Bogolyubov Institute for Theoretical Physics of National Academy of Sciences of Ukraine
Walter Thirring International Institute for Mathematical Physics, Astrophysics and Nuclear Investigations (Ukraine)

PROCEEDINGS of the Vth Petrov International Symposium "High Energy Physics, Cosmology and Gravity"

29 April–05 May, 2012, BITP, Kyiv, Ukraine

Editor

Stepan S. Moskaliuk

TIMPANI

Supported by the Austrian Academy of Sciences in Vienna, the National Academy of Sciences of Ukraine, the Austro-Ukrainian Institute for Science and Technology, the Slovak Research Centre (Slovakia), the Czech Research Centre (Czech Republic), the Hadronic Press Inc. and the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund

---

PROCEEDINGS of the Vth Petrov International Symposium "High Energy Physics, Cosmology and Gravity" (29 April–05 May, 2012, BITP, Kyiv, Ukraine.– Edited by S. S. Moskaliuk, – Kyiv: TIMPANI, 2012.–306 p.

ISBN 978-966-8904-58-5 "Vth Petrov International Symposium"

Contents

Preface

The fifth Petrov International Symposium on High Energy Physics, Cosmology and Gravity was organized by the Bogolyubov Institute for Theoretical Physics of the National Academy of Sciences of Ukraine and the Walter Thirring International Institute for Mathematical Physics, Astrophysics and Nuclear Investigations (Ukraine); and supported by the Austrian Academy of Sciences, the National Academy of Sciences of Ukraine, the Austro-Ukrainian Institute for Science and Technology, the Slovak Research Centre (Slovakia), the Czech Research Centre (Czech Republic), the Hadronic Press Inc. and the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund. This Symposium is dedicated to the 85th anniversary of the outstanding Austrian physicist Walter Thirring. Walter Thirring international cooperation with N. Bogolyubov Institute for Theoretical Physics of the National Academy of Sciences of Ukraine, in Kyiv, and with the respected Walter Thirring Institute for Mathematical Physics, Astrophysics, and Nuclear Investigations in the Transcarpathian Region of Ukraine had an effect clearly transgressing scientific policy: it helped to establish contacts of Ukrainian scientific institutions with Western European scientific institutions. This international cooperation in the frame of Bogolyubov-Petrov and Thirring-Kummer-Wess scientific schools have helped to guarantee a stimulating atmosphere which continues to attract the bright students which the community of physicists in Europe needs to accomplish its further scientific goals. These Proceedings are limited to the applications of new mathematical methods in High Energy Physics, Cosmology and Gravity. There are based on invited talks given at the forum where scientists and students with different professional backgrounds can discus concepts which are relevant to more than one field, and propose new mathematical methods for solutions of yet unsolved fundamental problems.

This collection is also reprinted in Journal "Algebras, Groups and Geometries" (2012, Vol. 29, issues n. 1–3). And it is recommended to researchers in various areas of High Energy Physics, Cosmology and Gravity, on the one hand, and to graduate and postgraduate students as an introduction into self-consistent modern mathematical methods applications in High Energy Physics, Cosmology and Gravity too.

S. S. Moskaliuk   Kosivska Poliana, December 2012

Alexey Chopovsky[1],
M. Eingorn, A. Zhuk

Spherical compactification of two extra dimensions in Kaluza-Klein geometries: approximate soliton solutions[2]

Odessa National University named after I.I. Mechnikov,
2, Dvoryanskaya street, Odessa 65082, Ukraine
e-mail: alexey.chopovsky@gmail.com

Multidimensional Kaluza-Klein models with toroidal compactification of extra dimensions face a severe problem. It lies in contradiction with the gravitational tests (the perihelion shift, the deflection of light, the time delay of radar echoes and PPN parameters) for a dust-like matter source of the gravitational field. One of the alternative choices of

[1]This research was co-financed by the Austrian Academy of Sciences in the framework of the collaboration with the National Academy of Sciences of Ukraine on Modern Problems in Astroparticle Physics.

[2]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

the further development direction lies in the change of the compactification type. According to this possible choice, we consider two extra dimensions compactified on a two-sphere, representing the curved internal space. In order to provide this curvature, we introduce the background matter in the form of a perfect fluid with the vacuum equation of state in the external space and an arbitrary equation of state in the internal space, as well as a bare multidimensional cosmological constant. Then we perturb the background by the non-dust-like matter source of the gravitational field, possessing tension in the internal space. As a result, we arrive at approximate soliton solutions and impose experimental constraints on their parameters in order to satisfy the gravitational tests.

1 Introduction

The search of a common principle, which describes the phenomenological plurality of the physical world, brought modern theoretical physics into the deep ontological crisis. This crisis is underlain by the impudent infringement of the Occam's principle, according to which "entities must not be multiplied beyond necessity". It is clear that any theoretical model may be called a physical one only relative to a special sphere of reality, where its predictions are experimentally verified or, at least, do not contradict the observational data. Beyond such sphere, the model represents just an abstractive logical construct that is absolutely separated from physical reality. However, nowadays we diagnose the impetuous growth of a tumor of new theoretical essences, while their ontological status remains indefinite. In this situation the problems of revealing of physically inadequate theories become really relevant. As a vivid example we can consider multidimensional theories in general and Kaluza–Klein models in particular.

On the one hand, the significant increase of Kaluza–Klein models popularity in last decades was caused by the well-known problems, which arise in the Standard $SU(3)\times SU(2)\times U(1)$ model, such as the hierarchy problem [1], or the fact that the Standard model does not include gravity. An attempt to solve these problems has been undertaken in supersymmetric models, such as superstring and M-theories

[2]. These theories can be consistently formulated only in the dimensionalities of the fundamental space-time $\mathcal{D} = 10$ and $\mathcal{D} = 11$ correspondingly. In both theories additional spatial dimensions are compactified on the energy scale unattainable within the limits of sub-Planckian physics, or, in other words, they are based on the Kaluza–Klein approach.

On the other hand, the Kaluza–Klein models in their original formulation [3, 4] face a serious problem. As it was shown in [5], the models with toroidal compactification of extra spatial dimensions contradict the experimental data. In particular, the gravitational field of a point-like massive source with dust-like equation of state was considered in this article in the weak field approximation. In General Relativity we can use this approach to derive formulas for such effects, as the Mercury perihelion shift, the deflection of light by the Sun, the frequency shift, and the Shapiro time delay of radar echoes [6].

General Relativity is in brilliant accordance with experiments that check these effects, or, in other words, with classical gravitational tests. Also these tests impose strong restrictions on the numerical values of the so-called PPN parameters β and γ [7–10], which are the coefficients in the Robertson–Eddington expansion of the metrics in powers of a small perturbation parameter $2\varphi/c^2$ in isotropic spherical coordinates r_3, θ, ϕ:

$$ds^2 \approx \left(1 + \frac{2\varphi}{c^2} + \beta\frac{2\varphi^2}{c^4} + ...\right) c^2dt^2 - \\ - \left(1 - \gamma\frac{2\varphi}{c^2} + ...\right) \left(dr_3^2 + r_3^2 d\theta^2 + r_3^2 \sin^2\theta d\phi^2\right), \tag{1}$$

where φ is the gravitational potential. For example, in the first order in perturbation the deflection of light is defined by the expression

$$\delta\phi = (1+\gamma)\,\frac{r_g}{\rho}, \tag{2}$$

where r_g is the gravitational radius. For the perihelion shift per one revolution we have the formula

$$\delta\phi = \frac{1}{3}(2 - \beta + 2\gamma)\,\frac{3\pi r_g}{a\,(1-e^2)}. \tag{3}$$

If a planet (or a satellite) is on the superior conjunction (the far side of the Sun from the Earth), then the formula for the Shapiro time delay is defined by the formula

$$\delta t = (1+\gamma)\,\frac{r_g}{c}\ln\left(\frac{4r_{\oplus}r_{pl}}{R^2_{Sun}}\right). \tag{4}$$

The tightest constraint on the parameter γ comes from the Shapiro time delay experiment using the Cassini spacecraft

$$\gamma = 1 + (2.1 \pm 2.3)\times 10^{-5}. \tag{5}$$

This value is completely consistent with General Relativity, where γ is equal to the unity. Thus, separating the linear in perturbation metric mode in a certain multidimensional model, we can detect the deviation of theoretical predictions from experimental data comparing the obtained value for γ with the unity. It is clear that the significant difference between this numbers points to a certain flaw in the considered model.

Such analysis of Kaluza–Klein models with toroidal topology of additional dimensions has been carried out in [5]. The authors have shown that in the case of three-dimensional external (non-compact) space and dust-like equations of state in both internal and external spaces the following relation fulfills for the parameter γ:

$$\gamma = \frac{1}{D-2}, \tag{6}$$

where D is the total number of spatial dimensions. This result does not depend on the sizes of the extra dimensions. Therefore, point-like gravitating sources are in concordance with experiments only in the three-dimensional space.

The exact soliton solutions of the Einstein equations were investigated in [11, 12]. In these solutions a gravitating source is uniformly smeared over the internal space and its nonrelativistic gravitational potential exactly coincides with the Newtonian one. A new class of solutions, called latent solitons, which are indistinguishable from General Relativity and at the same time are the only objects which satisfy the gravitational experiments at the same level of accuracy as General Relativity, was obtained in [10]. To get these solutions, the matter source must have tension in the internal space instead of the dust-like equation of state, and this is a distinctive feature of these solutions. In particular, black strings and black branes belong to this class. However, the physical meaning of such strange characteristic as tension in the internal space for ordinary astrophysical objects is not clear.

Thus, in the case of toroidal compactification, on the one hand we arrive at the contradiction with the experimental data for the physically reasonable gravitating source in the form of a point-like mass, and on the other hand we have no problem with the experiments for black strings or branes but arrive at very strange equation of state in the internal spaces. How common is this problem for the Kaluza-Klein models? To understand it, we investigate a model with spherical compactification of the internal space.

Our first goal is to get a black brane solution with spherical topology of two extra dimensions.

2 Black brane with spherical compactification

So, to start with, let us consider the six-dimensional static metrics in the form

$$ds^2 = \tilde{A}(\tilde{r}_3)c^2dt^2 + \tilde{B}(\tilde{r}_3)d\tilde{r}_3^2 + \tilde{C}(\tilde{r}_3)\left(d\theta^2 + \sin^2\theta d\phi^2\right) + \\ + \tilde{E}(\tilde{r}_3)\left(d\xi^2 + \sin^2\xi d\eta^2\right), \tag{7}$$

where tilde denotes the "Schwarzschild-like" parameterization for the metrics and the three-dimensional radial coordinate. Similar to the black strings or branes with the flat internal space, here the metric coefficients depend only on the absolute value of the three-dimensional radius-vector. These coefficients can be found with the help of the six-dimensional Einstein equation

$$R_{ik} = \kappa_6\left(T_{ik} - \frac{1}{4}Tg_{ik} - \frac{1}{2}\Lambda_6 g_{ik}\right), \quad \kappa_6 \equiv 2S_5\tilde{G}_6/c^4\,, \tag{8}$$

here Λ_6 is a bare cosmological constant, S_5 is the total solid angle and $\tilde{G}_6$ is the gravitational constant in the six-dimensional space-time. However, in the case of the six-dimensional space-time with spherical compactification of the internal space, we should introduce additional matter which provides the nonzero internal space curvature. Let the components of the energy-momentum tensor of this matter have the form

$$T_{ik} = \begin{cases} \varepsilon(\tilde{r}_3) g_{ik} & \text{for} \quad i = k = 0, \ldots, 3; \\ -\omega_1 \varepsilon(\tilde{r}_3) g_{ik} & \text{for} \quad i = k = 4, 5. \end{cases} \tag{9}$$

Its trace reads $T = 2(2-\omega_1)\varepsilon(\tilde{r}_3)$. In the language of a perfect fluid, we have a vacuum-like equation of state in the external space, but an arbitrary equation of state with the parameter ω_1 in the internal space. Then the Einstein equations reduce to the following system of fundamentally different equations:

$$\frac{R_{00}}{\tilde{A}} = -\frac{1}{4\tilde{A}'\tilde{C}^2\tilde{E}^2}\left(\frac{\tilde{A}'^2\tilde{C}^2\tilde{E}^2}{\tilde{A}\tilde{B}}\right) = \frac{\kappa_6}{2}(\omega_1\varepsilon - \Lambda_6), \tag{10}$$

$$\frac{R_{11}}{\tilde{B}} = -\frac{1}{4\tilde{A}'}\left(\frac{\tilde{A}'^2}{\tilde{A}\tilde{B}}\right) - \frac{1}{2\tilde{C}'}\left(\frac{\tilde{C}'^2}{\tilde{B}\tilde{C}}\right) - \frac{1}{2\tilde{E}'}\left(\frac{\tilde{E}'^2}{\tilde{B}\tilde{E}}\right) = \frac{\kappa_6}{2}(\omega_1\varepsilon - \Lambda_6), \tag{11}$$

$$\frac{R_{22}}{\tilde{C}} = \frac{1}{\tilde{C}} - \frac{1}{4\tilde{C}'\tilde{A}\tilde{C}\tilde{E}^2}\left(\frac{\tilde{C}'^2\tilde{A}\tilde{E}^2}{\tilde{B}}\right) = \frac{\kappa_6}{2}(\omega_1\varepsilon - \Lambda_6), \tag{12}$$

$$\frac{R_{44}}{\tilde{E}} = \frac{1}{\tilde{E}} - \frac{1}{4\tilde{E}'\tilde{A}\tilde{E}\tilde{C}^2}\left(\frac{\tilde{E}'^2\tilde{A}\tilde{C}^2}{\tilde{B}}\right) = -\frac{\kappa_6}{2}[(2+\omega_1)\varepsilon + \Lambda_6], \tag{13}$$

where prime denotes the derivative with respect to the coordinate $\tilde{r}_3$. In the case of black strings or branes with toroidal compactification, the internal space is flat. Now, we require that the internal space is exactly the two-sphere, that is the function $\tilde{E}$ is constant. Therefore, from the equation (13) we get the relation

$$-\frac{1}{a^2} = -\frac{\kappa_6}{2}[(2+\omega_1)\varepsilon + \Lambda_6], \tag{14}$$

which is valid for identically constant ε, which we denote as epsilon with a bar. On the other hand, equations (10), (11) and (12) exactly coincide with the vacuum four-dimensional Schwarzschild equations if the following condition holds:

$$\bar{\varepsilon} = \Lambda_6/\omega_1 . \tag{15}$$

Substituting this value of the background matter energy density into the equation (14), we get the relation

$$\bar{\varepsilon} = \frac{1}{(1+\omega_1)\kappa_6 a^2}, \tag{16}$$

The obtained equalities allows to conclude that because of positiveness of the background energy density $\omega_1 > -1$. The parameter ω_1, which determines the state in the internal state, is not fixed and takes part in fine-tuning between $\bar{\varepsilon}$ and Λ_6. Choosing different values of this parameter (with the vacuum-like equation of state in the external space), we can simulate different forms of matter. In particular, $\omega_1 = 1$ and $\omega_1 = 2$ correspond to the monopole form-fields (the Freund-Rubin scheme of compactification) and the Casimir effect, respectively. As an example, let's consider the case of the Freund-Rubin stable compactification with two-forms

$$F_{ik} = \begin{cases} \sqrt{g_2}\,\varepsilon_{ik} f & \text{for } i = k = 4,\ 5; \\ 0 & \text{otherwise}; \end{cases} \tag{17}$$

where g_2 is the determinant of the metrics on the internal sphere, ε_{ik} is the totally antisymmetric Levi-Civita tensor and f is a constant which we define below.

Hence, the energy-momentum tensor is determined by

$$T_{ik} = \begin{cases} \dfrac{f^2}{8\pi}\cdot g_{ik}, & \text{for} \quad i,\ k = 0,\ \ldots,\ 3; \\ -\dfrac{f^2}{8\pi}\cdot g_{ik}, & \text{for} \quad i,\ k = 4,\ 5. \end{cases} \tag{18}$$

The comparison of this expression with the background energy-momentum tensor shows that the parameter of the equation of state in the internal space ω_1 is equal to the unity. Similarly, we can consider the stabilization by means of the Casimir effect where $\omega_1 = 2$. It is also worth noting that in the case of the zero cosmological constant, the parameter ω_1 should also be equal to zero. Therefore, the homogeneous matter with the received energy-momentum tensor provides spherical compactification of the internal space.

In the usual four-dimensional space-time, the Schwarzschild metrics is created by a compact (for example, point-like) spherically symmetric gravitating matter source. Thus, to get the external spacetime in the form of the Schwarzschild metrics, we have to introduce such object which is spherically symmetric in the external space and uniformly smeared over the internal space. Let the energy-momentum tensor of this perturbation have the following nonzero components:

$$\begin{array}{rcl} \hat{T}_{00} & = & \hat{\varepsilon} g_{00}, \quad \hat{T}_{\alpha\alpha} = 0\,, \ \alpha = 1,2,3 \\ \hat{T}_{44} & = & -\hat{p}_1 g_{44}\,, \quad \hat{T}_{55} = -\hat{p}_1 g_{55}\,. \end{array} \tag{19}$$

Note that 44 and 55 components of the energy-momentum tensor are generally nonzero. Then the total energy-momentum tensor is the superposition of the background one and the energy-momentum tensor of the perturbation. In the weak-field limit we can suppose that the energy density is approximately $\hat{\rho}c^2$, where rho with a hat is the multidimensional rest mass density, and for the particle uniformly smeared over the internal space, multidimensional and three-dimensional rest mass densities are proportional to each other, namely, they're connected by the relation

$$\hat{\rho} = \frac{\hat{\rho}_3}{V_2}, \tag{20}$$

where $V2$ is the volume of the internal sphere. Also in the case of a pointlike gravitating mass the three-dimensional rest mass density is proportional to the delta-function of the position vector in the external space.

Now it is the crucial point of our reasoning.

Taking into account only the gravitating matter source and keeping in mind that we want to get the Schwarzschild solution in the external space, it can be easily realized that the only non-zero components of the Ricci tensor should have the following form:

$$R_{00} = \frac{1}{2}\kappa_6 \hat{\varepsilon} g_{00} \approx \frac{1}{2}\kappa_N \hat{\rho}_3 c^2 g_{00}\,, \tag{21}$$

$$R_{\alpha\alpha} = -\frac{1}{2}\kappa_6 \hat{\varepsilon} g_{\alpha\alpha} \approx -\frac{1}{2}\kappa_N \hat{\rho}_3 c^2 g_{\alpha\alpha},\ \alpha = 1\,,2\,,3\,, \tag{22}$$

$$R_{44} = 1, \quad R_{55} = \sin^2 \xi\,, \tag{23}$$

where $\kappa_6/V_2 = \kappa_N \equiv \dfrac{8\pi G_N}{c^4}$, G_N is the Newton's gravitational constant. And now, substituting these components of the Ricci tensor as well as the components of the total energy-momentum tensor of the perturbed system in the Einstein–Hilbert equations, one can see that these equations are compatible only if the following equation of state holds:

$$\hat{p}_1 = -\frac{1}{2}\hat{\varepsilon} \tag{24}$$

For example, the 00-component of the Einstein equation is given by (10), and we see that the left-hand side of the equation identically coinsides with the right-hand side only if the source has the equation of state (24):

$$R_{00} = \frac{1}{2}\kappa_6 \hat{\varepsilon} g_{00} = \kappa_6 \left[\hat{\varepsilon} - \frac{1}{4}(\hat{\varepsilon} - 2\hat{p}_1)\right] g_{00} \tag{25}$$

Similarly, all other nontrivial components also give the same equation of state. That is the gravitating matter source should have tension in the internal space as it takes place for the black strings or branes with toroidal compactification.

