Robustness of GR, Attempts to Modify Gravity

Lecturer: Prof. Nima Arkani-Hamed
Transcriber: *Alexander Chen*

August 7, 2011

Contents

1 Lecture 1

We are going to talk about attemps to modify gravity, and why it is so hard. We are interested in attempts that are motivated by astrophysics at large distances. We seem to have some breakdown of gravity at large distances, like galaxy rotation curves. Dark matter and dark energy were dreamed up to explain these, but it could be that gravity is wrong. There is one big motivation to throw away dark matter and dark energy is the cosmological constant problem. The vacuum energy is easily weak scale to the 4th, but we know that the vacuum energy is extremely small. There is a fine-tuning explanation, but we might want a fundamental or mechanism way to explain it. This is by far the best motivation for modifying gravity.

The literature is filled with ideas of modifying gravity. There are some non-ideas, some silly ideas, and a few are very good ideas. We will examine them. We will first show why gravity is special, and then talk about the good ideas and their main issues. The slogan for these lectures is “don’t modify gravity, but try to understand it instead!”

The first reason not to modify gravity is that it is the unique theory which is generally covariant and nice and beautiful, but that is a dumb reason. Another reason is that it is a gauge theory, but we know we can add a mass to a gauge theory and modify it in the infrared. We will instead show the statement *GR is the unique low energy theory for interacting massless spin 2 particles.* Similarly Yang-Mills theory is the unique low energy theory for interacting massless spin 1 particles. Therefore the two dominating strutures that we know are completely determined structures in long distances.

All the drama has to do with the fact that massless particles only have *helicity* as their spin degrees of freedom. Therefore photons and gravitons only have two helicity states. This is equivalent to the fact that you can’t boost into the rest frame of a massless particle. So there is a discontinuous transition from massive particles to massless particles. The other assumption is that physics is local, and what happens on the other planet doesn’t affect the local physics here. Therefore we group the ladder operators into fields, and this is the reason why field theory occurs in many branches of physics, because it makes locality manifest.

Now if we want to describe a spin 1 particle we use a vector field $A_\mu(x) = \epsilon_\mu(p)e^{ipx}$. But there are 4 indices and we have 4 degrees of freedom. We need some constraints to restrict the number to 2. One is that $p^\mu\epsilon_\mu(p) = 0$, but this only reduces to 3 degrees of freedom, but then we are stuck. In fact there does not exist objects like $\epsilon^\pm_\mu(p)$ which transform like 4-vectors. To be explicit, say we construct

$$p^\mu = (E, 0, 0, E)\,, \quad \epsilon^\pm_\mu = \left(0, \frac{1}{\sqrt{2}}, \pm\frac{i}{\sqrt{2}}, 0\right) \tag{1}$$

But this set of polarization vector do not transform into themselves. Under Lorentz transformations we have instead

$$\epsilon_\mu \to (\Lambda\epsilon)_\mu + \alpha(\Lambda p)_\mu \tag{2}$$

This is natural because $p^2 = 0$, therefore $\epsilon + \alpha p$ will still be orthogonal to p. So what we have to do is to introduce a redundancy, meaning that $\epsilon_\mu(p)$ and $\epsilon_\mu(p) + \alpha(p)p_\mu$ should be *identified*. This is nothing other than the usual misnomer “gauge symmetry”, but this is just a redundancy. So this reduces our degrees of freedom down to 2.

Similar things can be done for spin 2 particles. We can define $h_{\mu\nu}$ and require that $p^\mu h_{\mu\nu} = 0$ and ${h^\mu}_\mu = 0$, but we are left with 5 degrees of freedom. Therefore we need to have the redundancy

$$h_{\mu\nu} \leftrightarrow h_{\mu\nu} + \alpha_\mu p_\nu + \alpha_\nu p_\mu \tag{3}$$

where $\alpha \cdot p = 0$ so we have 3 redundancy and the degrees of freedom is reduced to 2. Remember that these are just an artifact from the requirement that we need to describe things in the settings of field theory, where everything is local.

We can define an invariant quantity $F_{\mu\nu} = p_\mu \epsilon_\nu - p_\nu \epsilon_\mu$ under this redundancy. For gravity we should produce the Riemann tensor. Let's see how we can reinvent it. Let's say we can decompose $h_{\mu\nu}$ into something like

$$h_{\mu\nu} = a_\mu^A a_\nu^B \eta_{AB} \tag{4}$$

This only says that any symmetric rank two tensor can be decomposed into a quadratic form. With the transformation of α to be $\delta a_\mu^A = \zeta^A(p) p_\mu$, we can have the previous transformation for $h_{\mu\nu}$. Then we can construct invariant quantities. For example $p_\mu a_\nu^A - p_\nu a_\mu^A$ is invariant under $\delta a_\mu^A = \zeta^A(p) p_\mu$, but this quantity can't be expressed using $h_{\mu\nu}$ only, because $h_{\mu\nu}$ has A and B indices contracted. So we are naturally led to

$$\frac{1}{2}\left(p_\mu a_\nu^A - p_\nu a_\mu^A\right)\left(p_\alpha a_\beta^B - p_\beta a_\alpha^B\right)\eta_{AB} \tag{5}$$

and this actually is the linearized Riemann tensor $R_{\mu\nu\alpha\beta}$.

