A) = 0$$

with

$$h(A) = -\frac{1}{2r}A'A^{-2}$$

$$s(A) = -\frac{1}{r^2}(1 - A^{-1})$$

yields

$$-\frac{1}{2r}A'A{-}2 - \frac{1}{2r}A'A^{-2} - \frac{1}{r^2}(1 - A^{-1}) = 0$$

Thus

$$-A'A{-}2 - A'A^{-2} - \frac{2}{r}(1 - A^{-1}) = 0$$

$$rA' + A(A - 1) = 0$$

This equation has the solution

$$A(r) = (1 + \frac{k}{r})^{-1}$$

for some k, *Schwarzschild's exact solution to Einstein's equations.*