Therefore, the required exact solution of the field equations, which is called in the considered case **the black brane** with spherical compactification, is presented here:

$$ds^2 = \left(1 - \frac{r_g}{\tilde{r}_3}\right) c^2 dt^2 - \left(1 - \frac{r_g}{\tilde{r}_3}\right)^{-1} d\tilde{r}_3^2 - \tilde{r}_3^2 d\Omega_2^2 - a^2 \left(d\xi^2 + \sin^2 \xi d\eta^2\right) , \tag{26}$$

So, the matter source of this black brane consists of two parts. First, it is the homogeneous component with fine-tuning conditions, which provides spherical compactification of the internal space. Second, it is the gravitating object which is spherically symmetric and compact in the external space and uniformly smeared over the internal space. It has negative pressure in the extra dimensions. This component provides the Schwarzschild-like metrics in the external spacetime.

To calculate formulas for the famous gravitational experiments or expressions for parameterized post-Newtonian (PPN) parameters, it is usually convenient to rewrite the metrics in isotropic (with respect to our three-dimensional space) coordinates. The Schwarzschild-like radial coordinate and the isotropic radial coordinate are connected by the relation:

$$\tilde{r}_3 = r_3 \left(1 + \frac{r_g}{4r_3}\right)^2 . \tag{27}$$

For example, in isotropic coordinates the linear in perturbation expression for the metrics, similar to the Eddington–Robertson expansion in General Relativity, is given by the approximate equality:

$$ds^2 \approx \left(1 + \frac{2\varphi_N}{c^2}\right) c^2 dt^2 - \left(1 - \frac{2\varphi_N}{c^2}\right) \left(dx^2 + dy^2 + dz^2\right) -$$
$$-a^2 \left(d\xi^2 + \sin^2 \xi d\eta^2\right). \tag{28}$$

.

This equality shows, that the PPN-parameter $\gamma = 1$. It's also not difficult to demonstrate, that the PPN parameter β is also equal to the unity, similar to General Relativity. Therefore, our black brane satisfies the gravitational experiments at the same level of accuracy as General Relativity.

3 Approximate soliton solutions

Now we shall consider the other problem. It arises from the following question: are the objects, which provide the black brane metrics, the only sources which satisfy the gravitational experiments in the Kaluza–Klein models with spherical topology of additional dimensions? To get the answer, let's note, that in General Relativity, the weak-field limit is a good approximation to calculate the above-mentioned gravitational experiments. In this limit, a gravitating massive body (e.g., a point-like mass) has dust-like equations of state. Obviously, the physical sense of such approach should be preserved in multidimensionality, and it's natural to generalize this approach to our model.

Let us investigate the most general case, where, instead of the dust-like equations of state in all spatial dimensions, we suppose the following energy-momentum tensor of the perturbation:

$$\begin{gathered} \hat{T}_{00} \approx \hat{\rho} c^2, \quad \hat{T}_{\alpha\alpha} = 0, \quad \alpha = 1,2,3 \\ \hat{T}_{44} \approx \Omega \hat{\rho} c^2 a^2, \quad \hat{T}_{55} \approx \Omega \hat{\rho} c^2 a^2 \sin^2 \xi. \end{gathered} \tag{29}$$

Here Ω denotes a certain parameter. Concerning the energy-momentum tensor of the background matter, we suppose that perturbation does not change the equations of state in the external and internal spaces, i.e. $\bar{\varepsilon}$ and ω_1 are constant and still satisfy the fine-tuning conditions (15). For example, if we had a monopole form-fields (corresponding to $\omega_1 = 1$) before the perturbation, the same type of matter we shall have after the perturbation. Therefore, the energy-momentum tensor of the perturbed background is defined by

$$\tilde{T}_{ik} \approx \begin{cases} \left(\bar{\varepsilon} + \varepsilon^1\right) g_{ik}, & i, k = 0, ..., 3; \\ -\omega_1 \left(\bar{\varepsilon} + \varepsilon^1\right) g_{ik}, & i, k = 4, 5, \end{cases} \tag{30}$$

where the correction ε^1 is of the same order of magnitude as the perturbation. Further we shall see that existence of such correction provides the field equations consistency. The total energy-momentum tensor is the superposition of the corresponding tensors of the perturbed background and the perturbation: $T_{ik} = \tilde{T}_{ik} + \hat{T}_{ik}$.

In the case of uniformly smeared (over the internal space) perturbation, the perturbed metrics preserves its diagonal form and in isotropic coordinates is given by formulas

$$ds^2 = Ac^2dt^2 + Bdx^2 + Cdy^2 + Ddz^2 + Ed\xi^2 + Fd\eta^2, \tag{31}$$

$$\begin{array}{ll} A \approx 1 + A^1(r_3), & B \approx -1 + B^1(r_3), \\ C \approx -1 + C^1(r_3), & D \approx -1 + D^1(r_3), \\ E \approx -a^2 + E^1(r_3), & F \approx -a^2 \sin^2 \xi + F^1(r_3), \end{array} \tag{32}$$

where we take into account the spherical symmetry of the perturbation with respect to the external space. All the terms indexed by the unity are of the order of perturbation. To find these coefficients, we should solve the Einstein–Hilbert equation, which is reduced now to the system of linear equations:

$$\triangle_3 A^1 = \kappa_6 \omega_1 \varepsilon^1 + \left(\frac{3}{2} + \Omega\right) \kappa_6 \hat{\rho} c^2, \tag{33}$$

$$\triangle_3 B^1 = \triangle_3 C^1 = \triangle_3 D^1 = -\kappa_6 \omega_1 \varepsilon^1 + \left(\frac{1}{2} - \Omega\right) \kappa_6 \hat{\rho} c^2, \tag{34}$$

$$\triangle_3 E^1 = (2 + \omega_1) \kappa_6 a^2 \varepsilon^1 - \frac{2}{a^2} E^1 + \left(\frac{1}{2} + \Omega\right) \kappa_6 \hat{\rho} c^2 a^2, \tag{35}$$

where the triangle is the three-dimensional Laplace operator. It is very helpful to analyze the non-diagonal components of the field equation. Using the geometric properties of the perturbation (namely, the spherical symmetry with respect to the external space and the uniform smearing over the internal sphere) we also obtain the relations

$$F^1 = E^1 \sin^2 \xi, \quad \triangle_3 E^1 = \frac{a^2}{2} \left(\triangle_3 A^1 - \triangle_3 B^1\right) =$$
$$= \frac{a^2}{2} \left[2\kappa_6 \omega_1 \varepsilon^1 + (1 + 2\Omega)\, \kappa_6 \hat{\rho} c^2\right]. \tag{36}$$

The comparison of (35) and (36) gives

$$\kappa_6 \varepsilon^1 = \frac{E^1}{a^4}. \tag{37}$$

Hence the introduction of the background energy-momentum tensor perturbation is totally legitimate, because only in the case of nonzero ε^1 the system of linearized field equations is consistent in the general case. The substitution of the relation (37) back into (36) gives the Helmholtz equation:

$$\triangle_3 E^1 - \frac{\omega_1}{a^2} E^1 = \left(\frac{1}{2} + \Omega\right) \kappa_6 \hat{\rho} c^2 a^2 = \left(\frac{1}{2} + \Omega\right) \frac{8\pi G_N}{c^2} a^2 m \triangle(r_3), \tag{38}$$

where for the smeared extra dimensions the perturbation rest mass density is proportional to the delta-function of the position vector in the external space. If the parameter $\Omega \neq -1/2$, then the negative value of ω_1 results in the nonphysical oscillating solution. Hence, in the case of positive ω_1, the solution of this equation is given by

$$E^1 = a^2 \frac{\varphi_N}{c^2} (1 + 2\Omega)\, e^{-r_3/\lambda}, \quad \lambda = a \Big/ \sqrt{\omega^1}\,, \tag{39}$$

φ_N denotes the Newton's potential of a perturbation. It's easy now to obtain the metric corrections A^1 and B^1:

$$A^1 = \frac{2\varphi_N}{c^2} + \frac{E^1}{a^2} = \frac{2\varphi_N}{c^2} \left[1 + \left(\frac{1}{2} + \Omega\right) \exp\left(-r_3/\lambda\right)\right], \tag{40}$$

$$B^1 = \frac{2\varphi_N}{c^2} - \frac{E^1}{a^2} = \frac{2\varphi_N}{c^2}\left[1 - \left(\frac{1}{2} + \Omega\right)\exp\left(-r_3/\lambda\right)\right]. \tag{41}$$

Thus, to get agreement with gravitational experiments, coefficients A^1 and B^1 should be very close to each other. In General Relativity, A^1 is exactly equal to B^1. In our model, we can satisfy this condition in two cases.

First case: $\Omega = -1/2$. Obviously, this is the case of the previous problem with the black brane, and we just reproduce this exact solution in the weak-field limit. Here, the parameter ω_1 is not fixed and satisfies the condition $\omega_1 > -1$ including the case of the zero ω_1, when a bare cosmological constant is absent.

Second case: $\Omega \neq -1/2$, $r_3 \gg \lambda$. Here, the metrics asymptotically approaches to (28), including the physically reasonable case of the dust-like equation of state $\Omega = 0$. Therefore, the second case is called the asymptotic black brane. The positiveness of the state parameter ω^1 is the necessary condition of the considered case.

The metric correction term A^1 describes the nonrelativistic gravitational potential: $A^1 = 2\varphi/c^2$. Therefore, this potential acquires the Yukawa correction term. The Yukawa interaction is characterized by two parameters: the parameter λ, which defines the characteristic range of this interaction, and the parameter α in front of the exponential function. In our case $\alpha = 1/2 + \Omega$. There exists a strong restriction on these parameters from the inverse square law. If, for example, omega is not equal to $-1/2$, the upper limit for λ is given by the relation

$$\lambda_{\max} = (a/\sqrt{\omega_1}) \approx 6 \times 10^{-3}\mathrm{cm}. \tag{42}$$

In view of this relation we have also a possibility to estimate the upper limit of the size of the internal space for a fixed value of the state parameter. Let us estimate now the Yukawa correction term for the gravitational experiments in the Solar system. We can take astrophysical external distances, for example, comparable with the radius of the Sun. Therefore, with very high accuracy we can drop the Yukawa correction term, and arrive at the case of the asymptotic black brane.

4 Conclusions

Thus, now let us summarize all the results of the present work in the form of a short conclusion. In this work we found a metrics for a black brane with spherical compactification of the internal space. This is the exact solution of the Einstein equations. To get such solution, we should first prepare the corresponding background with the flat external space-time and the curved internal space (the two-sphere). For this purpose, we should include a matter source in the form of a homogeneous perfect fluid with the vacuum equation of state in the external (our) space and an arbitrary equation of state in the internal space. The model can also contain a bare multidimensional cosmological constant. To get spherical compactification, parameters of the perfect fluid should be fine-tuned. The presence of such perfect fluid is the main difference from the well-known black branes with toroidal compactification. In the latter case we do not need to introduce an additional perfect fluid, because the background here is flat for both external and internal spaces.

The next step is to construct a Schwarzschild-like metrics in the external space-time. To perform it, we included a gravitating object which is spherically symmetric and compact in the external space as well as uniformly smeared over the internal space. We have shown that the Einstein equations are compatible only if this object has negative pressure (i.e. tension) in the internal space. It should be noted that the gravitating matter source for black branes with toroidal compactification has precisely the same equation of state in the internal space.

Then, we generalized our investigations to the case where the background with spherical compactification is perturbed by a matter source which has the dust-like equation of state in the external space and an arbitrary equation of state in the internal space. In the weak-field limit, we found solutions of the linearized Einstein equations. One case of the parameter choice reproduces the weak-field limit of the exact solution. In the other case the metric coefficients acquire the Yukawa correction terms which are negligibly small at three-dimensional distances much greater than the characteristic range of the Yukawa interaction. At these distances, the metrics asymptotically tends to the weak-field limit of the exact black brane solution. We named the second case the asymptotic black brane. Obviously, in the case of spherical compactification, the exact black branes and the asymptotic black branes satisfy the gravitational experiments at the same level of accuracy as General Relativity. Hence, we have two theoretical possibilities to satisfy observational restrictions, but we still don't know which of them is closer to physical reality.

References

1 N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Rev. D **59**, 086004 (1999); arXiv:hep-ph/9807344.

2 J. Polchinski, *String Theory, Volume 2: Superstring Theory and Beyond* (Cambridge University Press, Cambridge, 1998).

3 Th. Kaluza, *Zum Unitätsproblem der Physik*, Sitzungsber. d. Preuss. Akad. d. Wiss., 966 (1921).

4 O. Klein, *Quantentheorie and fünfdimensionale Relativitätstheorie*, Zeitschrift fr Physik **37**, 895 (1926).

5 M. Eingorn and A. Zhuk, Class. Quant. Grav. **27**, 205014 (2010); arXiv:gr-qc/1003.5690.

6 C.M. Will, *Was Einstein Right? Testing Relativity at the Century* (*100 Years of Relativity: Spacetime Structure – Einstein and Beyond*, World Scientific, Singapore, 2005); arXiv:gr-qc/0504086.

7 N. Straumann, *General Relativity and Relativistic Astrophysics* (Springer-Verlag, Berlin, Heidelberg, 1984).

8 C.M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge, 2000).

9 Hooft, Gerard âĂŸt *Introduction to General Relativity*, (Princeton: Rinton Press Inc., 2002).

10 Itzykson, C., and Zuber, J.B. (1980) Quantum Field Theory, New York: McGraw-Hill.

11 M. Eingorn and A. Zhuk, Phys. Rev. D **83**, 044005 (2011); arXiv:gr-qc/1010.5740.

12 M. Eingorn, O.R. de Medeiros, L.C.B. Crispino and A. Zhuk, Phys. Rev. D **84**, 024031 (2011); arXiv:gr-qc/1101.3910.

György Darvas

Finsler geometry in the presence of isotopic field charges applied for gravity[3]

Institute for Research Organisation of the Hungarian Academy of Sciences,
18, Nádor St., Budapest, H-1051 Hungary
e-mail: darvasg@iif.hu

The paper specifies the conservation of the isotopic field-charge spin on the gravitational interaction, and discusses one of the consequences. First, the isotopic field-charges of the gravitational field will be defined, followed by a short presentation how the conservation of the isotopic field-charge spin has been derived. It will be shown that in the presence of a kinetic gauge field the metric of the gravitational field and its curvature should follow a Finsler geometry, that means in the presence of an isotopic mass field, the metric and the curvature depend also on velocity. In particular, the $g_{\mu\nu}$ metric tensor, and consequently the affine connection field and the curvature tensor formed from its derivatives, depend on space-time plus velocity co-ordinates. We insert this metric in the formula of the affine connection field, and the Ricci tensor.

[3]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

These extended formulas will be applied to one side of the Einstein equation, while on the other side there appears the stress-energy tensor specified in the presence of an isotopic mass field. We conclude that due to the velocity dependence, the Schwarzschild solution cannot be applied! It must be replaced by a Finslerian solution. Finally, as an example, we will predict that this solution may give a more accurate calculation for the precession of the perihelion of Mercury. General Theory of Relativity was based in a significant part on the equivalence principle. This principle states equivalence between the mass of gravity and the mass of inertia. At the same time, this statement declares that gravitational and inertial masses are not identical, for equivalence can be observed between non-identical physical properties. Papers [1]-[12] dealt in detail with this ambiguity. It is analysed in [13]. In [14]-[15] is formulated an assumption that the field charges of the gravitational field – gravitational and inertial are not only equivalent in their measured quantity, there exists an invariance between them. This means, that there is a symmetry according to which they are interchangeable. They can be considered as macroscopically indistinguishable physical properties of matter, which behave as isotopes of each other. Since they are qualitatively not identical, we have the right to distinguish them in our physical equations, although we can calculate with them like with equivalent value quantities. As shown in [13]-[15], mass of gravity is associated with the potential part V of an object's Hamiltonian and mass of inertia is associated with the kinetic part T of an object's Hamiltonian. According to this observation we can call the mass of gravity as (scalar) potential mass, and the mass of inertia as kinetic mass. For kinetic mass is associated with (and, according to STR, depending on) the velocity of a massive object in a given reference frame, it can be described in a velocity dependent (kinetic) field, while gravitational mass belongs to a (scalar) field depending solely on the space-time co-ordinates. [16]-[17] proved the mathematical existence of a gauge invariance in a velocity dependent gauge field. This mathematical derivation led to two conserved Noether currents that exist simultaneously. This result [14]-[15] predicted at first, a conserved

quantity - called isotopic field charge spin - and, at second, the exchange of two gauge quanta (bosons) between interacting mass units. One of them can be identified with the earlier assumed graviton. The other boson - let's call it 'dion' [14] - which appears in the equations due to the consideration of a velocity dependent gauge field, is new, and is a consequence of the conservation of the isotopic field charge spin. The paper presents how did the equivalence principle applied in GTR lead to the assumption of the isotopic field charge spin and its conservation, and to the prediction of an additional boson exchange. The invariance between isotopic field charges in the presence of a velocity dependent gauge field, and the conservation of the isotopic field charge spin were extended to the field charges of other physical interaction fields [14]-[15], since the mathematical proof [16]-[17] allowed general interacting kinetic gauge fields. This predicted the exchange of additional gauge bosons in electroweak and strong interactions as well. The result is part of the 'new physics' expected for many years in high energy physics [18]-[22], and is a candidate to replace the SUSY assumption. The difference between SUSY and the isotopic field charge spin assumption is that the former renders fermion-boson pairs as newborn brothers to each other, while the latter does fermion-fermion and boson-boson twins. There are only the boson twins new and to be discovered, since the twin brothers of fermions originate in splitting the existing ones and are assumed to be identified with the long 'known' pairs defined by the equivalence principle.

1 INTRODUCTION

This paper treats fundamental physical interactions starting from two preliminary assumptions.

i) Although *mass of gravity* and *mass of inertia* are *equivalent* quantities in their measured values, they are *qualitatively not identical* physical entities. We take into consideration this difference in our equations. Then this *'equivalence is not identity' principle* is extended to sources of further fundamental interaction fields, other than gravity.

ii) *Physical interactions occur between these* qualitatively different *entities.*

These two assumptions do not contradict to any known physical theory, however, they allow another interpretation of facts built in our explanations of physical experience. Based on them we demonstrate the existence of an invariance between the two isotopic forms of the field charges, and formulate certain consequences in our view on the physical structure of matter.

2 THE NOTION OF ISOTOPIC FIELD CHARGES AND THEIR DYNAMICS

2.1 Equivalence does not mean identity

In a strict sense, *identical objects cannot be equivalent.* Only *qualitatively different* objects can be compared to conclude a quantitative equivalence between them. Equivalence always presumes the existence of at least one property, in which the compared objects differ. (Isotopic spin is a good example how to avoid ambivalence.)

The equivalence principle is one of the main pillars of the general theory of relativity (GTR). It states the equivalence of the gravitational and inertial masses. Let's consider the mass of gravity and the mass of inertia as two different properties of matter. For the same massive object can behave once as a source of gravity, then as a measure of inertia, we will imagine them as two isotopic states of the same property, called mass of the object.