We can then write down the equation of motion. In spin 1 gauge theory we have $p^\mu F_{\mu\nu} = 0$. Now for spin 2 theory we will get instead

$$R_{\mu\ \ \alpha\beta}^{\ \ \beta} = 0 \tag{6}$$

This is the old approach to GR as a gauge theory, but it becomes ungainly when we come to higher nonlinear level. For example to first order we couple A_μ to a current J^μ and $h_{\mu\nu}$ to the energy momentum tensor $T^{\mu\nu}$. Then the redundancy requires that $p_\mu T^{\mu\nu} = 0$, which is just the conservation law. Remember that the redundancy is introduced only to gaurantee Lorentz invariance and unitarity. For example, if we consider the diagram like The amplitude has to be like

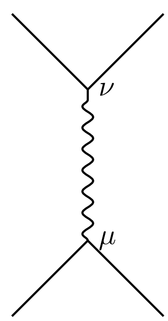

Figure 1: Simplest Diagram

$$\frac{i}{p + i\varepsilon} J^\mu N_{\mu\nu} J^\nu \tag{7}$$

Now let's figure out what $N_{\mu\nu}$ is. Because $p^2 = 0$, we would expect that

$$N_{\mu\nu} = \sum_{h=\pm 1} \epsilon_\mu^{*h}(p) \epsilon_\nu^h(p) \tag{8}$$

But remember that ϵ_μ is not even Lorentz invariant! If we evaluate this quantity using the above definition of ϵ_μ, we will see that this quantity is identity matrix in the x-y subspace. But then we can write it in another way

$$N_{\mu\nu} = -\eta_{\mu\nu} + \frac{p_\mu \bar{p}_\nu + p_\nu \bar{p}_\mu}{2p \cdot \bar{p}} \tag{9}$$

where $\bar{p}$ is defined by reversing the sign of the time component. Now this is Lorentz invariant, and this is what the redundancy buys us. So we can exploit the redundancy to write $N_{\mu\nu} = -\eta_{\mu\nu}$, or we can write $N_{\mu\nu} = -\eta_{\mu\nu} + \xi p_\mu p_\nu / p^2$. This is the familiar "choosing the gauge".

Now let's do the same thing for spin 2. Now we have $N_{\mu\nu,\alpha\beta}$. Remember that $\epsilon^+_{\mu\nu} = \epsilon^+_\mu \epsilon^+_\nu$, which is easy to see because of the rotation properties. Similar for the $-$ component. So now

$$\begin{aligned} N_{\mu\nu,\alpha\beta} &= \epsilon^{+*}_{\mu\nu}\epsilon^{+}_{\alpha\beta} + \epsilon^{-*}_{\mu\nu}\epsilon^{-}_{\alpha\beta} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{\mu\nu} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{\alpha\beta} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{\mu\nu} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{\alpha\beta} \\ &= C\left[-\eta_{\mu\nu}\eta_{\alpha\beta} + \eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}\right] \end{aligned} \tag{10}$$

where μ,ν,α,β are restricted to 1 and 2, and evaluating the constant we get $C = 1$. It can be shown that in D dimensions we have instead

$$N_{\mu\nu,\alpha\beta} = \left[-\frac{2}{D-2}\eta_{\mu\nu}\eta_{\alpha\beta} + \eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}\right] \tag{11}$$

Note also that we have also shown that like charges repel and opposite charges attract. And note that in $D = 3$ which means $2 + 1$ dimensions we don't have graviton interactions. Now because everything has an energy momentum tensor the graviton must have nonlinear coupling and we need to build up all the nonlinear coupling in our Lagrangian.

Let's see another argument which is by Weinberg. Imagine a process of a bunch of particles coming in and a bunch of particles coming out. We want to study the amplitude of this process involving an extra massless particle with momentum q. Say this is associated to a particle with momentum p_i. Let's study the limit where this particle is very soft $q \to 0$. There will be a term in the amplitude which looks like

$$\frac{1}{(p_i + q)^2 - M_i^2} \xrightarrow{q \to 0} \frac{1}{2p_i \cdot q} \tag{12}$$

That's not done yet. The numerator of this amplitude would be proportional to $\epsilon^\mu(q) p_{i\mu}$. Therefore the amplitude can be decomposed as

$$M(p_i, q) = M(p_i) \times \sum_i \frac{\epsilon^\mu(q) p_{i\mu} e_i}{2p_i \cdot q} \tag{13}$$

in the limit that q is very soft, and e_i is the charge of the particle i. Now by our redundancy, when we replace ϵ with q we should get zero in the amplitude. Then we get

$$S \to \frac{1}{2}\sum e_i = 0 \tag{14}$$

Therefore we must conclude that charge is conserved. Now let's do this for graviton. We have instead $\epsilon^{\mu\nu}$ in the numerator and we need to have

$$S \to \alpha^\nu \sum \kappa_i p_{i\nu} = 0 \tag{15}$$

where κ_i is what is analogous to e_i. But now either we have scattering only at discrete angles, or that all κ_i are equal. This is how we discover the principle of equivalence.