As much as the mass is the field charge of the gravitational field, we will call its two isotopic states as *isotopic field charges* for the *gravitational interaction.* The gravitational mass is associated with the (scalar) potential part of that interaction, while the inertial mass with the kinetic part. In GTR the latter is attributed to the momentum densities, while the former is associated with the gravitational field energy. They are separated within the stress-energy tensor ($T_{\mu\nu}$), but according to the general relativity principle they can be transformed into each other; - we should add, at least in their quantitatively equivalent values. GTR does not make any statement about the qualitative transformation of the two kinds of masses into each other. This was a reason to identify them. The need for a qualitative transformation simply has not emerged. Nevertheless, we show that it cannot be avoided. So, we introduce distinction between masses of gravity and inertia in our equations. (In a similar way, the electric charge - i.e., the source of the electromagnetic field - is the field charge of the electromagnetic interaction; flavour and lepton charge - are the sources of the weak field; the colour charge - i.e., the source of the strong field - is the field charge of the strong interaction.) The sources - field charges - are assumed to be realised in the matter field, while they serve as sources for gauge fields. Are they really the same, or can one distinguish the two agents? The mass of gravity and the mass of inertia are considered as two equivalent quantity *isotopic states* of the field charge of the gravitational field. They represent two different qualities. Their concepts express two properties of matter, whose existence originates in different experiences. Physics established quantitative relations between them (i.e., equal values), however this fact does not vanish their qualitative difference. We argue that we have all reason to make distinction between them in our theories.

When we introduce the two isotopic filed charges in our equations, they will destroy certain symmetries of those equations. This contradicts to our experience. Therefore, there must be an invariance that compensates and restores the spoiled symmetry. To avoid the contradiction between experience and theory, we assume that the two kinds of charges of the gravitational field, should be transformed into each other by a gauge transformation. Such a gauge transformation should involve the existence of a conserved property that we define in the following way.

Since the required transformation affects the *isotopic states* of the individual *field-charges* (we mark it with ד ['dalet' the fourth letter of the Hebrew alphabet]), this transformation must be performed in a special gauge field; and

since these states can occupy two positions in that gauge field, it must be a *spin-like property*, therefore, we will call this property as *Isotopic Field-Charge Spin (IFCS)* and denote it by Δ, and we will refer to the invariance transformation what we are seeking for as *isotopic field-charge gauge transformation*. This assumption assumes the existence of a local gauge field, in which the isotopic field-charge spin can rotate and occupy two states and concludes a conserved (non-Abelian) current and a corresponding class of $SU(2)$ type invariances.

For the same object can behave, e.g., in the gravitational field, once as the source of a gravitational force, and in another frame of reference as a source of a (kinetic) inertial force (cf., covariance principle), they must be able to get transformed into each other. Non-Abelian character and arbitrariness involve that the orientation of the isotopic field-charge spin is of no physical significance. If we determine the proper form of this invariance transformation, it will counteract the loss of symmetry between the two kinds of field-charges, and bring our equations in compliance with the experimental observations.

The required invariance shows certain formal similarities to YM-type invariances [23]-[24]. However, it must differ from them in at least two features. Once, the concerned physical property, namely the isotopic field charge (IFC, $\daleth$), is a quite different physical property than the isotopic states of nucleons. Secondly, the gauge field, and consequently the gauge transformation that rotates the isotopic field charge spin (IFCS, $\varDelta$) in this gauge field, are quite different from the isotopic gauge field derived for the isotopic spin transformation. (For specification, see section 2.)

The existence of such an invariance transformation provides us with a symmetry, and consequently with a conservation law, with the conservation of the introduced new property ($\varDelta$) of the field-charges. The conservation of isotopic field-charge spin is identical with the requirement of invariance of all interactions under isotopic field-charge spin rotation (in the gauge field where it is interpreted). Accordingly, all physical interactions should be invariant under a transformation in a specific gauge field, more precisely, under a rotation of the property, called isotopic field-charge spin ($\varDelta$). [15]-[17] proved that invariance transformation.

2.2 Interaction between the isotopic field charges

When we take a measure on an object, we have no experience that we found it in one or the opposite isotopic state. Would we observe a single particle, it were either in one or in the other IFCS state. We can call the two states as potential and kinetic, scalar and vector, or bound and free states. However, our measurement records a mixture of the two states. Nevertheless, we do not observe the individual IFCS states. Our observation suggests that they behave as being in both states, each measured object can occupy both a potential (bound) and a kinetic (free) IFCS state. In the lack of experience to catch a particle in one or the other stable state, we have good reason to assume that they permanently change their states. (Randomly or with a stable frequency, they may probably follow a similar mechanism like quarks do during their colour change via gluon exchange).

Let us consider a model of a doublet, when a particle can be in a potential state (V) and in a kinetic state (T). According to its actual state it has potential or kinetic energy respectively. According to our observation all particles possess both. We can interpret the phenomenon in the following way: In a *probabilistic model* we can consider that the wave function of the given particle may be in a potential state with amplitude ψ_T, or in a kinetic state with amplitude ψ_T. We detect a probabilistic mixture in a measurement. In a large set of particles (e.g., in the case of a massive body consisting of many particles) the probabilities reach a stable proportion and we observe stabilised measurable potential and kinetic energies in a given reference frame. A *harmonic oscillator model* presumes the permanent change of a single particle between its two isotopic field charge states. A particle *in a potential state* plays the role of the *source of a scalar field.* Therefore a potential isotopic field charge (we denote by $\urcorner_V$) is a scalar quantity. A particle *in a kinetic state* serves as a *current source of a vector field.* So a kinetic isotopic field charge (we denote by $\urcorner_T$) plays roles in three vector components according to three, directed, independent components of a field charge current. An important consequence of the switch between the two IFC states is that the isotopic field charges must commute between a scalar and three components of a vector quantity, according to the velocity components of the kinetic state in the given reference frame.

2.3 Isotopic field charges in the gravitational field

As a consequence of the distinction between m_V and m_T, as well as the association of the energy content with the mass m_V and the components of the momentum with m_T, we lose also the symmetry of the $T_{\mu\nu}$ energy-momentum tensor. To retain symmetry in Einstein's field equations we must require again the invariant transformation of m_V and m_T into each other in an appropriate gauge field. We refer to Mills [24] who foresaw the possible generalisation of YM type gauge invariance in general relativity "in close analogy with the curvature tensor". If we consider the energy-momentum tensor (in which both isotopic states of mass appear) as the source of the gravitational field, then - in the usual way - a scalar and a vector potential can be separated. (A hypothetic vector potential is justified by a non-static effect, e.g., acceleration, in the field.) Although, unlike QED, there is no analogy with the meaning of a vector potential of the electromagnetic field, the consideration of the kinetic (inertial) mass as an individual physical property against the gravitational mass may lend certain meaning to a gravitational vector potential. We can explain this so, that m_4 in T_{44} does not compose a fourth component of a four-vector in the classical theory of gravitation where there is a single scalar mass, while if we consider now $m_4 = m_V$, the three components of the kinetic mass m_T can compose a three-vector, however T_{i4} will not form a four vector either.

To maintain the Lorentz invariance of our physical equations in the gravitational field, we must demand to restore the invariance of $\begin{pmatrix} \vec{m}_T \\ m_V \end{pmatrix}$ under an *additional transformation* that should *counteract the loss of symmetry caused by the introduction of two isotopic states of mass.* We discuss that transformation in section 2. Further, in the case of gravitation the relation of the scalar and the vector fields are not linear even if we have not made distinction between the potential and kinetic masses. The non-linearity is coded in the relation of the tensors [25] at the left side of the Einstein equation (in units $c = 1$),

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}$$

or $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}$ where the Einstein tensor is defined as $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$ whose covariant derivative must vanish.

Since our $T_{\mu\nu}$ tensor on the right side has already lost its symmetry, we can take $\Lambda g_{\mu\nu}$ into account within a modified $T'_{\mu\nu}$ - handling the gravitational and kinetic masses in it together with the dark energy - and we get the following formally symmetric equation:

$$G_{\mu\nu} = 8\pi G T'_{\mu\nu} \,.$$

(The disadvantage of this apparently quasi-symmetric form is that the metric tensor $g_{\mu\nu}$ appears in the expressions at both sides of the equation.) It is only our enigmatic hope that the asymmetry hidden inside $T'_{\mu\nu}$ will be restored with the conservation of the IFCS for the isotopic gravitational field charges together with the dark energy. Nevertheless, even if the latter fails, the symmetry of the energy-momentum tensor can be saved by the invariant gauge transformation of the IFCS. The most important analogy is between the behaviour of the potential and the kinetic field charges of the individual fields that makes probable to conjecture that a unique transformation will assure their invariance (cf., section 2).[4]

(See in details in sections 3-4.)

3 CONSERVATION OF THE ISOTOPIC FIELD CHARGE SPIN

Distorted symmetry of our equations[5] - what is not in accordance with experience - can be restored by proving that there exists an invariance between the twin brothers of the field charges (sources of the fields) split according to the introduced new property (Δ). Invariance means that particles, disposed with these properties, can be exchanged. The "exchange rate" (gauge) depends on the velocity of the kinetic field charge compared to the respective matter field (i.e., to the scalar potential field charge in rest in that field). The validity of the assumption can be verified by demonstrating the existence of the gauge bosons that mediate the

[4]We must add to the conjecture of the "unique" transformation a few remarks. As [26] stated, "In contrast to the symmetry or invariance requirement in STR, the principle in GTR is most often presented as strictly speaking a covariance requirement." Gauge theories behave like GTR, at least in this respect. General covariance "is not tied to any geometrical regularity of the underlying spacetime, but rather the form invariance (covariance) of laws under arbitrary smooth coordinate transformations" [26, p. 34]. Weyl [27] found that the more general geometry resulting from admitting local changes called gauges described not only gravity but also electromagnetism. He showed also that the conservation laws of Noether follow in two distinct ways in theories with local symmetries. This led to the Bianchi identities, which hold between the coupled equations of motion, and which are due to the local gauge invariance of action. Later [28] demonstrated that the conservation of the electric charge followed from the local gauge invariance in the same way as does energy-momentum conservation from co-ordinate invariance in GTR.

[5]According to Higgs [29]: "The idea that the apparently approximate nature of the internal symmetries of elementary-particle physics is the result of asymmetries in the stable solutions of exactly symmetric dynamical equations, rather than an indication of asymmetry in the equations themselves, is an attractive one." Please, compare this notice with Wigners concern [30]!

exchange. This invariance as soon as proven means a new symmetry principle of nature. This perspective is challenging!

Section 2.2 below presents the main lines of the mathematical proof [17] of such invariance. The demonstration of the predicted gauge bosons is left to the experience.

3.1 Velocity dependent phenomena

We know certain phenomena in classical physics that depend on velocity in a given reference frame. As examples, there can be mentioned first the kinetic energy, then the Lorentz force, and the covariant effect of the Lorentz transformation $[(x^\mu)' = \Lambda^\mu_\nu(\nu)x^\nu$ for space-time vectors, and $(F^{\mu\nu})' = \Lambda^\mu_\alpha(\nu)\Lambda^\nu_\beta(\nu)F^{\alpha\beta}$ for the electromagnetic field tensor]. Descriptions of the mentioned phenomena handle the space-time co-ordinates as indirect variables. The Lorentz invariance depends only on the velocity difference between the compared systems. In general, kinetic quantities depend first on velocity in the chosen reference frame, and only indirectly, through $\nu = \nu(x_i, t)$ on the space-time variables. As [24] observed, "Hamilton's principle was first discovered in connection with mechanical systems, where the Lagrangian turns out to be the difference between the kinetic and potential energies, but the principle is easily extended to include velocity-dependent forces of certain types", including, e.g., the magnetic force on a moving, electrically charged particle.

It is not surprising that phenomena related solely to the kinetic part of the Hamiltonian (T) can be described in a velocity dependent, i.e., kinetic field $D_T = D[\nu(x_i, t)]$ where the dependence on the local co-ordinates is indirect. This does not disclose the possibility of localisation of the theory in space-time, however, it does not ensure it automatically. Local symmetry in a kinetic field means that the objects, fields or physical laws in question are invariant under a local transformation, namely under a set of continuously infinite number of separate transformations with an arbitrarily different one at every velocity in the given reference frame.

The isotopic field charge (IFC, $\daleth$) as a property can be identified in the case of the gravitational field with the properties of the masses of gravity and inertia respectively. The potential isotope of $\daleth(\daleth_V)$ depends directly on space-time co-ordinates. The physical state of the kinetic isotope of $\daleth(\daleth_T)$ depends primarily on the components of its velocity (and indirectly on its space-time co-ordinates).

When we try to specify physical phenomena that distinguish kinetic behaviour of objects from their behaviour in a field caused by another, potential source (i.e., $\daleth_V$) we should make attempt to seek for a description in a velocity dependent field.

3.2 Mathematical background of the conservation of Δ

For the sake of the description of the mentioned distinction, we introduce a gauge field D_μ, that depends primarily on velocity. We derived a set of conserved currents in such a field [17]. The mathematical treatment is as much general as possible, while we made a specification. Namely, Noethers second theorem allows the dependence of the concerned fields (on which the Lagrangian depends) on any, general co-ordinate. Certain physical theories restrict themselves on the four space-time co-ordinates as dependent variables. We discuss fields that depend on co-ordinates in the velocity four-space, (and handle the space-time co-ordinates as indirect variables).

For the effects of a general non-Abelian group on the local gauge invariance are to be described, we refer to the [24] review paper. We partially use the methods of his description of YM type gauge fields. We introduce a new type of localised gauge field that does not coincide with the isotopic spins YM field, marked by **B** in [23] and [24]; this field, marked by **D**, is *per defintionem* different from the YM field.[6] In our discussion, the **D** gauge field, introduced below, depends directly on the velocity-space coordinates, while the matter field depends directly on the four dimensional space-time co-ordinates. In other words, this means that although we primarily use coordinates of the velocity-space, our derivations are indirect and include derivatives with respect to the space-time co-ordinates (cf., the introduction of the relativistic λ_μ^ν tensor below) and play important role in our conclusions. This is an expression of the facts that we observe the physical events (occurring even in the velocity space) with respect to the 4D space-time, on the one hand, and that our operators should effect complex $\psi(x_\nu)$ fields which depend on the four space-time co-ordinates, on the other.

We extend the role of the co-ordinates to a set of generalised variables alike Noether [31] did. These variables may be the four space-time co-ordinates or

[6] Although we use the letter "**D**" to denote this gauge field, in [24] and many other publications that letter denoted the covariant derivative, which we will mark by careted (capped) derivation mark $\hat{\partial}$.

they may be others (and their number may vary). In her mathematical terms of invariant variational problems, the space-time co-ordinates did not play a distinguished role. According to her second theorem, other variables, among others (e.g., velocity-space co-ordinates), are allowed which may implicitly depend on the space-time co-ordinates. For practical reasons we replace the $f(\dot{x}_\mu, x_\nu)$ dependence with a $f(\dot{x}_\mu(x_\nu))$ dependence. The localisation is present here too (in the above generalised, Noetherian sense), although it makes us possible another way of calculating it.

We were seeking for invariance between scalar fields and (gauge) vector fields that describe kinetic processes, the latter depending therefore primarily on velocity. For this reason, we consider Lagrangians which depend on matter fields φ_k, and gauge fields $D_{\dot{\mu},\alpha}$, which all depend - in simple mathematical terms - on parameters. In physical terms these parameters are generally identified with the four space-time co-ordinates. In our specific case the dependence of $\mathbf{D}$ on x_μ will be given by the formula: $D_{\dot{\mu}} = D_\mu\left(\frac{\partial x^\mu}{\partial x_4}\right)$, or in another form $D_{\dot{\mu}} = D_\mu\left[\dot{x}^\mu(x_\nu)\right]$. The 2nd theorem of Noether is just about Lagrangians, which depend on arbitrary number of fields with arbitrary finite number of derivatives by arbitrary number of parameters. We can apply her theorem here because in mathematical terms she did not specify either the physical-mathematical character or the number of applicable parameters. Our consideration will be justified by the final result, which demonstrates that in a boundary situation, namely in the absence of a velocity-dependent gauge field we obtain the same currents that were derived in a space-time dependent field, (cf. Eqs. (4) and (7) below). In other words, in the absence of relativistically high velocities or acceleration, the effect of the velocity dependent gauge field can be neglected, and we get back to the same currents as derived in the semi-classical, only space-time dependent gauge's case. At the same time, in the presence of a velocity dependent gauge field, we derived new conserved Noether currents [17].

3.2.1 Noether's currents for gauge invariance localised in the velocity space

The presentation discusses general, non-Abelian case. Let's first introduce a (kinetic) $\mathbf{D}$ field localised in the velocity space, with components $D_{\dot{\mu}} = D_{\dot{\mu}}(\dot{x}^\mu)$, where $\dot{x}^\mu = \dot{x}^\mu(x_\nu)$; $(\mu, \nu = 1, 2, 3, 4)$; (dotted indices denote the velocity-space components).

We introduce a λ_μ^ν tensor defined as $\lambda_\mu^\nu = \partial_\mu \dot{x}^\nu = \frac{\partial \dot{x}^\nu}{\partial x_\mu}$ (Lorentz invariant acceleration), which characterises the changes of the velocity-space components in the space-time.

Localisation will be taken into consideration in this way (we refer to the generalised interpretation of localisation as defined above).[7]

In general, we base on a transformation group G and the transformations of its elements into each other $T\left[G_{\infty,\rho}\right] = T\left[p_\alpha(x_\beta)\right]$, where the number of parameters are arbitrary finite numbers $(\alpha = 1, \ldots, \rho))$; , $(\beta = 1, \ldots, \sigma)$. The p are parameters on which the transformations, constituting the group elements, depend. They take the form of functions $p_\alpha(x_\beta)$ and their derivatives. The group transformations depend on p and are finitely differentiable. G may take the form of different groups, depending on the concrete form of interaction in subject, namely $SO(3,1)$, $U(1)$, $SU(2)$, $SU(3)$ in the cases of the fundamental physical interactions.

We consider a Lagrangian density $L(\varphi_k, D_{\dot{\mu},\alpha})$, where φ_k, $(k = 1, \ldots, n)$ are the matter fields - which also includes the velocity field $\dot{x}^\mu = \dot{x}^\mu(x_\nu)$ –, and $D_{\dot{\mu},\alpha}$, $(\alpha = 1, \ldots, N)$, are the (kinetic) gauge fields. We assume, that $L(\varphi_k, D_{\dot{\mu},\alpha})$ is invariant under the local transformations of a compact, simple Lie group G generated by T_α, $(\alpha = 1, \ldots, N)$, where $[T_\alpha, T_\beta] = iC^\gamma_{\alpha\beta} T_\gamma$, and $CY^\gamma_{\alpha\beta}$ are the so-called structure constants, corresponding to the actually considered individual physical interactions symmetry group.[8] For examples, in the case of $SU(2)$ symmetry, G consists of 2×2 matrices with 3 independent components, representing a state doublet, and in the case of $SU(3)$ its matrix has 8 independent components, representing a state triplet. For simplicity we assume that the matter fields belong to a single, n-dimensional representation of G.

Let us consider a local transformation $V(\dot{x}) \in G$ parameterised by $p_\alpha(\dot{x})$ that acts on ψ as $\psi = V\psi'$

$$V(\dot{x}) = e^{-ip_\alpha(\dot{x})T_\alpha}$$

The infinitesimal transformations of the matter- and the gauge fields determine

[7] Relativistic covariance under Lorentz transformation $S(\Lambda)$ and its consequences are a standard part of quantum field theory textbooks for long, e.g., [32, Sec. 2.1.3]. Here we take into account time derivatives of Lorentz transformed velocities.