Let's see how non-Abelian gauge theories work in this language. Consider the Compton scattering, again we should have $M \to 0$ when $\epsilon_\mu \to q_\mu$, but the individual diagrams do not vanish, but the sum of the two diagrams vanish. Now if we have an additional index a at the external gauge legs. If we just write the vertex for interaction as $(T^a)_{ik}$ then the amplitude will be

$$\left(T^a T^b\right)_{ij} - \left(T^b T^a\right)_{ij} = \left[T^a, T^b\right]_{ij} \tag{16}$$

So we have to include the interaction of three gauge bosons with coefficient f^{abc} and do

$$\left[T^a, T^b\right] = i f^{abc} T^c \tag{17}$$

In the above argument we can't have spin 3 or higher spins because then we will only have scattering at specific angles, because of the structure of the initial data we have. Another thing is that we can construct theories that couple to matter as $F_{\mu\nu}$ directly, but they can't mediate inverse square law forces.

2 Lecture 2

Last time we showed that the massless spin 1 and spin 2 particles are described by a unique theory at low energies. The reason we always say low energies is that we nail the 2 derivative action which is leading at long distances, and from here the other structures are determined. We work at weak interactions because the names like "spin 2 particles" suggests that there exists free asymptotic states and the interaction goes to zero. So we have implicitly assumed weak coupling. Fortunately for gravity we know that this weak coupling limit is adequate, because gravity is weak to start with.

We have always emphasized that the gauge symmetry is a redundancy, and real physics should not be dependent on this thing. But we all know that gauge invariance seriously restricts our ways of describing the physics. This is because although there is no physical content in the redundancy by themselves, the structure makes it very difficult to *only* describe a spin 1 or spin 2 particle without introducing extra degrees of freedom. It is in this respect that the redundancy limits us, and we have to invent the well-known ways of describing gauge theories.

Now let's consider scattering amplitudes for massless high-spin particles. Remember we were calculating physical amplitudes like $M^{h_1\dots h_n}(p_1,\dots,p_n)$ and there is no Lorentz indices. But the way we computed them using a redundant form

$$M = \epsilon^{h_1}_{\mu_1}\dots\epsilon^{h_n}_{\mu_n}M^{\mu_1\dots\mu_n} \tag{18}$$

And we know that this is not necessary. Now how to describe it in a formalism-free way?

In order to describe a massless particle we usually use a four vector for the momentum p^μ with a constraint $p^\mu p_\mu = 0$. We would like to introduce a way to represent it without this constraint. We dot it with the Pauli matrices

$$p_\mu\sigma^\mu_{A\dot{A}} = \begin{pmatrix} p_0+p_2 & p_1+ip_2 \\ p_1-ip_2 & p_0-p_2 \end{pmatrix} \tag{19}$$

and the determinant of this matrix is just $p^\mu p_\mu = p^2$. This is good because if we do any 2×2 unitary transformation on the matrix the determinant is not modified. This gives us the spinor representation of the Lorentz group. Now if $p^\mu p_\mu = 0$ then this matrix is either zero, or of rank 1. If it is rank 1 then the matrix can be written as

$$P_{A\dot{A}} = \lambda_A\tilde{\lambda}_{\dot{A}} \tag{20}$$

where λ_A is a complex 2-vector. The reason for having A and $\dot{A}$ indices is because they can transform differently under Lorentz transformation, and invariants can be constructed as

$$\langle 12\rangle = \epsilon^{AB}\lambda_{1A}\lambda_{2B}, \quad [12] = \epsilon^{\dot{A}\dot{B}}\tilde{\lambda}_{1\dot{A}}\tilde{\lambda}_{2\dot{B}} \tag{21}$$

If the momentum is real then we should have $\tilde{\lambda} = \lambda^*$. Now these vectors are unconstrained degrees of freedom. But these λ are not unique, because we can scale

$$\lambda \to t\lambda, \quad \tilde{\lambda} \to t^{-1}\tilde{\lambda} \tag{22}$$

For real momenta t can only be a phase $e^{i\theta}$. But this is intuitively clear because for a photon with a specific helicity we can rotate around the axis of p and preserve the helicity. Therefore now we can write the scattering amplitudes as functions of λ and $\tilde{\lambda}$ and require that

$$M^{h_1\dots h_n}[t_i\lambda_i, t_i^{-1}\tilde{\lambda}_i] = t_i^{-2h_i}M[\lambda_i,\tilde{\lambda}_i] \tag{23}$$

This is to count for the invariance under the above scaling. For example, say

$$M(1^-2^+3^-4^+) = C\frac{\langle 13\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle} \tag{24}$$

This is an actually amplitude. If we scale λ_1 to $t_1\lambda_1$, then this amplitude scales like $t_1^2 M$, so we can see that λ_1 has helicity -1, and it agrees with the requirement. We are now starting with the very opposite side from the previous formalism. But this is a little more general because whatever formalism we start with, we have to produce the above structure.