[8] We partly follow the clues by Higgs [29] and Weinberg [33] at the beginning of their papers with the exception that we consider different dependencies in the potential and kinetic Hamiltonian terms.

the covariant derivatives of ψ in the gauge field. (For invariance, we can require that the derivatives of ψ coincide with the derivatives of $V\psi'$). The infinitesimal transformations can be formulated as follows:

$$\delta\varphi_k = -ip_\alpha(\dot{x})(T_\alpha)_{kl}\varphi_l(\dot{x}) \qquad (k = 1, \ldots, n)\,, \tag{1}$$

where the T_α are matrix-representation operators generating the group G, with the above commutation rule $[T_\alpha, T_\beta] = iC^\gamma_{\alpha\beta}T_\gamma$, and

$$\delta D_{\mu,\alpha} = \frac{1}{ℷ}\partial^{\dot{\rho}}p_\alpha(\dot{x})\partial_\mu\dot{x}^\rho + C^\gamma_{\alpha\beta}p_\beta(\dot{x})D_{\dot{\mu},\gamma}(\dot{x}) \qquad (\alpha = 1, \ldots, N) \tag{2}$$

where $\partial^{\dot{\rho}} = \frac{\partial}{\partial\dot{x}^\rho}$, and ℷ (Hebrew g, gimel) denotes a general coupling constant, which can be replaced by a concrete coupling constant for each individual physical interaction.

For the induced infinitesimal transformation δL of the Lagrangian density $L(\varphi_k, D_{\dot{\mu},\alpha})$, on using the field equations for both the matter and the gauge fields, one obtains

$$\delta L = \partial_\mu\left(\frac{\partial L}{\partial(\partial_\mu\varphi_k)}\delta\varphi_k + \frac{\partial L}{\partial(\partial_\mu D_{\nu,\alpha})}\delta D_{\nu,\alpha}\right)\,. \tag{3}$$

One would like to describe the events, resulted in the interaction between the matter field and the kinetic (velocity-space dependent) gauge field, as they are observed from the usual 4D space-time. Therefore one needs to apply derivatives by the space-time co-ordinates. Substituting from (1) and (2) into (3), using the notation $\frac{\partial\dot{x}^\nu}{\partial x_\mu} = \partial_\mu\dot{x}^\nu = \lambda^\nu_\mu$ (Lorentz invariant acceleration), and a permutation of the indices, one can obtain

$$\begin{aligned}\delta L =& \partial_\mu\left(\frac{\partial L}{\partial(\partial_\mu\varphi_k)}(-i)p_\alpha(\dot{x})(T_\alpha)_{kl}\varphi_l(\dot{x})\right) + \\ &+ \partial_\mu\left(\frac{\partial L}{\partial(\partial_\mu D_{\dot{\nu},\alpha})}\frac{1}{ℷ}\partial^{\dot{\rho}}p_\alpha(\dot{x})\lambda^\rho_\nu\right) + \delta_\mu\left(\frac{\partial L}{\partial(\partial_\mu D_{\dot{\mu},\alpha})}C^\gamma_{\alpha\beta}p_\beta(\dot{x})D_{\dot{\nu},\gamma}(\dot{x})\right)\,.\end{aligned}$$

We have derived from here the following two sets of equations:

$$J^{(1)\nu}_\alpha(x) = \partial_\mu F^{(1)\mu\nu}_\alpha(x) \quad \partial_\nu J^{(1)\nu}_\alpha = 0 \tag{4}$$

$$J^{(2)\nu}_\alpha = \partial_\mu F^{(2)\mu\nu}_\alpha \qquad \partial_\nu J^{(2)\nu}_\alpha = 0 \tag{5}$$

completed with

$$\frac{\partial L}{\partial(\partial_\mu D_{\dot{\mu},\alpha})}\lambda_\nu^\rho + \frac{\partial L}{\partial(\partial_\nu D_{\mu,\alpha})}\lambda_\mu^\rho = 0 \tag{6}$$

this set (4)-(6) demonstrates, that in the presence of a kinetic (velocity-dependent) gauge field, there exist two (families of) conserved Noether currents. Although the two conserved currents are not independent, in the presence of a kinetic gauge field they exist simultaneously. (One can easily see, that λ_μ^ν mixes the components of the gauge-field currents depending on the 4D velocity space in a similar way, like the Lorentz transformation mixes the co-ordinates of four-vectors in the 4D space-time; since the λ_μ^ν tensor was defined to characterise the changes of the velocity-space components - accelerations - in the space-time.) Taking into account the conditions how we have obtained these currents, one can write $J_\alpha^{(1)\mu}$ as

$$J_\alpha^{(1)\mu}(\dot{x}) = i\daleth\frac{\partial L}{\partial(\partial_\nu, \varphi_k)}(T_\alpha)_{kl}\varphi_i(\dot{x})\,. \tag{7}$$

The most significant conclusion of the above cited derivation (cf., [17]) is that in the presence of a kinetic gauge field $\mathbf{D}$, there appear extra $J_\alpha^{(2)\nu}$ conserved currents. Taking into account conditions of the derivation of $J_\alpha^{(2)\nu}$, one can write it in the form

$$J_\alpha^{(2)\nu} = i\daleth\left[\frac{\partial L}{\partial(\partial_\mu\varphi_k)}(T_\alpha)_{kl}\varphi_l(\dot{x})\lambda_\mu^\nu - C_{\alpha\beta}^\gamma D_{\dot{\omega},\beta}(\dot{x})\lambda_\mu^\omega \times F_\gamma^{(2)\mu\nu}(x)\right]\,. \tag{8}$$

Their dependence on the velocity-space gauge is apparent, although, none of the conserved vector currents involve the gauge parameters $p_\alpha(\dot{x})$ and their derivatives.

$>$From (4) and (7), considering consequences of (6), one obtains

$$\partial_\mu F_\alpha^{(1)\mu\nu}(\dot{x}) = i\daleth\frac{\partial L}{\partial(\partial_\nu\varphi_k)}(T_\alpha)_{kl}\varphi_l(\dot{x})\,. \tag{9}$$

$>$From (5) and (8), considering the concrete forms of the covariant derivatives, one obtains

$$\hat{\partial}_\mu F_\alpha^{(2)\mu\nu}(x) = i\daleth\frac{\partial L}{\partial(\partial_\mu\varphi_k)}(T_\alpha)_{kl}\varphi_l(\dot{x})\lambda_\mu^\nu\,. \tag{10}$$

3.3 Mathematical conclusions

First conclusion - of the conserved Noether current (4) - is a conserved quantity: Conservation of the field charge ($\daleth$).

Second conclusion - of the conserved Noether current (5) - is another conserved quantity: Conservation of the isotopic field charge spin (Δ).

Further, we could derive, in the usual way, the total isotopic field charge spin

$$\Delta = \frac{i}{\gimel} \int J^{(2)4} \mathrm{d}^3 x \, .$$

which is independent of time and independent of Lorentz transformation. $J^{(2)\mu}$ does not transform as a vector, while $\varDelta$ transforms as a vector under rotations in the isotopic field charge spin field.

3.4 Physical conclusions

Coupling of the two conserved quantities ($\daleth$ and $\varDelta$), what is based on the dependence of the two currents $J_\alpha^{(1)\mu}$ and $J_\alpha^{(2)\mu}$ on each other, has physical consequences. The quantities, whose conservation they represent, and which are coupled (by $\lambda_\mu^\nu = \partial_\mu \dot{x}^\nu$), exist simultaneously. The *derived conservation law verifies* just *the invariance between two isotopic states of the field charges, namely between the potential* $\daleth_V$ *and the kinetic* $\daleth_T$ what we intended to prove. We obtained, that in *the presence of kinetic fields we have two conserved currents that are effective simultaneously.* The kinetic gauge field $\mathbf{D}$ is present simultaneously with the interacting matter $[\varphi]$ and gauge $[\mathbf{B}]$ fields. The presence of $\mathbf{D}$ corresponds to the property of the field charges $\daleth$ of the individual fields that they split in two isotopic states, and analogously to the isotopic spin, we named these two states *isotopic field charge spin* what we denoted by $\varDelta$. The source of the isotopic field charge spin ($\varDelta$) is the field $\varphi(\dot{x})$, in interaction with the kinetic gauge field $\mathbf{D}$.

The physical meaning of $\varDelta$ can be understood, when we specify the transformation group associated with the $\mathbf{D}$ field, which describes the transformations of $\daleth$ (i.e., the isotopic field charges). $\daleth$ can take two (potential and kinetic) isotopic states $\daleth_V$ and $\daleth_T$ in a simple unitary abstract space. Their symmetry group is

$SU(2)$, that can be represented by 2×2 T_α matrices. There are three independent T_α that may transform into each other, following the rule $[T_\alpha, T_\beta] = iC^{\gamma}_{\alpha\beta}T_\gamma$, where the structure constants can take the values 0, ± 1. Let T_1 and T_2 be those which do not commute with T_3; they generate transformations that mix the different values of T_3, while this "third" component's eigenvalues represent the members of a $\varDelta$ doublet. For the isotopic field charges compose a $\daleth$ doublet of $\daleth_V$ and $\daleth_T$, the field's wave function can be written as

$$\psi = \begin{pmatrix} \psi_T \\ \psi_V \end{pmatrix} . \tag{11}$$

(11) is the wave function for a single particle which may be in the "potential state", with amplitude ψ_V, or in the "kinetic state", with amplitude ψ_T. ψ in (11) represents a mixture of the potential and kinetic states of the $\daleth$, and there are T_α that govern the mixing of the components ψ_V and ψ_T in the transformation. T_α $(\alpha = 1, 2, 3)$ are representations of operators which can be taken as the three components of the isotopic field charge spin, $\varDelta_1$, $\varDelta_2$, $\varDelta_3$ that follow the same (non-Abelian) com-mutation rules as do the T_α matrices, $[\varDelta_1, \varDelta_2] = i\varDelta_3$, etc. These operators represent the charges of the isotopic field charge spin space, and ψ are the fields on which the operators of the gauge fields act.

The quanta of the **D** field should carry isotopic field charge spin $\varDelta$. The $\varDelta$ doublet, as a conserved quantity, is related to the two isotopic states of field charges ($\daleth$), and the associated operators ($\varDelta_i$) induce transitions from one member of the doublet to the other.

3.5 Interpretation of the isotopic field charge spin conservation

Invariance between $\daleth_V$ and $\daleth_T$ means that they can substitute for each other arbitrarily in the interaction between field charges of any given fundamental physical interaction. They appear at a probability between [0, 1] in a mixture of states in the wave function ψ (11) so that the Hamiltonian of a *single particle* oscillates between V and T, while the Hamiltonian of a *composite system* is a mixture of the oscillating components of the particles that constitute the system. The individual particles in a *two-particle system* are either in the V or in the T state respectively, and switch between the two roles permanently; while the observable value of H is the expected value of the mixture of the actual states of the two, always opposite state particles.

The invariance between $\daleth_V$ and $\daleth_T$ (what is ensured by the conservation of $\varDelta$), and their ability to swap, means also that they can restore the symmetry in the physical equations which was lost when we replaced the general $\daleth$ (in our case mass m) by their isotopes $\daleth_V$ and $\daleth_T$ (concretely m_V and m_T).[9]

We denote the *predicted* quanta of the **D** field by δ. We call this hypothetical boson *"dion"*, after the Greek term meaning 'flee', 'flight', 'rout' in English. The δ quanta (dions) carry the $\varDelta$ (isotopic field charge spin as a physical property: charge of the **D** field). According to the IFCS model, gravitational interaction takes place between two massive particles wit the simultaneous exchange of a graviton and a dion.

Starting from the equivalence principle, through the qualitative distinction of the masses of gravity and inertia as isotopic field charges of the gravitational field and interaction between them, we concluded the prediction of a boson that mediates their interaction.

4 Finsler geometry in the presence of isotopic field charges

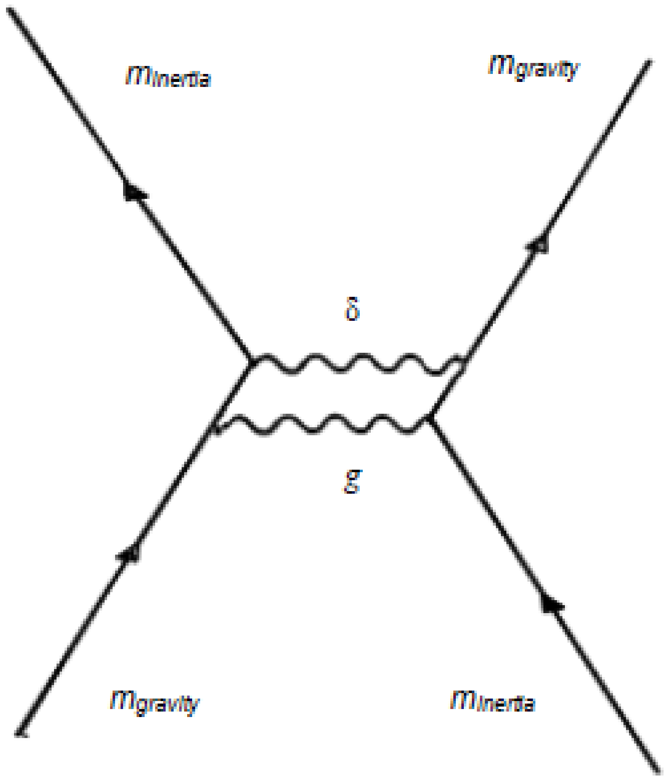
Let us specify (9) for the gravitational field [35]. The right side of the equation contains the scalar field that serves for the source of the gravitational field. The $\gimel$ can be replaced by the gravitational coupling constant g. As we noticed, the dependence on the gauge fields is on the left side of the equation (9). $F^{(1)\mu\nu}(\dot{x})$ must satisfy the

$$T_{\mu\nu} = F_{\mu\lambda}F_{\lambda\nu} + \frac{1}{4}\delta_{\mu\nu}g^{\kappa\sigma}F_{\lambda\sigma}g^{\lambda\rho}F_{\kappa\rho}$$

identity for the energy-momentum tensor $T_{\mu\nu}$. (In order to bring this form in compliance with the indices in (9), one should raise the indices by multiplying with the metric tensor g_{β_γ} in the right side.) This energy-momentum tensor $T_{\mu\nu}$ can be expressed by the way of the Einstein equation

$$T_{\mu\nu} = -\frac{1}{8\pi G_N}(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu}) \tag{12}$$

[9] Consequences of the application of effective field theories were analysed e.g., in philosophy by E. Castellani [34] and in physics by S. Weinberg [22].



where $R_{\mu\nu}$ is the Ricci tensor defined by the help of the derivatives of the metric tensor $g_{\mu\nu}$, $\mathbf{R}$ is the Ricci scalar formed from the Ricci tensor (Riemann curvature) and the metric tensor, and Λ is a constant of Nature, as well as G_N the constant of Newton.

The metric tensor $g_{\mu\nu}$ and its derivatives depend on the localisation of the given point in the space-time in the General Theory of Relativity (GTR), and are subject of Riemann geometry. In the presence of a kinetic field, that means, isotopic mass field $\mathbf{D}$ (mass being the field-charge of the gravitational field), however, the curvature depends also on velocity. (Whose velocity? On the actual inertial velocity of atest unit-mass placed in a given space-time point in the referenceframe fixed to the source of a scalar gravitational field φ which appears on the right side of (9).) The $g_{\mu\nu}$ metric tensor, and consequently the affine connection field and the curvature tensor formed from its derivatives, depend on space-time and velocity co-ordinates. With the appearance of the dependence on the velocity vector, the curvature becomes dependent on its direction in each space-time point. The direction (additional parameter) attributed to each space-time point is defined by the orientation of the velocity of a test unit-mass in the given space-time point, $\frac{\mathbf{v}}{|\nu|}$. The curvature can no more follow a “simple” Riemann geometry, it follows a Finsler geometry whose metric is defined by the dependence of $g_{\mu\nu}$ on (x_σ and) $\dot{x}_\rho$.

Of course, the space-time plus four-velocity dependence of the metric tensor $g_{\mu\nu}$ affects its all derivatives, including the formation of the affine connection field (from first derivatives) and the Riemann curvature (or Ricci tensor, second, covariant derivative)

$$\Gamma_{\lambda\mu\nu} = \frac{1}{2}\left[\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right] \qquad \Gamma^\lambda_{\mu\nu} = g^{\lambda\rho}\Gamma_{\rho\mu\nu}$$

and

$$R_{\mu\nu} = \partial_\mu \Gamma^\lambda_{\nu\lambda} - \partial_\lambda \Gamma^\lambda_{\mu\nu} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\lambda} - \Gamma^\lambda_{\sigma\lambda}\Gamma^\sigma_{\mu\nu}\,.$$

The solution of the Einstein equation in velocity dependent field with Finsler geometry must necessarily lead to solutions different from that of Schwarzschild.

5 The role of the isotopic field charge spin conservation

The role of equation (12) is to retain the invariance between the two isotopic forms, namely gravitational and inertial, of masses. The importance of this is to save the covariance of our equations. Since there appear two different kinds of (isotopic) masses in the energy-momentum “four-vector” (in the fourth column of $T_{\mu\nu}$) it does no more transform as a vector, and Lorentz transformation can no more guarantee alone the covariance of our equations.

As a consequence of the distinction between m_V and m_T, as well as the association of the energy content with the mass m_V and the components of the momentum with m_T, we lose also the symmetry of the $T_{\mu\nu}$ energy-momentum tensor. To retain symmetry in Einstein's field equations we must require again the invariant transformation of m_V and m_T into each other in an appropriate gauge field, namely in **D**. We refer to [24] who foresaw the possible generalisation of YM type gauge invariance in general relativity "in close analogy with the curvature tensor". If we consider the energy-momentum tensor (in which both isotopic states of mass appear) as the source of the gravitational field, then - in the usual way - the scalar and the vector potential can be separated. See, m_4 in T_{44} does not compose a fourth component of a four-vector in the classical theory of gravitation where there is a single scalar mass. If we consider now $m_4 = m_V$, the three components of the kinetic mass m_T can compose a three-vector, however $T_{\mu 4}$ will not form a four vector either. To maintain the Lorentz invariance of our physical equations in the gravitational field, we must demand to restore the invariance of $\begin{pmatrix} \overrightarrow{m}_T \\ m_V \end{pmatrix}$ under an *additional transformation* that should *counteract the loss of symmetry caused by the introduction of two isotopic states of mass.* We discussed that transformation in section 2. Further, in the case of gravitation the relation of the scalar and the vector fields are not linear even if we have not made distinction between the potential and kinetic masses. The non-linearity is coded in the relation of the tensors [25] at the right side of the Einstein equation (12) (in units $c = 1$), or we can write $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}$ where the Einstein tensor is defined as $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$ whose covariant derivative must vanish.

Since our $T_{\mu\nu}$ tensor has already lost its symmetry, we can take $\Lambda g_{\mu\nu}$ into account within a modified $T_{\mu\nu}$ - handling the gravitational and kinetic masses in it together with the dark energy - and we get the following formally symmetric equation: $G_{\mu\nu} = 8\pi G T'_{\mu\nu}$.

The symmetry of the energy-momentum tensor can be saved by the invariant gauge transformation of the IFCS. The most important analogy is between the behaviour of the potential and the kinetic field charges of the individual fields that makes probable to postulate a unique transformation to assure their invariance (cf., section 2).[10] So the invariance under the Lorentz transformation combined

[10] As [26] stated, "In contrast to the symmetry or invariance requirement in STR, the principle in GTR is most often presented as strictly speaking a covariance requirement." Gauge theories behave like GTR, at least in this respect. General covariance "is not tied to any geometrical regularity of the underlying spacetime, but rather the form invariance (covariance) of laws under arbitrary smooth coordinate transformations" [26, p. 34]. [27] found that the more general geometry resulting from admitting local changes called gauges described not only gravity but also electromagnetism. He showed also that the conservation laws of Noether follow in two

with the invariance of the isotopic field charge spin field provide together the covariance of the gravitational equation. However, this combined transformation should now be taken into consideration in a field with a metric depending on all space-time and velocity co-ordinates, following a Finsler geometry.