Now let's look at 3-particle amplitudes. We never talk about this because for massless particles because by relativity a massless particle can't split into two and conserve momentum. But this is an artifact of working with real momentum. Now if we generalize to talk about complex momentum then we can work with nonzero 3-particle amplitudes. In 2 to 2 scattering we have the Mendelstam variables s, t, and u, and the amplitudes are functions of them. But for 3-particle amplitudes we don't have these variables because

$$p_1\cdot p_2 = (p_1+p_2)^2 = p_3^2 = 0 \tag{25}$$

So the scattering amplitude can only depend on helicity states. Now because $p_1+p_2+p_3=0$ we have

$$\lambda_{1A}\tilde{\lambda}_{1\dot{A}} + \lambda_{2A}\tilde{\lambda}_{2\dot{A}} + \lambda_{3A}\tilde{\lambda}_{3\dot{A}} = 0 \tag{26}$$

If we contract the above equation with $\lambda_{1B}\epsilon^{AB}$ then we have

$$\langle 12\rangle\,\tilde{\lambda}_2 + \langle 13\rangle\,\tilde{\lambda}_3 = 0 \tag{27}$$

If we assume $\langle 12\rangle \neq 0$, then $\langle 13\rangle$ had better not be zero or this amplitude will be very singular. Therefore $\tilde{\lambda}_2$ and $\tilde{\lambda}_3$ are proportional, and we have $[23]=0$. By momentum conservation then we can conclude that either all $\langle ij\rangle = 0$ and $[ij]\neq 0$ or $[ij]=0$ and the other way around. So when the momentum is real both must vanish and this condition can't be satisfied.

Now note that $p_1\cdot p_2 = \langle 12\rangle[12] = 0$. So writing 3-particle amplitudes we can either use angular brackets or use square brackets. Let's say use angular brackets, and by symmetry we should have

$$M = C\,\langle 12\rangle^{a_1}\langle 23\rangle^{a_2}\langle 31\rangle^{a_3} \tag{28}$$

If the spin state of particle i is $-s_i$ then we have by scaling

$$a_2+a_3 = 2s_1,\quad a_3+a_1 = 2s_2,\quad a_1+a_2 = 2s_3 \tag{29}$$

So we can solve for the powers to get

$$a_1 = s_2+s_3-s_1,\quad \text{and cyclic for else} \tag{30}$$

If we started with square brackets the the powers just become minus these. So just from Poincare invariance we have nailed down the exact form of the amplitude! Specifically we have for example

$$M\left[1^-,2^-,3^+\right] = C\left(\frac{\langle 12\rangle^3}{\langle 13\rangle\langle 23\rangle}\right)^s \tag{31}$$

where s is the overall spin. Now we can also have a linear combination with the term for square brackets. But in the limit that the amplitude goes real we should recover that amplitude go to zero, and this gets rid of the square case.

Now if all the particles are in the spin state $-s$ then the amplitude is

$$M\left[1^{-},2^{-},3^{-}\right]=C\left(\langle 12\rangle\langle 23\rangle\langle 31\rangle\right)^{3} \tag{32}$$

Note this amplitude has different dimension as the above one, and they are suppressed by the energy level. Now if we compare the amplitude for $s=1$ and $s=2$, the gravity amplitude is the square of the Yang-Mills amplitude. The amplitude for Yang-Mills is dimensionless and proportional to g, and that of gravity is proportional to $1/M_p$. And these amplitudes include all the quantum effects because we started with only Poincare invariance. And thus we can write down anything up to spin anything. Note also that the above amplitude for 1^- and 2^- and 3^+ is antisymmetric under exchange of 1 and 2, but this violates Bose statistics. This tells us that we can't have this amplitude for single spin 1 particle, but we can have it for many, labeled by a, b, and c, and the amplitude is proportional to f_{abc}, as is in Yang-Mills theories.

Now let's try to write down an amplitude for 4 particles with 2 plus and 2 minus. This is equal to

$$M\left[1^{-},2^{-},3^{+},4^{+}\right]=\left(\langle 12\rangle[34]\right)^{2S}F(s,t,u) \tag{33}$$

By unitarity we know that when $s\to 0$ then we should have a pole, and the amplitude should factorize into two parts. This is where unitarity and locality come in, which determine the pole structure of F. At $s\to 0$ we should have

$$M\left[1^{-},2^{-},3^{+},4^{+}\right]=\sum_{h}M[1,2,h]\times M[-h,3,4]\times\frac{1}{s} \tag{34}$$

This is the leading order contribution to the 4 particle amplitude. Similarly for the other channels. Let's send s to 0 by sending $[12]\to 0$, and the 3-point amplitude going into the intermediate states would be

$$M=\left(\frac{\langle 12\rangle^{3}}{\langle 1I\rangle\langle 2I\rangle}\right)^{S}\left(\frac{[34]^{3}}{[3I]\,[4I]}\right)\frac{1}{s} \tag{35}$$

We can combine the denominator using the contraction of λ and $\tilde{\lambda}$'s. So we have

$$M=\left(\frac{\langle 12\rangle^{3}[34]^{3}}{\langle 12\rangle[23]\langle 23\rangle[34]}\right)^{S}\frac{1}{s}=\frac{\left(\langle 12\rangle[34]\right)^{2S}}{t^{S}}\frac{1}{s} \tag{36}$$

Therefore we conclude that

$$F(s,t,u)\xrightarrow{s\to 0}\frac{1}{s}\frac{1}{t^{S}} \tag{37}$$

Similarly for s and t changed. Now the problem has been changed to a math problem, namely *is it possible to find a mathematical function F that satisfy this requirement*? The answer is yes, but only for $S=0$ or $S=2$. When $S=0$ we have

$$F(s,t,u)=\frac{1}{s}+\frac{1}{t}+\frac{1}{u} \tag{38}$$

and for $S=2$ we have

$$F(s,t,u)=\frac{1}{stu} \tag{39}$$

Now we have determined the scattering amplitude for gravitons completely without even introducing Lagrangians and stuff.

Let's look at F by introducing a parameter z and define

$$f(z) = F(s + \alpha z, t + \beta z, u - \alpha z - \beta z) \tag{40}$$

We need that $f(z) \to 0$ when $z \to \infty$. Now assume that F exists with the above described pole structure, then $f(z)$ has poles on the complex plane when $s + \alpha z = 0$, $t + \beta z = 0$ or $u + \gamma z = 0$. These are all the poles of the function F therefore the function f. Now we know from complex analysis that if we know all the poles and the residues then we can construct the whole function by Cauchy's theorem. We are interested in F which is just when $f(0)$

$$f(z = 0) = \oint_{z=0} \frac{dz}{z} f(z) = \sum_{\text{res}} \frac{1}{z_{\text{res}}} \text{Res}\, f(z_{\text{res}}) \tag{41}$$

In the end we will find as an exercise that

$$f(0) = F(s, t, u) = \left(\frac{\alpha^S}{s} + \frac{\beta^S}{t} + \frac{\gamma^S}{u} \right) \frac{1}{(\alpha t - \beta s)^S} \tag{42}$$

Now if $S = 0$ we are done, and we have a nice function. We can show that when $S = 2$ then the pole at $\alpha t = \beta s$ is false, because the coefficient at the pole is zero. For any other S we have a contradiction that this function has an extra pole, and depend on α, β and γ which it should not depend on. Similarly we can add a, b, c indices for spin 1 particles and write down the most general amplitude, and we will see that we need to introduce f_{abc} which satisfy the Jacobian identity, and we find the scattering amplitude for gluons. For multiple massless spin 2 particles we can show that their interaction can be diagonized, and different amplitudes decouple.

Suppose we have spin S particles interacting with gravity, the scattering amplitude $M[\varphi_1^-, \varphi_2^+, 3^{\pm}]$ has units of $1/M_p$ and is the leading interaction. Then we consider the consistency of $M[\varphi_1^-, \varphi_2^+, \varphi_3^-, \varphi_4^+]$ requires that the spin S can only be 0, 1/2, 1, or 3/2.

3 Lecture 3

Today we will start to talk about the possible ways of modifying gravity and the challengings. What does it mean to modify gravity? The problem is not about adding a term in the Lagrangian, but about modifying the actual force between the distant planets. 50 years ago Brans-Dicke gave a model which is just GR plus a scalar field with an interaction like

$$(\partial\varphi)^2 + \frac{c\varphi}{M_{\mathrm{pl}}} T^{\mu}{}_{\mu} \tag{43}$$

But this in effect does not significantly change the force between distant planets, because it only changes the effective value of G_N. But this will not couple to radiation, and will change the way light is bent by gravity. This gives the constraint on the coefficient $c < 10^{-3}$. This theory is theoretically boring, and experimentally $c \sim 1$ has been ruled out, and $c \sim 10^{-3}$ is just irrelevant. There is a class of $F(R)$ theory which is completely content-free, and is equivalent to a scalar-tensor theory. There is a class of $F(R)$ theory which is completely content-free, and is equivalent to a scalar-tensor theory.

We will see that in the end of the day any theory boils down to adding scalar degrees of freedom coupled to matter in the above way with $c \sim 1$ and not ruled out. The differences lie in the self interaction sector. One example is the Galileon theory which introduces a novel symmetry structure. There are other examples like spontaneously breaking Lorentz symmetry, ghost condensate, etc. The moral is that a new class of low energy effective theories at long distances that have novel symmetry features is found and studied.

One of the reasons to modify gravity is to solve the cosmological constant problem. We need to modify gravity such that at long distances the huge cosmological constant do not gravitate and it does not ruin the large scale structure. One model is to have a brane which is our spacetime, which is living in a number of extra dimensions, and vacuum energy just give tension on the brane and give rise to gravitation in the bulk, but not in the brane. But it's hard to modify only the cosmological constant but nothing else, so nothing gravitates now. What about compactifying the extra dimensions? Then we have gravitation, but then the low energy effective theory would not know any structure like that, so this do not essentially solve the problem. The idea of DGP is to trap gravity on the brane, in order to make gravity 4-dimensional. Their concrete proposal is to add one extra dimension, and write $M_{\mathrm{pl}}^2 R^{(4)}$ for the brane and $M_5^2 R^{(5)}$ in the bulk. The propagator looks like

$$\frac{1}{p^2 M_{\mathrm{pl}}^2 + |p|\, M_5^2} \tag{44}$$

so there is a crossing scale $p_c^{-1} \sim r_c \sim M_{\mathrm{pl}}^2/M_5^2$, and beyond this gravitates as 5D. By tuning parameters we can make this scale large enough. This is a concrete proposal, but this does not solve the cosmological constant problem, because the gravitation in the brane is 4D, so it behaves like 4D and the cosmological constant still gravitates as usual. There are about 10 variation to avoid this problem, but all of them failed.

Let's say we have a magical way of taking the cosmological constant out, and degravitate it. But in the early universe we have only a single fluid $T_{\mu\nu} = \mathrm{diag}(\rho, P, P, P)$. And there is no way we can distinguish vacuum energy piece from radiation and matter. Now we need some crazy thing because the crucial difference between vacuum energy and the other things is that vacuum energy does not get diluted. So this crazy thing must be acausal. So either we introduce something that is acausal, or we introduce something that is causal and extra degrees of freedom but do not solve the CC problem.