6 Appendix

Comparison of the invariance properties in classical GTR and in the IFCS model

In classical physics, conservation laws - as consequences of the invariance properties of the investigated systems - can be obtained by integration of the Euler-Lagrange equations of motion of classical mechanical point systems. According to Hamilton's principle the variation of the action integral of the systems Lagrangian must be zero. These conservation laws include the conservation of the energy - invariance under translation in time. That conserved energy is equivalent with a well determined amount of mass $E = mc^2$, where $m = m_V$ is gravitational mass, and this conservation law does not provide any information on the quantity of kinetic mass.

In general relativistic treatment, the source of the gravitational field is the $T_{\mu\nu}$ momentum-energy stress tensor, which includes the sources of inertial and gravitational effects as well. Applying the same variational method and integration for the Einstein equation (using $[+ + + -]$ signature) we derive the conservation of the $-T_{44}$ element of the $T_{\mu\nu}$ momentum-energy stress tensor. $-T_{44}$ is energy density of the gravitational field, and is proportional to a certain amount of mass. According to invariance under translations in the Minkowski space (Lorentz transformation) the conserved current can be written in the form

$$\partial_\mu T_{\mu\nu} \equiv \partial_\mu \left(L\delta_{\mu\nu} - \partial_\nu \varphi_r \frac{\partial L}{\partial \partial_\mu \varphi_r} \right) = 0$$

where φ_r denote functions on which (and their first derivatives) the Lagrangian may depend.

---

distinct ways in theories with local symmetries. This led to the Bianchi identities, which hold between the coupled equations of motion, and which are due to the local gauge invariance of action. Later [28] demonstrated that the conservation of the electric charge followed from the local gauge invariance in the same way as does energy-momentum conservation from co-ordinate invariance in GTR.

The Einstein equation

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}$$

provides the elements of $T_{\mu\nu}$ in which - according to the left side - the contribution of the kinetic and potential components are mixed by the $g_{\mu\nu}$ curvature tensor. Applying the usual integration method and Gauss' theorem, we get the fourth column of the momentum-energy stress tensor for a conserved quantity, what is no else than the four-momentum density, which behaves like a four-vector and whose individual components are

$$P_\nu = \frac{1}{ic}\int T_{4\nu}\mathrm{d}V$$

or separated

$$P_k = \frac{1}{ic}\int T_{4k}\mathrm{d}V = \frac{1}{ic}\int \partial_k\varphi_i\frac{\partial L}{\partial\partial_4\varphi_i}\mathrm{d}V \qquad (k = 1, 2, 3)\,;$$

$$H = -icP_4 = -\int T_{44}\mathrm{d}V = \int(\partial_4\partial_i\frac{\partial L}{\partial\partial_4\varphi_i} - L)\mathrm{d}V$$

what are considered the conserved total momentum and energy of the field respectively.

If we take into account the qualitative difference between the masses m_T (what appear in the components of P_k) and m_V (what appears in H) that are mixed by the curvature tensor $g_{\mu\nu}$ in the elements of $T_{\mu\nu}$, this consideration will involve the mixed m_T and m_V dependence of the Lagrangians as well. As a consequence, P_k and H cannot be considered separately, and independently of each other, conserved quantities. (We do not investigate here the ambiguous interpretations of invariant mass.) The covariance of the gravitational equation can no more be secured by the Lorentz invariance alone. The lost symmetry of nature can be restored only with the shown invariance between the isotopic mass states (as field charges of the gravitational field, conservation of $\mathit{\Delta}$) which are rotated in an isotopic field charge spin gauge field. The covariance of the gravitational equation is a result of invariance under the combination of the Lorentz transformation and rotation in the isotopic field charge field. In the latter case the four components of $(P_k[m_T], H[m_V])$ transform as isovectors. Due to the IFCS gauge transformation, the transformation of the field components can be described in a (space-time +) velocity dependent gauge field, whose metric, consequently, depends also on the velocity components, and is subject of a Finsler geometry.

References

1 Einstein, A. (1921) A brief outline of the development of the theory of relativity, *Nature*, 106, 2677, Feb. 17, 782-784.

2 Einstein, A. (1935) Elementary derivation of the equivalence of mass and energy, *Bull. Amer. Math. Soc.*, 41, 223-230.

3 von Laue, M. (1911) *Ann. Phys.* 35, 524-542.

4 von Laue, M. (1955) *Relativitätstheorie*, 6th ed. Braunschweig.

5 Mie, G. (1912a) *Ann. Phys.*, 39, 1-40.

6 Mie, G. (1912b) Grundlagen einer Theorie der Materie, *Ann. Phys.*, 37, 511-534.

7 Mie, G. (1913) *Ann. Phys.*, 40, 1-66.

8 Hraskó, P. (2003) Ekvivalens-e egymással a tömeg és az energia? [Are mass and energy equivalent? In Hungarian] *Fizikai Szemle*, 53, 9, 330.

9 de Broglie, L. (1923) Note. Radiations. - Ondes et quanta, *Comptes Rendus*, 177, 507-510.

10 de Broglie, L. (1925) Recherches sur la théorie des quanta, *Annales de Physique*, 10, III, 22-128.

11 de Haas, E. P. J. (2004a) The combination of de Broglie's harmony of the phases and Mie's theory of gravity results in a principle of equivalence for quantum gravity, *Ann. Fond. L. de Brioglie* 29, 4, 707-726.

12 de Haas, E.P.J. (2005) From Laue's stress-energy tensor to Maxwell's Equations and the implications for Einstein's GTR, *Proceedings of the Int. Conference "hysical Interpretation of Relativity Theory (PIRT-05)"*, eds. M.C.Duffy et al., Moscow, Bauman Univ. Press.

13 Darvas, G. (2006) Invariance, identity, equivalence, *Symmetry: Culture and Science*, 17, 175-192.

14 Darvas, G. (2011) The Isotopic Field Charge Spin Assumption, *International Journal of Theoretical Physics*, 50, 10, 2961-2991. DOI: 10.1007/s10773-011-0796-9

15 Darvas, G. (2012, coming) *Another Version of Facts - on Physical Interactions.*

16 Darvas, G. (2007) A velocity dependent gauge invariance, In: *Symmetry in Nonlinear Physics*, Kiev, Institute of Mathematics, Ukrainian Academy of Sciences,
http://www.imath.kiev.ua/ appmath/Abstracts2007/Darvas.html

17 Darvas, G. (2009) Conserved Noether currents, Utiyama's theory of invariant variation, and velocity dependence in local gauge invariance, *Concepts of Physics*, VI, 1, 3-16.

18 Hooft, Gerard't (2005) "Conceptual Basis of Quantum Field Theory" In: Elsevier *"Handbook of the Philosophy of Science"*.

19 CERN workshop, (2008a) Report, http://arxiv.org/pdf/0801.1800

20 CERN workshop, (2008b) Report, http://arxiv.org/pdf/0801.1833

21 CERN workshop, (2008c) Report, http://arxiv.org/pdf/0801.1826

22 Weinberg, S. (1997) Changing Attitudes and the Standard Model, p. 38, in Hoddeson, L., L. Brown, M. Riordan and M. Dresden: *The Rise of the Standard Model*, Cambridge University Press, Cambridge., pp. 3644.

23 Yang, C. N., Mills, R. L. (1954) Conservation of isotopic spin and isotopic gauge invariance, *Phys. Rev.* 96 (1), 191-195.

24 Mills, R (1989) Gauge fields, *American Journal of Physics*, 57 (6) June, 493-507.

25 Hooft, Gerard't (2002) *Introduction to General Relativity*, Princeton: Rinton Press Inc.

26 Martin, C.A. (2003) On continuous symmetries and the foundations of modern physics, in Brading, K. and Castellani, E., editors, *Symmetries in Physics: Philosophical Reflections*, pp. 29-60, Cambridge: Cambridge University Press, 445 p.

27 Weyl, H. (1918) Gravitation und Elektrizität, *Sitzungsberichte Preussische Akademie der Wissenschaften*, 465-480.

28 Weyl, H. (1929) Electron and gravitation, (Engl. transl. in: O'Raifeartaigh: *The Dawning of Gauge Theory*, 1997, Princeton Univ. Press).

29 Higgs, P.W. (1966) Spontaneous symmetry breakdown without massless bosons, *Phys. Rev.* 145, 1156-1162.

30 Wigner, E.P. (1984) The meaning of symmetry, pp. 729-733, in Zichichi, A. ed., *Gauge interactions theory and experiment*, Proceedings of the 20^{th} international school of subnuclear physics, Erice, Sicily, New York: Plenum Press.

31 Noether, E. A. (1918) Invariante Variationsprobleme, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen: Mathematisch-physikalische Klasse*, 235-257.

32 Itzykson, C., and Zuber, J.B. (1980) *Quantum Field Theory*, New York: McGraw-Hill.

33 Weinberg, S. (1973) General theory of broken symmetries, *Phys. Rev.* D 7, 1068-1082

34 Castellani, E. (2001) Reductionism, emergence and effective field theories, *arXiv:physics/0101039 v1*, 18 p.

35 Darvas, G. (2011) GTR and the Isotopic Field Charge Spin Assumption, pp. 53-65, In: *Physical Interpretations of Relativity Theory*, Ed. M.C. Duffy, V.O. Gladyshev, A.N. Morozov, P. Rowlands, Bauman Moscow State Technical University: Moscow, Liverpool, Sunderland, 347 p.

M. Eingorn[11], *Zhuk A.*

Problematic aspects of toroidal and spherical compactifications of extra dimensions in Kaluza-Klein geometries[12]

Odessa National University named after I.I. Mechnikov,
2, Dvoryanskaya street, Odessa 65082, Ukraine
e-mail: maxim.eingorn@gmail.com

In our recent papers we have investigated linear (with respect to the scalar curvature) as well as non-linear f(R) Kaluza-Klein models with both toroidal and spherical compactifications of an arbitrary number of extra dimensions and non-dust-like matter sources of the gravitational field concerning compatibility with the experimental data (laboratory and astrophysical tests). In the case of toroidal extra dimensions only latent solitons (in particular, black strings and black branes), possessing relativistic negative tension, satisfy the observations. In the

[11]This research was co-financed by the Austrian Academy of Sciences in the framework of the collaboration with the National Academy of Sciences of Ukraine on Modern Problems in Astroparticle Physics.

[12]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

case of spherical compactification a dust-like matter source also satisfies them under certain conditions, imposed on the parameters of the model. However, there is the other problematic aspect in this case. Together with the perturbation of the background matter, which provides the internal space curvature, the dust-like source looks like an effective source, which has the non-dust-like equation of state in the external space. Again introducing tension, one can avoid this difficulty, but tension itself has no clear physical origin. Thus, this possibility is not satisfactory enough.

1 Introduction

The multidimensionality of spacetime is an essential property of the modern theories of unification such as superstrings, supergravity and M-theory, which have the most self-consistent formulation in spacetime with extra spatial dimensions [1]. Obviously, these physical theories should be consistent with observations. For example, in the weak field limit they must satisfy the gravitational experiments such as the perihelion shift, the deflection of light and the time delay of radar echoes. It is well known that general relativity is in good agreement with these experiments [1]. Therefore, in order to investigate the similar correspondence for multidimensional theories, in our recent papers [3]- [9], we investigated classical gravitational tests (the perihelion shift, the deflection of light and the time delay of radar echoes) in Kaluza-Klein (KK) models.
We paid attention mainly to theories with toroidal compactification of extra dimensions, i.e. with compact and flat internal spaces. These theories are very popular in the literature devoted to KK models. Generalizing the standard approach [7] of general relativity, we supposed that the background metrics (in the absence of the matter source) is flat for our external four-dimensional spacetime and all internal spaces, and a pointlike matter source has dustlike equations of state in all spatial dimensions. To our surprise, the obtained formulas strongly contradict the observations [3].
It turned out that in order to satisfy the experimental data, the matter source should have negative parameters in equations of state in the internal spaces (tension) [4, 5]. For example, latent solitons have tension and satisfy the gravitational

tests at the same level of accuracy as general relativity [5]. The uniform black strings and black branes are particular examples of the latent solitons. The similar situation takes place for nonlinear (with respect to the scalar curvature) KK theories with toroidal compactification [6,7]. Here, a pointlike mass with the dustlike equation of state in all spatial dimensions also contradicts the observations [6], but there are two classes of asymptotic latent solitons, which are in agreement with the observations at the same level of accuracy as general relativity [7]. For both of these classes, a gravitating mass has tension in the internal space with unclear physical origin.

Let us note that the metric coefficients for uniform black strings/branes depend only on the three-dimensional radius-vector. Therefore, a matter source is uniformly smeared over the extra dimensions and the nonrelativistic gravitational potential exactly coincides with the Newtonian one.

Then, we generalized our investigation to the case of KK models with spherical compactification of the internal space [8, 9]. Here, the background metrics is not flat because the internal space (e.g., the two-sphere) is curved. To create such background, we need to introduce the background matter. A pointlike (and, for example, dustlike) mass disturbs this background. In the presence of a bare cosmological constant the perturbed metric coefficients have the Yukawa type corrections with respect to the usual Newtonian gravitational potential. These corrections are negligible in the Solar system, and the considered model satisfies the gravitational tests.

Moreover, all models with spherical compactification, where a matter source has the dustlike equation of state $\hat{p}_0 = 0$ in the external (our) space and an arbitrary equation of state $\hat{p}_1 = \Omega\hat{\varepsilon}$ in the internal space, satisfy asymptotically (in the region of the negligibly small Yukawa interaction) the gravitational experiments [9]. However, in all models with $\Omega \neq -1/2$ a gravitating matter source acquires effective relativistic pressure in the external (our) space. Obviously, this situation can not be acceptable for ordinary astrophysical objects such as our Sun. Therefore, in spite of the agreement (asymptotical) with the gravitational experiments, such models fail with the observations. Only in the case of tension $\Omega = -1/2$, a matter source remains dustlike in the external space. Therefore, tension also plays a crucial role in models with spherical compactification as in the case of toroidal compactification. The only problem is to explain the physical origin of tension for ordinary astrophysical objects. In this paper we present a brief review of our results, which prejudice KK models.

2 Linear models with toroidal compactification

In this section we analyze very briefly linear with respect to the scalar curvature R KK models with toroidal compactification of the internal spaces.

2.1 Pointlike mass

First, we investigate a model with a pointlike massive gravitating source in the weak field limit. It means that the gravitational field is weak, i.e. the metrics is only slightly perturbed from its flat spacetime value:

$$1g_{ik} \approx \eta_{ik} + h_{ik}\,. \tag{1}$$

Here, the metric correction terms $h_{ik} \sim O(1/c^2)$, where c is the speed of light, $i, k = 0, 1, \ldots, D$, and D is a total number of the spatial dimensions. In the weak field limit, the only nonzero component of the energy-momentum tensor for a pointlike mass at rest is $T_{00} \approx \rho_D c^2 \sim O(c^2)$. ρ_D is a D-dimensional rest mass density, and for a pointlike mass m we have $\rho_D = m\delta(\mathbf{r}_D)$. Usually, we deal with the case of matter sources, which are uniformly smeared over the extra dimensions [11]. In this case, the metric coefficients may depend only on coordinates of the external space and the nonrelativistic three-dimensional rest mass density ρ_3 is connected with the D-dimensional one as follows: $\rho_D = \rho_3/V_{D'} = m\delta(\mathbf{r}_3)/V_{D'}$, where $D' = D - 3$ is a total number of the extra dimensions and $V_{D'}$ is a total volume of the unperturbed internal spaces. For example, if a_i are periods of tori, then we have $V_{D'} = \prod_{i=1}^{D'} a_i$. The Einstein equation

$$2R_{ik} = \frac{2S_D\tilde{G}_{\mathcal{D}}}{c^4}\left(T_{ik} - \frac{1}{D-1}g_{ik}T\right)\,, \tag{2}$$

where $S_D = 2\pi^{D/2}/\Gamma(D/2)$ is a total solid angle (a surface area of the $(D-1)$-dimensional sphere of the unit radius) and $\tilde{G}_{\mathcal{D}}$ is the gravitational constant in the

$(\mathcal{D}=D+1)$-dimensional spacetime, is reduced to a system of linearized equations with the following nonzero solutions [3]:

$$
\begin{aligned}
3h_{00} &= -\frac{2(D-2)}{D-1}\frac{2G_N m}{c^2 r_3}\,, & & (3)\\
h_{\alpha\alpha} &= -\frac{2}{D-1}\frac{2G_N m}{c^2 r_3}\,, \quad \alpha = 1,2,3\,, & & (4)\\
h_{\mu\mu} &= -\frac{2}{D-1}\frac{2G_N m}{c^2 r_3}\,, \quad \mu = 4,5,\ldots,D\,, & & (5)
\end{aligned}
$$

where we introduce the Newtonian gravitational constant

$$6G_N = \frac{S_D}{4\pi}\frac{\tilde{G}_{\mathcal{D}}}{V_{D'}}\,. \tag{6}$$

Hereafter, the Latin indices $i,k=0,\ldots,D$, the Greek indices $\alpha,\beta=1,2,3$ and the Greek indices $\mu,\nu=4,5,\ldots,D$.
It is well known that in order to satisfy the gravitational experiments (such as the deflection of light and the time delay of radar echoes) at the level of accuracy of general relativity, the metric coefficients h_{00} and $h_{\alpha\alpha}$ should coincide with each other. However, Eqs. (3) and (4) show that for the considered model $h_{00}/h_{\alpha\alpha} = D-2$ and this ratio does not depend on the size of the internal space. So, we can not make it equal to a unity. This is the first problematic aspect to be mentioned. On the other hand, h_{00} defines the nonrelativistic gravitational potential: $h_{00} = 2\varphi/c^2$. For example, in general relativity $\tilde{h}_{00} = 2\varphi_N/c^2 = -2G_N m/\left(c^2 r_3\right)$ and $\tilde{h}_{00} = \tilde{h}_{\alpha\alpha}$. In our case, the Newtonian gravitational potential acquires a prefactor $2(D-2)/(D-1)$.