If we want to degravitate cosmological constant, we look at Einstein equations

$$M_{\mathrm{pl}} G_{\mu\nu} = T_{\mu\nu} \tag{45}$$

On the right hand side we have enormous vacuum energy, and we don't want it to give rise to a huge curvature, so we want to modify the left hand side. We can change it like

$$M_{\rm pl}\left[1+F(L^2\Box)\right]G_{\mu\nu}=T_{\mu\nu} \tag{46}$$

where $L\sim H^{-1}$ and require that $F(x)$ be huge for $x\ll 1$ but $F(x)$ go to zero when $x>1$. But if we solve the equations of motion assuming maximal symmetry, we will find very small Hubble. The equation of motion can be written schematically as

$$(1+F(L^2\Box))\Box\varphi=-\frac{T}{M_{\rm pl}} \tag{47}$$

Now if cosmological constant degravitates, then it could be that $1+F=0$, then it does not solve the problem, but just adds new degrees of freedom. So we need to assume that $1+F$ does not have zero on the complex plane, then it is okay. But this violates causality. The Green function of this equation is

$$G(t)\xrightarrow{p\to 0}\int\frac{e^{-i\omega t}}{1+F(m^2L^2)} \tag{48}$$

Now because the Green function has no poles we can just compute the Fourier integral, and see that it is symmetric in time, and is not causal. This is because that the Green function has no poles. The poles are important because it is their existence that requires us to use contours and use advanced or retarded conditions. This is analogous as that in eternal inflation and anthropic argument, where they essentially choose a future boundary condition.

Now let's talk about modifying gravity in this more restricted sense, which is adding new degrees of freedom. Let's consider first gauge theories and consider a massive spin 1 particle. The propagator of this particle should be like

$$\frac{iN_{\mu\nu}}{p^2-M^2+i\epsilon} \tag{49}$$

where $N_{\mu\nu}$ is again given as a polarization sum. Here it is easier because we can go to the rest frame of the particle and see that we have 3 polarizations, and we can choose the polarization such that

$$N_{\mu\nu}=\delta_{ij} \tag{50}$$

where i and j are spatial indices, so we can write it in a Lorentz invariant way

$$N_{\mu\nu}=-\eta_{\mu\nu}+\frac{p_\mu p_\nu}{M^2} \tag{51}$$

Now we can talk about what the Lagrangian is, and the inverse of the propagator must give us the Lagrangian

$$\mathcal{L}^{\rm free}=-\frac{1}{4g^2}F_{\mu\nu}^2+\frac{1}{2}f^2A_\mu A^\mu \tag{52}$$

Now we can have objection that this is not gauge invariant, but that is irrelevant. Now a better question is that why the kinetic term is gauge invariant but only the mass term is not, and why we can't write down terms like $(\partial_\mu A^\mu)^2$.

Let's employ the Stuckelberg trick. There should be some limit where we can think of this particle as massless, like at very high energies. But we know that there is discontinuous change because we have one more degree of freedom. Plus in the limit we should have gauge redundancy at this limit for the massless particle. The above Lagrangian is not gauge invariant, and we can show it easily

$$\mathcal{L} \to -\frac{1}{4g^2}F_{\mu\nu}^2 + f^2(A_\mu + \partial_\mu \pi)^2 \tag{53}$$

So this is not gauge invariant. But it is trivial to make it gauge invariant by elevating π to a degree of freedom, which is the longitudinal degree of freedom. And there is no content in this elevation. This is useful because we have introduced gauge redundancy, and we have separated the longitudinal degree of freedom from the other two degrees. This is called the Stuckelberg trick.

In case we think that this is not as good as the usual gauge theory, let's integrate out π. If we do that, we have to get a Lagrangian which is gauge invariant, but because π is massless, we will get a nonlocal action

$$\frac{1}{4g^2}F^{\mu\nu}\left(1 + \frac{m^2}{\Box}\right)F_{\mu\nu} \tag{54}$$

This looks like our attempts at modifying gravity, but there is nothing here but repackaging. Like we said, gauge redundancy has no use unless we only want to describe the limited degrees of freedom dictated by 2 helicity states.

Now we can couple A to a fermion. Accidentally the only interaction we can write down is gauge invariant, so π does not couple to the fermion and at high energy limit it just decouples and acts like a free field. Now let's look at non-Abelian theories

$$-\frac{1}{4g^2}\operatorname{tr} F_{\mu\nu}^2 + \frac{1}{2}f^2 \operatorname{tr}(A_\mu A^\mu) \tag{55}$$

Again there is a scale $m^2 = g^2 f^2$ above which the gauge particle can be seen as massless. But there is another scale, which is f. If we make a gauge transformation

$$A_\mu \to U^\dagger(A_\mu + \partial_\mu)U, \quad U = e^{i\pi^a T^a} \tag{56}$$