2.2 Latent solitons, black strings and black branes

Above, we consider the case of a pointlike mass with the dustlike equations of state in all spatial dimensions. However, there is a class of exact static spherically symmetric (with respect to the external space) soliton solutions (see, e.g., [4, 5]) with nonzero equations of state in the extra dimensions:

$$10T_{\mu\mu} \approx \omega_\mu \frac{1}{V_{D'}}\rho_3(\mathbf{r}_3)c^2 \approx \omega_\mu T_{00}, \quad \mu = 4,5,\ldots,D\,. \tag{7}$$

These solutions are defined by the parameters γ_μ, which are connected with the equation of state parameters ω_μ [5]:

$$11\omega_\mu = \frac{\gamma_\mu - 1 + \tau}{2 - \tau}\,, \tag{8}$$

where $\tau = \sum_{\mu=4}^{D} \gamma_\mu$. In the weak field limit, the metric correction terms for these solutions read [5]:

$$12h_{00} = -\frac{2(D-2)}{D-1}\frac{2G_N m}{c^2 r_3} - \frac{2\Omega}{D-1}\frac{2G_N m}{c^2 r_3}\,, \tag{9}$$

$$h_{\alpha\alpha} = -\frac{2}{D-1}\frac{2G_N m}{c^2 r_3} + \frac{2\Omega}{D-1}\frac{2G_N m}{c^2 r_3}\,, \tag{10}$$

$$h_{\mu\mu} = -\frac{2}{D-1}\frac{2G_N m}{c^2 r_3} - 2\left(\omega_\mu - \frac{\Omega}{D-1}\right)\frac{2G_N m}{c^2 r_3}\,, \tag{11}$$

where $\Omega = \sum_{\mu=4}^{D} \omega_\mu$. Obviously, the second terms in Eqs. (12)-(14) are due to nonzero equations of state in the extra dimensions. Solutions with

$$16\Omega = -\frac{D-3}{2} \tag{12}$$

we call latent solitons. In this case h_{00} and $h_{\alpha\alpha}$ exactly coincide with the Newtonian expressions and with each other: $h_{00} = \tilde{h}_{00} = h_{\alpha\alpha} = \tilde{h}_{\alpha\alpha}$. Black stings ($D = 4$) and black branes ($D > 4$) are particular cases of the latent solitons with the same equations of state ($\omega_\mu = -1/2$) in all extra dimensions. For these particular cases, each $h_{\mu\mu} = 0$, i.e. each scale factor of the internal spaces is constant. Since $\Omega < 0$, all or some of ω_μ should be negative. This is the second problematic aspect to be mentioned.

3 Nonlinear models with toroidal compactification

In this section we briefly analyze nonlinear $f(R)$ KK models with toroidal compactification of the internal spaces. In the case of nonzero equations of state (10) in the extra dimensions, the correction terms read [7]

$$18h_{00} = -\frac{2(D-2)}{D-1}\frac{2G_N m}{c^2 r_3} - \frac{2\Omega}{D-1}\frac{2G_N m}{c^2 r_3} - \frac{4a}{D-1}R\,, \tag{13}$$

$$h_{\alpha\alpha} = -\frac{2}{D-1}\frac{2G_N m}{c^2 r_3} + \frac{2\Omega}{D-1}\frac{2G_N m}{c^2 r_3} + \frac{4a}{D-1}R\,, \tag{14}$$

$$h_{\mu\mu} = -\frac{2}{D-1}\frac{2G_N m}{c^2 r_3} - 2\left(\omega_\mu - \frac{\Omega}{D-1}\right)\frac{2G_N m}{c^2 r_3} + \frac{4a}{D-1}R\,, \tag{15}$$

where $a \equiv (1/2)f''(0)$ and the scalar curvature

$$21R = \frac{1-\Omega}{2aD}\frac{2G_N m}{c^2 r_3}\exp\left[-\left(\frac{D-1}{4|a|D}\right)^{1/2} r_3\right]\,. \tag{16}$$

It is clear that the second terms in (18)-(20) take place due to the nonzero equations of state in the extra dimensions ($\omega_\mu, \Omega \neq 0$) and the third terms originate from the nonlinearity of the model ($a \neq 0$). The Eq. (21) shows that nonlinearity generates the Yukawa interaction with the mass $[(D-1)/(4|a|D)]^{1/2}$ [6].

3.1 Pointlike mass

Let us first consider the case of a pointlike gravitating source at rest, i.e. with the dustlike equation of state in all spatial dimensions: $\omega_\mu = 0,\ \mu = 4, 5, \ldots, D \Rightarrow \Omega = 0$. In this case, the second terms in Eqs. (18)-(20) disappear and we arrive at equations of the subsection 2.1 with the admixture of the Yukawa interaction. Similar to the linear case, this situation also contradicts the observational data [6].

3.2 Asymptotic latent solitons, black strings and black branes

Now, we want to consider solutions, which are in agreement with the gravitational tests (the deflection of light and the time delay of radar echoes). In the case of the linear models, it takes place for the latent solitons. Therefore, we should take into account tension in the internal spaces: $\Omega \neq 0$. Unfortunately, it is impossible to get exact soliton solutions in the case of an arbitrary function $f(R)$. Therefore, in the paper [7], we proposed two types of asymptotic solutions with $h_{00} = \tilde{h}_{00}$ and $h_{\alpha\alpha} = \tilde{h}_{\alpha\alpha} \;\Rightarrow\; h_{00} = h_{\alpha\alpha}$. These asymptotic latent solitons exist in the regions $r_3 \gg \sqrt{|a|}$ and $r_3 \ll \sqrt{|a|}$. Let us consider these two regions separately.

3.2.1 $r_3 \gg \sqrt{|a|}$.

In this asymptotic region, the exponent in (21) is negligible and we can drop the third terms in the Eqs. (18)-(20). In other words, the effect of nonlinearity is negligibly small and we arrive at the case of the subsection 2.2. Here, $\Omega = -(D-3)/2$.

3.2.2 $r_3 \ll \sqrt{|a|}$.

In this case, we can replace the exponent in (21) by a unity. Here, the effect of nonlinearity is not negligible. After substitution (21) into (18) and (19) we get

$$22h_{00} = -\frac{2G_N m}{c^2 r_3}\left[1 + \frac{D-3}{D-1} + \frac{2\Omega}{D-1} + \frac{2(1-\Omega)}{D(D-1)}\right], \tag{17}$$

$$h_{\alpha\alpha} = -\frac{2G_N m}{c^2 r_3}\left[1 - \frac{D-3}{D-1} - \frac{2\Omega}{D-1} - \frac{2(1-\Omega)}{D(D-1)}\right]. \tag{18}$$

It can be easily seen that for

$$24\Omega = -\frac{D-2}{2} \tag{19}$$

we have $h_{00} = \tilde{h}_{00}$ and $h_{\alpha\alpha} = \tilde{h}_{\alpha\alpha} \ \Rightarrow \ h_{00} = h_{\alpha\alpha}$.

4 Spherical compactification of the internal space

In this section we consider a model with spherical compactification of the internal space, where the background metrics is defined on a product manifold $M_4 \times M_2$. Here, M_4 describes the external four-dimensional flat spacetime and M_2 corresponds to a two-dimensional sphere with the radius (the internal space scale factor) a. To create such spacetime with the curved internal space, we should introduce the background matter. As we have shown in our papers [8, 9], this matter simulates a perfect fluid with the vacuum equation of state in the external space. In the internal space (the two-sphere) the parameter of the equation of state is

$$27\omega_1 = \frac{\Lambda_6}{1/[(2S_5\tilde{G}_6/c^4)a^2] - \Lambda_6}, \tag{20}$$

where Λ_6 is a bare multidimensional cosmological constant. Different forms of matter can simulate such perfect fluid. For example, $\omega_1 = 1$ and $\omega_1 = 2$ correspond to the Freund-Rubin mechanism of compactification and the Casimir effect, respectively. In the case $\Lambda_6 = 0$ we get the dustlike equation of state with the parameter $\omega_1 = 0$. For $\omega_1 > 0$ the internal space is stabilized [9]. In this model, the Eq. (6) takes the form $4\pi G_N = S_5\tilde{G}_6/(4\pi a^2)$, where we take into account that the volume of the internal space $V_{D'} \equiv V_2 = 4\pi a^2$. The background metrics and matter are perturbed by a pointlike massive source with dustlike equations of state in all spatial dimensions. In the case $\omega_1 > 0$ the correction terms read [9]

$$28h_{00} = -\frac{2G_N m}{c^2 r_3} - \frac{G_N m}{c^2 r_3}\exp\left(-\frac{\sqrt{\omega_1}}{a}r_3\right), \tag{21}$$

$$h_{\alpha\alpha} = -\frac{2G_N m}{c^2 r_3} + \frac{G_N m}{c^2 r_3}\exp\left(-\frac{\sqrt{\omega_1}}{a}r_3\right). \tag{22}$$

So, we have the Yukawa interaction with the mass squared ω_1/a^2. Obviously, the admixture of such interaction to h_{00} and $h_{\alpha\alpha}$ is negligible for sufficiently large Yukawa masses. Exactly this situation takes place for the gravitational tests in the Solar system [9]. Here, $h_{00} = h_{\alpha\alpha}$ with very high accuracy, and we achieve good agreement with the gravitational tests for the considered model. Obviously, models with $\omega_1 \le 0$ do not satisfy the experimental data.

However, even for $\omega_1 > 0$ we have the third problematic aspect to be mentioned. As we pointed out in the papers [8, 9], the matter source in the KK models with spherical compactification should consist of two parts. First, it is the homogeneous perfect fluid which provides spherical compactification of the internal space. Second, it is the gravitating mass, which is finite (i.e. pointlike) in the external space and uniformly smeared over the internal space. The total energy-momentum tensor is the sum of these parts with the following nonzero components:

$$a1 \qquad T_0^0 \approx \bar{\varepsilon} + \varepsilon^1 + \hat{\rho}(\mathbf{r}_3)c^2\,, T_\alpha^\alpha \approx \bar{\varepsilon} + \varepsilon^1\,, \quad \alpha = 1, 2, 3\,, \tag{23}$$

$$T_4^4 = T_5^5 \approx -\omega_1\bar{\varepsilon} - \omega_1\varepsilon^1 - \Omega\hat{\rho}(\mathbf{r}_3)c^2\,, \tag{24}$$

where $\bar{\varepsilon}$ is the energy density of the homogeneous perfect fluid, $\hat{\rho}(\mathbf{r}_3)$ is the rest mass density of the finite gravitating mass and ε^1 is the excitation of the background matter energy density by this mass. The background matter is fine-tuned with the radius a of the two-sphere: $\bar{\varepsilon} = [(1+\omega_1)\kappa_6 a^2]^{-1}$, and a free parameter

ω_1 defines the equation of state of this matter in the internal space. The model may also include a six-dimensional cosmological constant Λ_6, which is fine-tuned with the parameters of the model: $\Lambda_6 = \omega_1 \bar{\varepsilon}$. This bare cosmological constant is absent if $\omega_1 = 0$. It is assumed that the gravitating matter source has the dustlike equation of state in the external (our) space $\hat{p}_0 = 0$ and an arbitrary equation of state $\hat{p}_1 \approx \Omega \hat{\rho}(\mathbf{r}_3) c^2$ in the internal space. We also suppose that this source is uniformly smeared over the internal space: $\hat{\rho}(\mathbf{r}_3) = \hat{\rho}_3(\mathbf{r}_3)/V_2$. In the case of a pointlike mass in the external space $\hat{\rho}_3(\mathbf{r}_3) = m\delta(\mathbf{r}_3)$.
The metrics for the considered model in isotropic coordinates takes the form (see for details [8, 9])

$$ds^2 = Ac^2dt^2 + Bdx^2 + Cdy^2 + Ddz^2 + E(d\xi^2 + \sin^2 \xi d\eta^2)$$

with $A \approx 1 + A^1(r_3)$, $B \approx -1 + B^1(r_3)$, $C \approx -1 + C^1(r_3)$, $D \approx -1 + D^1(r_3)$, $E \approx -a^2 + E^1(r_3)$, where all metric perturbations A^1, B^1, C^1, D^1, E^1 are of the order $O(1/c^2)$ and can be found with the help of the Einstein equations. They read

$$a4A^1 = \frac{2\varphi_N}{c^2} + \frac{E^1}{a^2}\, , B^1 = C^1 = D^1 = \frac{2\varphi_N}{c^2} - \frac{E^1}{a^2}\, , \tag{25}$$

$$E^1 = a^2 \frac{\varphi_N}{c^2}\, (1 + 2\Omega)\, e^{-r_3/\lambda}, \quad \lambda \equiv a/\sqrt{\omega_1}\, , \tag{26}$$

where the Newton's potential is $\varphi_N = -G_N m/r_3$. The solution (26) takes place for $\omega_1 > 0$. In the opposite case $\omega_1 < 0$, we get the nonphysical oscillating solution. If $\Omega \neq -1/2$, the Eq. (26) demonstrates that the Yukawa interaction is generated. The admixture of such interaction to A^1, B^1, C^1, D^1 is negligible at distances $r_3 \gg \lambda$ (i.e. for the large Yukawa mass $\sqrt{\omega_1}/a$), and we achieve good agreement with the gravitational tests in this region. Exactly this situation takes place in the Solar system [8].
The Einstein equations also lead to the following important relationship: $\varepsilon^1 = E^1/\left(\kappa_6 a^4\right)$. The Eq. (26) shows that this background perturbation is localized around the gravitating mass and falls off exponentially with the distance r_3 from it. Therefore, the bare gravitating mass is covered by this "coat". For an external observer, this coated gravitating mass is characterized by the effective energy-

momentum tensor with the following nonzero components:

$$\begin{aligned}
a8T_0^{0(eff)} &\approx \varepsilon^1 + \hat{\rho}(\mathbf{r}_3)c^2 = -(1+2\Omega)\frac{mc^2}{2V_2^2 r_3}\exp\left(-\frac{\sqrt{\omega_1}}{a}r_3\right) + \frac{1}{V_2}mc^2\delta(\mathbf{r}_3), \\
T_\alpha^{\alpha(eff)} &\approx \varepsilon^1 = -(1+2\Omega)\frac{mc^2}{2V_2^2 r_3}\exp\left(-\frac{\sqrt{\omega_1}}{a}r_3\right), \quad \alpha = 1,2,3\,, \qquad (27) \\
T_4^{4(eff)} &= T_5^{5(eff)} \approx -\omega_1\varepsilon^1 - \Omega\hat{\rho}(\mathbf{r}_3)c^2 \\
&= (1+2\Omega)\frac{\omega_1 mc^2}{2V_2^2 r_3}\exp\left(-\frac{\sqrt{\omega_1}}{a}r_3\right) - \frac{\Omega}{V_2}mc^2\delta(\mathbf{r}_3)\,.
\end{aligned}$$

These components define the effective energy density and pressure of the coated gravitating mass. From the Eq. (27) we conclude that this mass acquires relativistic pressure $\hat{p}_0^{(eff)} = -T_\alpha^{\alpha(eff)}$ in the external space. We see that the effective energy density $\hat{\varepsilon}^{(eff)} = T_0^{0(eff)}$ and effective pressure in the external (our) space $\hat{p}_0^{(eff)} = -T_\alpha^{\alpha(eff)}$ depend on the parameter Ω, which defines the equation of state of the bare gravitating mass in the internal space. It can be easily seen that the equality $\Omega = -1/2$ is the only possibility to achieve $\hat{p}_0^{(eff)} = 0$ for our model. It means that the bare gravitating mass should have tension with the equation of state $\hat{p}_1 = -\hat{\varepsilon}/2$ in the internal space. Then, the effective and bare energy densities coincide with each other and the gravitating mass remains pressureless in our space. In the internal space the gravitating mass still has tension with the parameter of state $-1/2$. Therefore, tension plays a crucial role for models with spherical compactification.

5 Conclusions

In this paper we investigated failure with the gravitational tests for (linear as well as nonlinear with respect to the scalar curvature) KK models with toroidal and spherical compactifications of the internal spaces, when the gravitating matter source is dustlike with respect to all spatial dimensions. We demonstrated that introduction of tension (negative relativistic pressure in the internal spaces) can save the situation, but only in mathematical sense.

References

1. J. Polchinski, *String Theory, Volume 2: Superstring Theory and Beyond*, (Cambridge University Press, Cambridge, 1998).

2. C.M. Will, *Theory and Experiment in Gravitational Physics*, (Cambridge University Press, Cambridge, 2000).

3. M. Eingorn and A. Zhuk, Class. Quant. Grav. **27**, 205014 (2010); (arXiv:gr-qc/1003.5690).

4. M. Eingorn and A. Zhuk, Phys. Rev. D **83**, 044005 (2011); (arXiv:gr-qc/1010.5740).

5. M. Eingorn, O. de Medeiros, L. Crispino and A. Zhuk, Phys. Rev. D **84**, 024031 (2011); (arXiv:gr-qc/1101.3910).

6. M. Eingorn and A. Zhuk, Phys. Rev. D **84**, 024023 (2011); (arXiv:gr-qc/1104.1456).

7. M. Eingorn and A. Zhuk, Phys. Rev. D **85**, 064030 (2012); (arXiv:gr-qc/1112.1539).

8. A. Chopovsky, M. Eingorn and A. Zhuk, Phys. Rev. D **85**, 064028 (2012); (arXiv:gr-qc/1107.3388).

9. A. Chopovsky, M. Eingorn and A. Zhuk, *Exact and asymptotic black branes with spherical compactification*, submitted to Phys. Rev. D (2012); (arXiv:gr-qc/1202.2677).

10. L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields, Fourth Edition: Volume 2 (Course of Theoretical Physics Series)*, (Pergamon Press, Oxford, 2000).

11. M. Eingorn and A. Zhuk, Class. Quant. Grav. **27**, 055002 (2010); (arXiv:gr-qc/0910.3507).

Seyed Hossein Fakhr[13], *M. Eingorn, A. Zhuk*

Spherical compactification of three and more extra dimensions in Kaluza-Klein geometries[14]

Odessa National University named after I.I. Mechnikov,
2, Dvoryanskaya street, Odessa 65082, Ukraine

In our recent papers we have investigated the Kaluza-Klein model with spherical compactification of two extra dimensions and a non-dust-like matter source of the gravitational field concerning its compatibility with the experimental data of astronomical observations (namely, the time delay of radar echoes and the light deflection). We have found conditions for the parameters of the model, under which it satisfies the gravitational tests. Now we generalize our previous results to the

[13]This research was co-financed by the Austrian Academy of Sciences in the framework of the collaboration with the National Academy of Sciences of Ukraine on *Modern Problems in Astroparticle Physics*.

[14]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

case of spherical compactification of three and more extra dimensions in Kaluza-Klein geometries. In detail we consider the internal space in the form of the three-sphere and arrive at corresponding approximate solutions, satisfying the experimental data.

1 Introduction

Any physical theory is correct until it does not contradict the experimental data. Obviously, the Kaluza-Klein model is no exception to this rule. There is a number of the well-known gravitational experiments in the Solar system, e.g., the deflection of light and the time delay of radar echoes. In the weak field limit, all these effects can be expressed via parameterized post-Newtonian (PPN) parameters β and γ [1,2]. These parameters take different values in different theories of gravity. There are strict experimental restrictions on these parameters [3–6]. The tightest constraint on γ comes from the Shapiro time-delay experiment using the Cassini spacecraft: $\gamma - 1 = (2.1 \pm 2.3) \times 10^{-5}$. General Relativity is in good agreement with all gravitational experiments [7]. Here, the PPN parameters $\beta = 1$ and $\gamma = 1$. The Kaluza-Klein model should also be tested by the above-mentioned experiments.

In our previous papers [8, 9] we have investigated this problem in the case of spherical compactification of the 2-dimensional internal space. In contrast to the case of toroidal compactification the background metrics was not flat but had topology $\mathbb{R} \times \mathbb{R}^3 \times S^2$. To make the internal space curved, we introduced background matter. We demonstrated that this matter can be simulated by a perfect fluid with the vacuum equation of state in the external space and an arbitrary equation of state with the parameter ω_1 in the internal space. Our model contained also a bare multidimensional cosmological constant Λ_6. We perturbed this background by a point-like mass and calculated the perturbed metric coefficients in the weak field approximation up to the order $1/c^2$. In the case $\omega_1 > 0$, these metric coefficients acquired the Yukawa correction terms with respect to the usual Newtonian gravitational potential. The Yukawa interaction was characterized by its mass which was proportional to $\sqrt{\omega_1}$. The terrestrial inverse square law experiments [10] restrict such corrections and provide strong bounds on parameters of the model, e.g., on the radius of the internal two-sphere. This radius is in many orders of magnitude less than the radius of the Sun. Obviously, in the Solar system we could drop the Yukawa correction terms with very high accuracy, and the parameterized post-Newtonian parameter γ was equal to 1 similar to General Relativity. Therefore, our model satisfied the gravitational experiments (the deflection of light and the time delay of radar echoes) at the same level of accuracy as General Relativity.

In the present paper we generalize our previous results to the case of spherical compactification of three and more extra dimensions (and a non-dust-like matter source of the gravitational field as perturbation).

2 Background metrics and matter required for spherical compactification

Let us start from the 7-dimensional diagonal metrics

$$ds^2 = Adt^2 + Bdx^2 + Cdy^2 + Ddz^2 + Ed\xi^2 + Fd\eta^2 + Gd\zeta^2\,,$$

where A, B, C, D, E, F, G are functions of $t, x, y, z, \xi, \eta, \zeta$, and find the corresponding diagonal covariant Ricci tensor components for the background metric coefficients

$$A = 1, \quad B = C = D = -1\,, \quad E = -a^2 \sin^2\zeta \sin^2\eta, F = -a^2 \sin^2\zeta, G = -a^2\,,$$

where a is a radius of the three-sphere, representing the internal space. For these values of the metric coefficients we get

$$R_{00} = R_{11} = R_{22} = R_{33} = 0\,, \quad R_{44} = 2\sin^2\eta \sin^2\zeta = -\frac{2}{a^2}E = -\frac{2}{a^2}g_{44}\,,$$
$$R_{55} = 2\sin^2\zeta = -\frac{2}{a^2}F = -\frac{2}{a^2}g_{55}\,, \quad R_{66} = 2 = -\frac{2}{a^2}G = -\frac{2}{a^2}g_{66}\,.$$

The corresponding scalar curvature reads

$$R = R_{ik}g^{ik} = R_{44}g^{44} + R_{55}g^{55} + R_{66}g^{66} = -\frac{6}{a^2}\,.$$

Obviously, in the case of the n-dimensional internal space in the form of the n-sphere

$$R_{\mu\mu} = -\frac{n-1}{a^2}g_{\mu\mu}, \quad \mu = 3+1, 3+2, ..., 3+n\,,$$

and

$$R = -\frac{n(n-1)}{a^2}\,.$$

In our present particular case $n = 3$. In the previous particular case [8, 9] $n = 2$.

The Einstein equation, written down in the form

$$R_{ik} - \frac{1}{2}Rg_{ik} = \kappa T_{ik} + \kappa\Lambda_7 g_{ik}, \quad \kappa = \mathrm{const} > 0\,,$$

where Λ_7 is a bare 7-dimensional cosmological constant, allows to find immediately the corresponding covariant energy-momentum tensor components for matter required for spherical compactification:

$$\tilde{T}_{00} = g_{00}\left(\frac{3}{\kappa a^2} - \Lambda_7\right) = \frac{3}{\kappa a^2} - \Lambda_7\,,$$
$$\tilde{T}_{11} = \tilde{T}_{22} = \tilde{T}_{33} = g_{\alpha\alpha}\left(\frac{3}{\kappa a^2} - \Lambda_7\right) = -\frac{3}{\kappa a^2} + \Lambda_7, \;\; \alpha = 1, 2, 3\,,$$
$$\tilde{T}_{44} = g_{44}\left(\frac{1}{\kappa a^2} - \Lambda_7\right) = -\frac{\sin^2\zeta\sin^2\eta}{\kappa} + \Lambda_7 a^2\sin^2\zeta\sin^2\eta\,,$$
$$\tilde{T}_{55} = g_{55}\left(\frac{1}{\kappa a^2} - \Lambda_7\right) = -\frac{\sin^2\zeta}{\kappa} + \Lambda_7 a^2\sin^2\zeta\,,$$
$$\tilde{T}_{66} = g_{66}\left(\frac{1}{\kappa a^2} - \Lambda_7\right) = -\frac{1}{\kappa} + \Lambda_7 a^2\,.$$

Obviously, in the general case

$$\tilde{T}_{00} = g_{00}\left(\frac{n(n-1)}{2\kappa a^2} - \Lambda_{4+n}\right) = \frac{n(n-1)}{2\kappa a^2} - \Lambda_{4+n}\,, \quad \alpha = 1, 2, 3\,,$$
$$\tilde{T}_{11} = \tilde{T}_{22} = \tilde{T}_{33} = g_{\alpha\alpha}\left(\frac{n(n-1)}{2\kappa a^2} - \Lambda_{4+n}\right) = -\frac{n(n-1)}{2\kappa a^2} + \Lambda_{4+n},$$
$$\tilde{T}_{\mu\mu} = g_{\mu\mu}\left(\frac{(n-1)(n-2)}{2\kappa a^2} - \Lambda_{4+n}\right), \quad \mu = 3+1, 3+2, ..., 3+n\,,$$

where Λ_{4+n} is a bare $(4+n)$-dimensional cosmological constant.

The transition $\tilde{T}_{ik} \;\; \to \;\; \tilde{T}_k^i$ allows to find background energy density and pressures as well as equations of state in external and internal spaces:

$$
\begin{aligned}
&\tilde{T}_0^0 = \tilde{T}_{00}g^{00} = \frac{3}{\kappa a^2} - \Lambda_7 = \bar{\varepsilon}, \quad \tilde{T}_1^1 = \tilde{T}_2^2 = \tilde{T}_3^3 = \frac{3}{\kappa a^2} - \Lambda_7 = -\bar{p}_0, \\
&\tilde{T}_4^4 = \tilde{T}_5^5 = \tilde{T}_6^6 = \frac{1}{\kappa a^2} - \Lambda_7 = -\bar{p}_1\,, \quad \bar{p}_0 = \omega_0\bar{\varepsilon}, \quad \omega_0 = -1, \\
&\bar{p}_1 = \omega_1\bar{\varepsilon} \quad \Rightarrow \omega_1 = \frac{\bar{p}_1}{\bar{\varepsilon}} = \frac{\Lambda_7\kappa a^2 - 1}{3 - \Lambda_7\kappa a^2}, \quad \Lambda_7 = \frac{3\omega_1 + 1}{\omega_1 + 1}\frac{1}{\kappa a^2}\,.
\end{aligned}
$$

In the general case

$$
\begin{aligned}
&\tilde{T}_0^0 = \frac{n(n-1)}{2\kappa a^2} - \Lambda_{4+n} = \bar{\varepsilon}, \quad \tilde{T}_1^1 = \tilde{T}_2^2 = \tilde{T}_3^3 = \frac{n(n-1)}{2\kappa a^2} - \Lambda_{4+n} = -\bar{p}_0\,, \\
&\tilde{T}_\mu^\mu = \frac{(n-1)(n-2)}{2\kappa a^2} - \Lambda_{4+n} = -\bar{p}_1, \quad \mu = 3+1, 3+2, ..., 3+n\,, \\
&\Lambda_{4+n} = \frac{n(n-1)\omega_1 + (n-1)(n-2)}{2(\omega_1+1)}\frac{1}{\kappa a^2}\,, \quad \bar{p}_0 = \omega_0\bar{\varepsilon}, \quad \omega_0 = -1, \\
&\bar{p}_1 = \omega_1\bar{\varepsilon} \quad \Rightarrow \omega_1 = \frac{\bar{p}_1}{\bar{\varepsilon}} = \frac{2\Lambda_{4+n}\kappa a^2 - (n-1)(n-2)}{n(n-1) - 2\Lambda_{4+n}\kappa a^2}\,.
\end{aligned}
$$

Thus, the background matter has the vacuum equation of state in the external space and an arbitrary equation of state in the internal space.

3 Perturbation in the form of a non-dust-like matter source

Now let us turn to the first-order approximation (the weak field limit) and find the corresponding approximate expressions for the metric coefficients, when perturbation represents a point-like mass at rest, uniformly smeared over the internal space. For

$$
\begin{aligned}
&g_{00} = A \approx 1 + A^1(\mathbf{r}_3), \quad g_{11} = B \approx -1 + B^1(\mathbf{r}_3), \quad g_{22} = C \approx -1 + C^1(\mathbf{r}_3), \\
&g_{33} = D \approx -1 + D^1(\mathbf{r}_3)\,, \quad g_{44} = E \approx [-a^2 + G^1(\mathbf{r}_3)]\sin^2\zeta\sin^2\eta\,, \\
&g_{55} = F \approx [-a^2 + G^1(\mathbf{r}_3)]\sin^2\zeta, \quad g_{66} = G \approx -a^2 + G^1(\mathbf{r}_3)
\end{aligned}
$$

let us write down the corresponding approximate expressions for the diagonal covariant Ricci tensor components:

$$
\begin{aligned}
&R_{00} \approx \frac{1}{2}\triangle A^1\,, \quad R_{11} \approx \frac{1}{2}\triangle B^1 + \frac{1}{2}\left(-A^1 - B^1 + C^1 + D^1 + \frac{3G^1}{a^2}\right)_{xx}\,, \\
&R_{22} \approx \frac{1}{2}\triangle C^1 + \frac{1}{2}\left(-A^1 + B^1 - C^1 + D^1 + \frac{3G^1}{a^2}\right)_{yy}\,, \\
&R_{33} \approx \frac{1}{2}\triangle D^1 + \frac{1}{2}\left(-A^1 + B^1 + C^1 - D^1 + \frac{3G^1}{a^2}\right)_{zz}\,, \\
&R_{44} \approx 2\sin^2\eta\sin^2\zeta + \frac{1}{2}\triangle E^1 \approx R_{66}\sin^2\eta\sin^2\zeta\,, \\
&R_{55} \approx 2\sin^2\zeta + \frac{1}{2}\triangle F^1 \approx R_{66}\sin^2\zeta\,, \quad R_{66} \approx 2 + \frac{1}{2}\triangle G^1\,.
\end{aligned}
$$

Now let us rewrite the previously used Einstein equation in the other form:

$$
R_{ik} = \kappa\left(T_{ik} - \frac{1}{5}Tg_{ik} - \frac{2}{5}\Lambda_7 g_{ik}\right)\,.
$$

In the general case ($\Lambda_7 \to \Lambda_{4+n}$, $g_{ik}g^{ik} = 4+n$)

$$
R_{ik} = \kappa\left(T_{ik} - \frac{1}{2+n}Tg_{ik} - \frac{2}{2+n}\Lambda_{4+n} g_{ik}\right)\,.
$$

In order to solve this equation, let us find the covariant components of the total energy-momentum tensor $T_{ik} = \tilde{T}_{ik} + \hat{T}_{ik}$, where the components $\hat{T}_{ik}$ correspond to the perturbation. The approximate expressions for the covariant energy-momentum tensor components of the background matter in presence of the gravitating mass read

$$
\begin{aligned}
&\tilde{T}_{00} \approx \left(\bar{\varepsilon} + \varepsilon^1\right) g_{00} \approx \left(\bar{\varepsilon} + \varepsilon^1\right)\left(1 + A^1\right) \approx \bar{\varepsilon} + \varepsilon^1 + \bar{\varepsilon}A^1\,, \\
&\tilde{T}_{11} \approx \left(\bar{\varepsilon} + \varepsilon^1\right) g_{11} \approx \left(\bar{\varepsilon} + \varepsilon^1\right)\left(-1 + B^1\right) \approx -\bar{\varepsilon} - \varepsilon^1 + \bar{\varepsilon}B^1\,, \\
&\tilde{T}_{22} \approx \left(\bar{\varepsilon} + \varepsilon^1\right) g_{22} \approx \left(\bar{\varepsilon} + \varepsilon^1\right)\left(-1 + C^1\right) \approx -\bar{\varepsilon} - \varepsilon^1 + \bar{\varepsilon}C^1\,, \\
&\tilde{T}_{33} \approx \left(\bar{\varepsilon} + \varepsilon^1\right) g_{33} \approx \left(\bar{\varepsilon} + \varepsilon^1\right)\left(-1 + D^1\right) \approx -\bar{\varepsilon} - \varepsilon^1 + \bar{\varepsilon}D^1\,, \\
&\tilde{T}_{44} = \tilde{T}_{66}\sin^2\eta\sin^2\zeta, \quad \tilde{T}_{55} = \tilde{T}_{66}\sin^2\zeta\,, \\
&\tilde{T}_{66} \approx -\omega_1\left(\bar{\varepsilon} + \varepsilon^1\right) g_{66} \approx -\omega_1\left(\bar{\varepsilon} + \varepsilon^1\right)\left(-a^2 + G^1\right) \approx \omega_1 a^2\bar{\varepsilon} + \omega_1 a^2\varepsilon^1 - \omega_1\bar{\varepsilon}G^1 \\
&\Rightarrow \quad \tilde{T} \approx \bar{\varepsilon} + \varepsilon^1 + 3\left(\bar{\varepsilon} + \varepsilon^1\right) - 3\omega_1\left(\bar{\varepsilon} + \varepsilon^1\right) = 4\bar{\varepsilon} - 3\omega_1\bar{\varepsilon} + 4\varepsilon^1 - 3\omega_1\varepsilon^1\,.
\end{aligned}
$$

In the general case instead of the last two approximate expressions we have respectively:

$$\tilde{T}_{DD} \approx -\omega_1\left(\bar{\varepsilon}+\varepsilon^1\right)g_{DD} \approx -\omega_1\left(\bar{\varepsilon}+\varepsilon^1\right)\left(-a^2+G^1\right) \approx \omega_1 a^2\bar{\varepsilon}+\omega_1 a^2\varepsilon^1-\omega_1\bar{\varepsilon}G^1\,,$$
$$\tilde{T} \approx \bar{\varepsilon}+\varepsilon^1+3\left(\bar{\varepsilon}+\varepsilon^1\right)-n\omega_1\left(\bar{\varepsilon}+\varepsilon^1\right) = 4\bar{\varepsilon}-n\omega_1\bar{\varepsilon}+4\varepsilon^1-n\omega_1\varepsilon^1\,.$$

Here and in what follows $D = 3+n$ is the total space dimensionality and we keep the convenient designation $g_{DD} = G$.

Further, the approximate expressions for the covariant energy-momentum tensor components of the gravitating mass read

$$\hat{T}_{00} \approx \hat{\rho}c^2, \quad \hat{T}_{\alpha\alpha} = 0, \quad \hat{T}_{44} = \hat{T}_{66}\sin^2\eta\sin^2\zeta, \quad \hat{T}_{55} = \hat{T}_{66}\sin^2\zeta, \;\; \alpha = 1,2,3\,,$$
$$\hat{T}_{66} \approx \Omega\hat{\rho}c^2a^2, \quad \Omega = \text{const}\,, \quad \Rightarrow \quad \hat{T} \approx \hat{\rho}c^2 - 3\Omega\hat{\rho}c^2 = \hat{\rho}c^2\left(1-3\Omega\right)\,,$$

where $\hat{\rho}$ is the rest mass density of the perturbation.

In the general case $\hat{T}_{DD} \approx \Omega\hat{\rho}c^2a^2$ and $\hat{T} \approx \hat{\rho}c^2 - n\Omega\hat{\rho}c^2 = \hat{\rho}c^2\left(1-n\Omega\right)$.

The approximate expressions for the covariant components of the total energy-momentum tensor read

$$T_{00} \approx \hat{\rho}c^2+\bar{\varepsilon}+\varepsilon^1+\bar{\varepsilon}A^1, \quad T_{11} \approx -\bar{\varepsilon}-\varepsilon^1+\bar{\varepsilon}B^1\,, \quad T_{22} \approx -\bar{\varepsilon}-\varepsilon^1+\bar{\varepsilon}C^1,$$
$$T_{33} \approx -\bar{\varepsilon}-\varepsilon^1+\bar{\varepsilon}D^1\,, \quad T_{44} = T_{66}\sin^2\eta\sin^2\zeta, \quad T_{55} = T_{66}\sin^2\zeta\,,$$
$$T_{66} \approx \Omega\hat{\rho}c^2a^2+\omega_1 a^2\bar{\varepsilon}+\omega_1 a^2\varepsilon^1-\omega_1\bar{\varepsilon}G^1$$
$$\Rightarrow \quad T \approx \hat{\rho}c^2\left(1-3\Omega\right)+4\bar{\varepsilon}-3\omega_1\bar{\varepsilon}+4\varepsilon^1-3\omega_1\varepsilon^1\,.$$

Obviously, in the general case $T_{DD} \approx \Omega\hat{\rho}c^2a^2+\omega_1 a^2\bar{\varepsilon}+\omega_1 a^2\varepsilon^1-\omega_1\bar{\varepsilon}G^1$ and $T \approx \hat{\rho}c^2\left(1-n\Omega\right)+4\bar{\varepsilon}-n\omega_1\bar{\varepsilon}+4\varepsilon^1-n\omega_1\varepsilon^1$.

Taking into account the relationships $\Lambda_7 = \bar{\varepsilon}\left(1+3\omega_1\right)/2$ and $\bar{\varepsilon} = 2/[\kappa a^2(1+\omega_1)]$ (or $\Lambda_{4+n} = \bar{\varepsilon}\left(n-2+n\omega_1\right)/2$ and $\bar{\varepsilon} = (n-1)/[\kappa a^2(1+\omega_1)]$ respectively in the general case) as well as the relationships $B^1 = C^1 = D^1$ and $-A^1+B^1+3G^1/a^2 = 0$ (or $-A^1+B^1+nG^1/a^2 = 0$), which follow directly from the non-diagonal components of the Einstein equation, one can equate both sides of the Einstein equation diagonal components and get

$$\triangle A^1 = \frac{2}{5}\kappa\left(4+3\Omega\right)\hat{\rho}c^2+\frac{2}{5}\kappa(1+3\omega_1)\varepsilon^1\,,$$
$$\triangle B^1 = \frac{2}{5}\kappa\left(1-3\Omega\right)\hat{\rho}c^2-\frac{2}{5}\kappa(1+3\omega_1)\varepsilon^1\,,$$
$$\triangle G^1 = \frac{2}{5}\kappa a^2\left(1+2\Omega\right)\hat{\rho}c^2+\frac{4}{5}\kappa a^2(2+\omega_1)\varepsilon^1-\frac{4}{a^2}G^1\,.$$

It can be easily seen that the relationship between G^1 and ε^1 reads

$$G^1 = \frac{\kappa a^4}{3}\varepsilon^1, \quad \varepsilon^1 = \frac{3}{\kappa a^4}G^1 \, .$$

The final expressions for A^1, B^1 and G^1 in the case $1+3\omega_1 > 0$ and $\hat{\rho} = m\delta(\mathbf{r}_3)/S_3$ (where m is the mass of the particle at rest and S_3 is the total volume of the internal space) read

$$\begin{aligned} A^1 &= \frac{2\varphi_N}{c^2} + \frac{3}{2a^2}G^1, \quad B^1 = \frac{2\varphi_N}{c^2} - \frac{3}{2a^2}G^1 \, , \\ G^1 &= a^2\frac{4\varphi_N}{5c^2}(1+2\Omega)\exp\left(-\frac{r_3}{\lambda_3}\right) \, . \end{aligned}$$

Here $\varphi_N = -G_N m/r_3$ is the standard Newtonian gravitational potential and $\lambda_3 = \sqrt{5}a/\left(2\sqrt{1+3\omega_1}\right)$.