Now we can again evelate π to a degree of freedom. Its Lagrangian will be

$$\mathcal{L} = f^2 \operatorname{tr} \partial_\mu U^\dagger \partial^\mu U = f^2\left[\operatorname{tr}(\partial\pi)^2 + (\partial\pi)^2\pi^2 + \dots\right] \tag{57}$$

and we have a whole tower of nontrivial interactions. We can go to a canonical normalization where $\pi^c = f\pi$ then the above self interactions are

$$\operatorname{tr}(\partial\pi^c)^2 + \frac{(\partial\pi^c)^2\pi^{c2}}{f^2} + \dots \tag{58}$$

The interaction becomes of order 1 when the energy scale approaches order f, so this description breaks down and new physics must occur. There is a separation of these two scales m and f which are parametrically separated by $1/g$. In between these scales we have a good effective field theory, but above the scale f we need to have new physics. This is because we introduced new degrees of freedom. This same argument can be repeated to show that there is something new at LHC energies because the scale is parametrically higher than the electroweak scale.

We could arrive at the above conclusion right away by brutally calculating the amplitudes. If we look at the propagator we computed using the polarization sum

$$\left(-\eta_{\mu\nu} + \frac{p_\mu p_\nu}{M^2}\right) \tag{59}$$

Now if we boost this to very high energy $E \gg M$, and the longitudinal polarization vector grows like $\epsilon_\mu^L \sim p_\mu/M$. Now if we look at the 4 particle scattering amplitudes and then the leading piece will give $g^2 E^2/m^2 \sim (E/f)^2$. The bottom line for modifying non-Abelian gauge theories at infrared is that we introduce a new degree of freedom and a new scale where new physics happens.

4 Lecture 4

Last time we considered the Lagrangian for π. It is a non-renormalizable theory, and we expect the effective theory to break down at a new UV scale f. Schematically we need to add new terms like

$$\mathcal{L}_\pi = f^2 \left[\partial_\mu U \partial^\mu U + (\partial^2 U)^2 + \dots\right] \tag{60}$$

But what does the new terms look like in the unitary gauge? The $(\partial^2 U)^2$ looks like $(\partial_\mu A^\mu)^2$, and we could also have A_μ^2 etc. Remember we asked the question that even though we don't have gauge symmetry, why don't we have terms like $(\partial_\mu A^\mu)^2$. Now these term show up because of the interactions of U. But these terms are parametrically small by a factor of g. What if we don't have this suppression? Suppose we have in the Lagrangian

$$\frac{1}{4g^2}F_{\mu\nu}^2 + \frac{c}{g^2}(\partial_\mu A^\mu)^2 + f^2 A_\mu A^\mu \tag{61}$$

Then under the gauge transformation we have a term like $(1/g^2)(\Box\pi)^2$, and this adds an extra pole in the propagator of π, and we have added an unphysical degree of freedom. The pathology of the ghost has to be small if we are to have a good effective theory, so the coefficient c has to be small, like the scale of the cutoff f, which is the same as what we have before.

Now let's talk about massive gravity. Let's again look at the propagator of the massive spin 2 particle which is $iN_{\mu\nu,\alpha\beta}/(p^2 - M^2)$ where

$$N_{\mu\nu,\alpha\beta} = \eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha} - \frac{2}{3}\eta_{\mu\nu}\eta_{\alpha\beta} + \frac{p_\mu p_\nu p_\alpha p_\beta}{M^4}\ \text{terms} \tag{62}$$

The first few terms are like that of the massless propagator, but we have a $2/3$ instead of 1, which is know as the VdVZ discontinuity. This theory does not give the correct light bending by the sun, because of this discontinuity. Vainschtein came in and computed a solution in this theory that is spherically symmetric. There is some modification at large scales, and at small distances we have some screening mechanism and make it like massless gravity. But then it was found that there were ghosts and stuff, and the field is a mess.

If we expand the metric around flat space then the action for h looks like

$$M_{\text{pl}}(\partial h)^2 + f^4(h_{\mu\nu}^2 - h^2) + \dots \tag{63}$$

where $M_g^2 = f^4/M_{\text{pl}}^2$, and the interaction is called a Fierz-Pauli structure. The dots is because that this is unlikely to be the whole story. This is not diffeomorphism invariant, but we will do the same trick and add the invariance back. We just do the diffeomorphism

$$\eta_{\mu\nu} \to \partial_\mu Y^\alpha(x)\partial_\nu Y^\beta(x)\eta_{\alpha\beta} \tag{64}$$

Then we know that

$$h_{\mu\nu} \to g_{\mu\nu} - \partial_\mu Y^\alpha \partial_\nu Y^\beta \eta_{\alpha\beta} \tag{65}$$

The unitary gauge corresponds to x^α. The dynamics of Y give the new degrees of freedom that are required. Instead of the unitary gauge we will go to a gauge that

$$Y^\alpha(x) = x^\alpha + \pi^\alpha(x) \tag{66}$$

and we have

$$H_{\mu\nu} = h_{\mu\nu} + \partial_\mu \pi_\nu + \partial_\nu \pi_\mu + \partial_\mu \pi_\alpha \partial_\nu \pi^\alpha \tag{67}$$

where $h_{\mu\nu}$ is the linearized metric in unitary gauge. Now we can put this into the Fierz-Pauli interaction and get

$$\begin{aligned} \mathcal{L} &= M_{\rm pl}^2(\partial h)^2 + f^4(Ah_{\mu\nu}^2 + Bh^2) \\ &= M_{\rm pl}^2(\partial h)^2 + f^2\left[A(\partial_\mu\pi_\nu + \partial_\nu\pi_\mu)^2 + B(\partial_\mu\pi^\mu)^2\right] \end{aligned} \tag{68}$$

Now we can make $\pi_\mu = A_\mu + \partial_\mu\phi$ and introduce a gauge-like redundancy on A and ϕ. Then the interaction part of the Lagrangian becomes

$$A(\partial_\mu\partial_\nu\phi)^2 + B(\nabla\phi)^2 \tag{69}$$

But this is no standard kinetic term, and this give us 2 degrees of freedom on ϕ which is not desirable. So to get rid of one degree of freedom we need to make a choice $A + B = 0$, so that this gives a total derivative. Now the mixing between A and ϕ and h look like

$$f^4\left[h_{\mu\nu}\partial^\mu\partial^\nu\phi - h\Box\phi\right] \to f^4\varphi(\partial_\mu\partial_\nu h^{\mu\nu} - \Box h) \tag{70}$$

This couping looks like that of a scalar field coupled to R, and we can decouple it by a Weyl transformation $h_{\mu\nu} \to h_{\mu\nu} - m_g^2\eta_{\mu\nu}\phi$, and get a standard kinetic term on ϕ. Therefore $h_{\mu\nu}T^{\mu\nu}$ will transform into $h_{\mu\nu}T^{\mu\nu} - m_g^2\phi T$ and we see that this scalar couples to the trace of the energy-momentum tensor. This is also the origin of the $2/3$ factor.

Let's continue this process and look at the self interaction of ϕ field. The interaction looks like $f^2\left[(\partial^2\phi)^3 + (\partial^2\phi)^4 + \dots\right]$, and the leading cubic interaction gives

$$\frac{(\partial^2\phi^c)^3}{\Lambda_5^5}, \quad \text{where} \quad \Lambda_5 = (m_g^4 M_{\rm pl})^{1/5} \tag{71}$$

Now Λ_5 gives a new energy scale. This nonlinear interaction becomes large at large distances. Let's compare it to $(\partial\phi)^2$

$$\left(\frac{M_{\rm pl}R_s}{r^2}\right)^2 \text{ vs } \left(\frac{M_{\rm pl}R_s}{r^3}\right)^3\frac{1}{\Lambda_5^5} \tag{72}$$

Therefore the nonlinear effect becomes important at $r \sim (M_{\rm pl}R_s)^{1/5}\Lambda_5^{-1}$ which is parametrically big. This effect kinetically screens itself at small distances. This classical nonlinear dynamics makes this scalar theory interesting. But exactly because of this interesting nonlinear behavior at low scales, we will see ghosts and problems at this lower scale, which make the solution unstable.

This is a theory for Fierz-Pauli massive gravity. The problem we encountered shows that we can do better. Note that in the Fierz-Pauli interaction, we actually should have all the higher order terms

$$f^4\left[h_{\mu\nu} - h^2\right] + c_3h^3 + c_4h^4 + \dots \tag{73}$$

Because we have $(\partial^2\phi)^3$ in the h^3, we expect the possibility of tuning the parameters to kill all of the $(\partial^2\phi)^n$ interactions. Of course we can carry out the same process as above, but let's do the same using a symmetry argument. The reason that we have $\partial^2\phi$ terms in the interaction is that it is a nonlinear realization of the original diffeomorphism symmetry, and it manifests itself in the form of

$$\phi \to \phi + c, \quad \partial_\mu\phi \to \partial_\mu\phi + b_\mu \tag{74}$$

This symmetry, for obvious reasons, is called Galileon symmetry. Anything like $\partial^2\phi$ is obviously invariant. At this point we can completely forget the starting point and construct theories of ϕ which has the Galileon symmetry. The first theory we would write down would be

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 + \frac{(\partial\phi)^2\Box\phi}{\Lambda_3^3} \tag{75}$$

where $\Lambda_3^3 = m_g^2 M_{\text{pl}}$. It is easy to chech that the interaction term transforms under the Galileon transformation by a total derivative. This effective Lagrangian was seen first in the DGP model of modifying gravity, and there the scalar field ϕ is understood geometrically as the perturbation of the brane.

Galileon symmetry requires that the equation of motion is invariant under it, which means that there should be two and only two derivatives on each field, which avoids the possibility of ghosts. The equation of motion can be thought of as $F(\partial_\mu\partial_\nu\phi) = \partial^\mu X_\mu$ where X_μ is a conserved current under the symmetry. In D dimensions there are D different kinds of possible terms in the Lagrangian.

This theory has some interesting features. One is that at large distances we have modification and at smaller distances this is hidden in an interesting way. There is no ghosts because there are precisely two derivatives on ϕ. This theory has a fatal flaw which is associated with the kind of shift symmetry, which is that we could have superluminal propagation. We can see why there might be a problem, because the leading order theory puts things exactly on the light cone. Because we can shift the derivative of the field to change the speed of propagation such that there can be superluminal propagation. This means that this theory is not local. This can also be seen from the scattering amplitudes, where Galileon symmetry makes the leading order $B(s^2 + t^2 + u^2)$ vanish, whereas basic assumptions of locality in quantum field theory requires that strictly $B > 0$. This is how the nonlocality is encoded in the formalism.