In general case one can equate both sides of the Einstein equation diagonal components and get

$$\begin{aligned} \triangle A^1 &= \frac{2(1+n+n\Omega)}{2+n}\kappa\hat{\rho}c^2 + \frac{2(n-2+n\omega_1)}{2+n}\kappa\varepsilon^1 \, , \\ \triangle B^1 &= \frac{2(1-n\Omega)}{2+n}\kappa\hat{\rho}c^2 - \frac{2(n-2+n\omega_1)}{2+n}\kappa\varepsilon^1 \, , \\ \triangle G^1 &= \frac{2(1+2\Omega)}{2+n}\kappa a^2\hat{\rho}c^2 + \frac{4(2+\omega_1)}{2+n}\kappa a^2\varepsilon^1 - \frac{2(n-1)}{a^2}G^1 \, . \end{aligned}$$

Again it can be easily seen that the relationship between G^1 and ε^1 reads

$$G^1 = \frac{2\kappa a^4}{n(n-1)}\varepsilon^1, \quad \varepsilon^1 = \frac{n(n-1)}{2\kappa a^4}G^1 \, .$$

The final expressions for A^1, B^1 and G^1 in the case $n-2+n\omega_1 > 0$ and $\hat{\rho} = m\delta(\mathbf{r}_3)/S_n$ (where S_n is the total volume of the internal space) read

$$\begin{aligned} A^1 &= \frac{2\varphi_N}{c^2} + \frac{n}{2a^2}G^1, \quad B^1 = \frac{2\varphi_N}{c^2} - \frac{n}{2a^2}G^1 \, , \\ G^1 &= a^2\frac{4\varphi_N}{(2+n)c^2}(1+2\Omega)\exp\left(-\frac{r_3}{\lambda_n}\right) \, . \end{aligned}$$

Here $\lambda_n = a\sqrt{(2+n)/\left[2(n-1)(n-2+n\omega_1)\right]}$.

Obviously, if the range λ_3 (or λ_n) of Yukawa interaction is small enough, we can drop the corresponding terms in the astrophysical problems and obtain the relationship $A^1 = B^1$ in agreement with the deflection of light and the time delay of radar echoes with the same accuracy as General Relativity.

4 Conclusions

In order to calculate the deflection of light by the Sun and the time delay of radar echoes, we need the metric coefficients in the weak field approximation. Performing the corresponding calculations in General Relativity, we usually assume that the background spacetime metrics is flat and perturbation has the form of a point-like mass (see, e.g., [7]).

In the present paper we considered the Kaluza-Klein model where the internal space is not flat but has the form of a three- (or n-) sphere with the radius a. Similar to General Relativity, the external spacetime background remains flat and the perturbation takes the form of a point-like mass. Additionally, we included a bare multidimensional cosmological constant. First, we found the background matter which corresponds to our unperturbed metrics. It was shown that this matter can be simulated by a perfect fluid with the vacuum equation of state in the external space and an arbitrary equation of state with the parameter ω_1 in the internal space. Then, in the weak field approximation we perturbed the background matter and metrics by a point-like mass. We assumed that such perturbation does not change the equations of state. We have shown that in the case $1+3\omega_1 > 0$ (or $n-2+n\omega_1 > 0$) the perturbed metric coefficients have the Yukawa type corrections with respect to the standard Newtonian gravitational potential. The inverse square law experiments restrict such corrections and provide the following bound on the parameters of the model: $\lambda_{\max} \sim 10^{-3}$ cm. Obviously, in the Solar system we can drop the Yukawa correction terms with very high accuracy, and the post-Newtonian parameter γ is equal to 1 similar to General Relativity. Therefore, our model satisfies the gravitational experiments (the deflection of light and the time delay of radar echoes) at the same level of accuracy as General Relativity. This is the main conclusion of our paper. The usual drawback of such models consists in fine tuning of their parameters.

References

1. C.M. Will, *Theory and Experiment in Gravitational Physics*, (Cambridge University Press, Cambridge, 2000).

2. N. Straumann, *General Relativity and Relativistic Astrophysics*, (Springer-Verlag, Berlin, Heidelberg, 1984).

3. B. Bertotti, L. Iess and P. Tortora, Nature **425**, 374 (2003).

4. C.M. Will *Was Einstein Right? Testing Relativity at the Century*, in *100 Years of Relativity: Space-time Structure – Einstein and Beyond, ed. Abhay Ashtekar, page 205*, (World Scientific, Singapore, 2005); (arXiv:gr-qc/0504086).

5. Bh. Jain and J. Khoury, *Cosmological Tests of Gravity* (2010); (arXiv:astro-ph/1004.3294).

6. T. Clifton, P.G. Ferreira, A. Padilla and C. Skordis, Physics Reports **513**, 1 (2012); (arXiv:astro-ph/1106.2476).

7. L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields, Fourth Edition: Volume 2 (Course of Theoretical Physics Series)*, (Pergamon Press, Oxford, 2000).

8. A. Chopovsky, M. Eingorn and A. Zhuk, Phys. Rev. D **85**, 064028 (2012); (arXiv:gr-qc/1107.3388).

9. A. Chopovsky, M. Eingorn and A. Zhuk, *Exact and asymptotic black branes with spherical compactification*, submitted to Phys. Rev. D (2012); (arXiv:gr-qc/1202.2677).

10. D.J. Kapner, T.S. Cook, E.G. Adelberger, J.H. Gundlach, B.R. Heckel, C.D. Hoyle and H.E. Swanson, Phys. Rev. Lett. **98**, 021101 (2007); (arXiv:hep-ph/0611184).

A. Kusevich[15], *M. Eingorn, A. Zhuk*

On Gibbs distribution and equations of state[16]

Odessa National University named after I.I. Mechnikov,
2, Dvoryanskaya street, Odessa 65082, Ukraine
e-mail: lika.kusevich@gmail.com

The black string represents the most interesting exact spherically symmetric "soliton" solution of the vacuum Einstein equation in five-dimensional space-time with a single finite (compact) extra dimension. Its metrics describes the gravitational field of a non-dust-like matter source, possessing relativistic negative tension and, hence, a very specific equation of state in the one-dimensional internal space. Such matter source has two main advantages. First of all, it is at the same level of agreement with gravitational tests as General Relativity. Secondly, its equations of state do not violate the necessary condition of the internal space stabilization. At the same time tension has no clear physical origin.

We investigate this challenge from the statistical physics and thermo-

[15]This research was co-financed by the Austrian Academy of Sciences in the framework of the collaboration with the National Academy of Sciences of Ukraine on *Modern Problems in Astroparticle Physics*.

[16]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

dynamics point of view, generalizing the standard Gibbs distribution and its consequences to the multidimensional case. Using quantum mechanics, we find the discrete part of the energy spectrum of a free black string. Then we consider an ideal gas of black strings and obtain corresponding non-relativistic equations of state.

1 Introduction

Present-day observable phenomena, such as dark energy and dark matter, represent the great challenge for modern cosmology, astrophysics and theoretical physics generally. Nowadays within the scope of standard models these phenomena have no satisfactory explanation. This critical situation stimulates the search of solutions of this very complicated and overwhelmingly important problem beyond all conventional models, for example, by introducing extra spatial dimensions (ESDs). This breathtaking generalization follows directly from modern theories of unification of all known fundamental interactions (such as superstring theory, supergravity and M-theory). Indeed, these theories have the most self-consistent formulation in multidimensional space-times with ESDs [1]. Obviously, it is extremely necessary to subject these and other non-standard physical theories to a procedure of hard-edged screening concerning their compatibility with experimental data.

In the well-known Kaluza-Klein models, based on two pioneering papers [2, 3] by Theodor Kaluza and Oskar Klein respectively, all ESDs are assumed to be finite/compact and microscopic (see, e.g., [4-6], where the authors involve such ESDs in solving of the well-known topical hierarchy problem). Let us note that in brane world models (see, e.g., [7, 8]) ESDs may be macroscopic and even infinite/non-compact.

In the recent paper [9] it was explicitly shown that Kaluza-Klein models with toroidal compactification of ESDs and a standard dust-like matter source of the gravitational field contradict experimental data of astronomical observations. In these models formulas for the classical gravitational tests of any theory of gravity (such as the perihelion shift, the deflection of light, the time delay of radar echoes

[10] and PPN parameters [11, 12]) are incompatible with observations in the Solar System.

Let us note that in the important case of non-toroidal (namely, spherical) compactification of ESDs the state of affairs improves noticeably because of the background with a non-dust-like equation of state in the internal space. When appropriately choosing model parameters, this background leads to stabilization of the internal space, and at first glance the corresponding Kaluza-Klein model becomes consistent with all known experimental data, including astronomical observations as well as laboratory tests of the Newton's inverse square law at small distances. One of the main characteristic features of spherical compactification lies in the fact that the internal space is curved and the background with the non-zero energy-momentum tensor as well as certain physical properties is necessary to provide this curvature. The natural topical question arises, whether Kaluza-Klein models with toroidal compactification also survive, when introducing non-dust-like matter sources of the gravitational field with non-dust-like equations of state in the internal space.

Such matter sources were considered in [13], where it was explicitly shown that among the exact "soliton" solutions of the vacuum Einstein equation in the 5-dimensional space-time with a single compact ESD [14-16], describing the static gravitational field of a finite spherically symmetric matter source at rest, there is only one solution, called "the black string", satisfying all observational data with the same accuracy as the Schwarzschild solution in General Relativity. This fact represents the main advantage of this solution. According to the considered Kaluza-Klein model, all ordinary non-relativistic particles must be identified exactly with the black strings.

A single black string at rest is characterized by the dust-like equation of state $p_0 = 0$ in the 3-dimensional external space and the very specific, strange and even unlikely equation of state $p_1 = -\varepsilon/2$ in the 1-dimensional internal space, where p_0 and p_1 are the corresponding pressures and ε is the rest energy density. Thus, the pressure p_1, sometimes called "tension", is negative and relativistic. Unfortunately, both these circumstances have unclear physical origin, and the corresponding burning issue remains open. This fact represents the main disadvantage of the black string.

In this work we produce consistent multidimensional generalization of standard methods of quantum mechanics, statistical physics and thermodynamics and apply it in order to derive different thermodynamic quantities, characterizing an ideal gas of black strings. Firstly, we solve exactly the 4-dimensional Schrodinger equation for the wave function of a free black string and find its energy spectrum. Secondly, we generalize the standard Gibbs distribution to the case of the multidimensional space and obtain the partition function of the considered ideal gas. Thirdly, with the help of this function and the first law of thermodynamics we arrive at the explicit expression for the pressure in the internal space and investigate its asymptotical behavior. This predictably positive and non-relativistic expression represents the usual temperature dependent contribution to the pressure.

In order to explain the unusual (negative and relativistic as well as temperature independent) contribution, we assume that each non-relativistic particle perturbs the hypothetical background matter in such a way that together with this nontrivial perturbation it looks like a black string with the tension. This strong requirement imposes severe restrictions on the parameters of the perturbation.

In conclusion we summarize our main results.

2 Multidimensional Gibbs distribution and an ideal gas

Let us start with the stationary 4-dimensional Schrodinger equation

$$\hat{H}_4\psi_4 = E_4\psi_4, \quad \hat{H}_4 = \hat{H}_3 - \frac{\hbar^2}{2m}\frac{\partial^2}{\partial\xi^2},$$

$$\hat{H}_3 = -\frac{\hbar^2}{2m}\triangle_3 = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right), \tag{1}$$

where $\hat{H}_4$ and $\hat{H}_3$ are 4- and 3-dimensional Hamilton operators respectively; ψ_4 is a wave function of a free black string (it depends on all spatial coordinates x, y, z, ξ, but does not depend on time t); the coordinate ξ corresponds to the ESD and $\triangle_3$ is a 3-dimensional Laplace operator. Let us note that subscripts 4, 3 and 1 indicate everywhere that the corresponding quantity relates to the total 4-dimensional, the external 3-dimensional or the internal 1-dimensional spaces

respectively.
Following the variable separation method, we seek for the solution of the equation (1) in the form $\psi_4(x,y,z,\xi)=\psi_3(x,y,z)\psi_1(\xi)$ and obtain

$$\hat{H}_3\psi_3=E_3\psi_3, \quad -\frac{\hbar^2}{2m}\frac{d^2\psi_1}{d\xi^2}=E_1\psi_1, \quad E_4=E_3+E_1, \tag{2}$$

where E_3 and E_1 represent the standard and the additional parts of the total energy E_4 respectively. Now our aim is to determine E_1 . Imposing periodic boundary conditions

$$\psi_1(0)=\psi_1(a), \quad \frac{d\psi_1}{d\xi}(0)=\frac{d\psi_1}{d\xi}(a), \tag{3}$$

where a is the period of the torus (the size of the ESD), one can explicitly show that

$$E_{1(n)}=\frac{2\pi^2\hbar^2}{ma^2}n^2, \quad n=0,1,2,.... \tag{4}$$

Thus, we have arrived at the additional energy spectrum, which is necessary for the subsequent determination of the corresponding partition function Z_1. For $n=0$ the wave function $\psi_{1(0)}=1/\sqrt{a}$ is constant. Therefore, we can draw an important side conclusion that in the ground state ($n=0$, $E_{1(0)}=0$) the black string is uniformly smeared over the ESD. Thus, the assumption of the uniform smearing, actually made in [14-16], means that the matter source is considered in its ground state.
For $n=1,2,3,...$ the wave function $\psi_{1(n)}$ can be expressed in the form of the linear combination of two orthogonal functions

$$\psi_{1s(n)}=\sqrt{\frac{2}{a}}\sin\left(\frac{2\pi n}{a}\xi\right), \quad \psi_{1c(n)}=\sqrt{\frac{2}{a}}\cos\left(\frac{2\pi n}{a}\xi\right). \tag{5}$$

Both these functions (as well as $\psi_{1(0)}$) are real and satisfy the normalization condition $\int\limits_0^a \psi_1^2 d\xi=1$.
Now let us turn to the multidimensional Gibbs distribution. Proceeding from the fundamental principles of quantum statistical physics, one can show that it preserves its standard form:

$$w_\nu=\frac{1}{Z}\exp\left(-\frac{\varepsilon_\nu}{kT}\right), \quad \sum_\nu w_\nu=1, \quad Z=\sum_\nu \exp\left(-\frac{\varepsilon_\nu}{kT}\right), \tag{6}$$

where w_ν represents the probability of finding a system, closed in the thermostat, in the ν-th quantum state with the energy ε_ν; ν denotes the full set of quantum numbers, unambiguously determining the considered quantum state; k is the Boltzmann constant and T is the temperature. Finally, Z represents the partition function.

Now let us consider an ideal gas of N identical black strings. Obviously, in view of (2) the partition function Z_4 of each of them can be expressed in the form of the product of two partition functions Z_3 and Z_1, corresponding to the external and the internal spaces respectively: $Z_4 = Z_3 Z_1$. Substituting the discrete spectrum (4) into (6), we obtain

$$Z_1 = \sum_{n=0}^{+\infty} \exp\left(-\frac{E_{1(n)}}{kT}\right) = \sum_{n=0}^{+\infty} \exp\left(-\frac{2\pi^2\hbar^2}{ma^2kT}n^2\right) =$$
$$= \sum_{n=0}^{+\infty} \exp\left(-\frac{T_c}{T}n^2\right) = \sum_{n=0}^{+\infty} q^{n^2} = \frac{1}{2} + \frac{1}{2}\theta_3(0,q), \tag{7}$$

where $\theta_3(z,q) = 1 + 2\sum_{n=1}^{+\infty} q^{n^2}\cos 2nz$ denotes the third of the theta-functions [17-19].

In (7) we have also introduced a convenient quantity q and a characteristic temperature T_c:

$$q = \exp\left(-\frac{2\pi^2\hbar^2}{ma^2kT}\right) = \exp\left(-\frac{T_c}{T}\right), \quad 0 < q < 1, \quad T_c = \frac{2\pi^2\hbar^2}{ma^2k}. \tag{8}$$

According to [20], the free energy $F = U - TS = -kTlnZ$, where U is the internal energy and S is the entropy, preserves its standard form, while the first law of thermodynamics now reads

$$TdS = dU + p_0 a dV_3 + p_1 V_3 da, \quad dF = -SdT - p_0 a dV_3 - p_1 V_3 da. \tag{9}$$

It follows from (9), in particular, that

$$p_0 = -\frac{1}{a}\left(\frac{\partial F}{\partial V_3}\right)_{T,a}, \quad p_1 = -\frac{1}{V_3}\left(\frac{\partial F}{\partial a}\right)_{T,V_3},$$
$$S = -\left(\frac{\partial F}{\partial T}\right)_{V_3,a}, \quad U = -T^2\left(\frac{\partial}{\partial T}\left(\frac{F}{T}\right)\right)_{V_3,a}. \tag{10}$$

For the considered ideal gas the existence of the ESD results in the additional (everywhere with respect to the standard 3-dimensional part) free energy

$$F_1 = -NkTlnZ_1 = -NkTln\left[\sum_{n=0}^{+\infty}\exp\left(-\frac{T_c}{T}n^2\right)\right] =$$

$$= -NkT\left[\frac{1}{2}+\frac{1}{2}\theta_3\left(0,\exp\left(-\frac{T_c}{T}\right)\right)\right]. \qquad (11)$$

From (10) and (11) we obtain the following additional pressures:

$$p_0 = 0, \quad p_1 = \frac{2NkT_c}{V_3 a}\frac{\sum\limits_{n=0}^{+\infty} n^2 q^{n^2}}{\sum\limits_{n=0}^{+\infty} q^{n^2}} = \frac{2NkT}{V_3 a}\frac{\theta_3'(0,q)}{1+\theta_3(0,q)}\exp\left(-\frac{T_c}{T}\right), \qquad (12)$$

where the prime denotes the derivative with respect to q. It is clear that p_1 is positive and non-relativistic. It has the following asymptotes:

$$p_1|_{T<<T_c} \approx \frac{2NkT_c}{V_3 a}\exp\left(-\frac{T_c}{T}\right), \quad p_1|_{T>>T_c} \approx \frac{NkT}{V_3 a} = n_4 kT, \quad n_4 = \frac{N}{V_3 a}. \quad (13)$$

The latter asymptote is predictable, since when the temperature is high enough, we can apply the classical approach instead of the quantum one.

3 Background matter perturbation and tension

In order to explain the tension of a single black string, let us consider the 5-dimensional Minkowski metrics, slightly perturbed by the ordinary non-relativistic particle of the mass m at rest, uniformly smeared over the ESD:

$$dS^2 \approx (1+A_1)c^2dt^2 + (-1+B_1)\left(dx^2+dy^2+dz^2\right) + (-1+C_1)d\xi^2, \qquad (14)$$

where small correction functions A_1, B_1 and C_1 depend only on $r_3 = \sqrt{x^2+y^2+z^2}$ in view of spherical symmetry and satisfy the following gauge and boundary conditions:

$$A_1 = B_1 + C_1; \quad \lim_{r_3\to+\infty} A_1 = 0, \quad \lim_{r_3\to+\infty} B_1 = 0, \quad \lim_{r_3\to+\infty} C_1 = 0. \qquad (15)$$

The choice of the perturbed metrics exactly in the form (14) with such metric coefficients g_{ik} is always possible in the considered case (see, e.g., [21], where a similar approach is evolved). Henceforth we adhere to the same accuracy everywhere. The non-zero covariant Ricci tensor components read

$$R_{00} \approx \frac{1}{2}\triangle_3 A_1, \quad R_{11} = R_{22} = R_{33} \approx \frac{1}{2}\triangle_3 B_1, \quad R_{44} \approx \frac{1}{2}\triangle_3 C_1. \tag{16}$$

Let us assume that the particle itself has no tension, and, consequently, its only non-zero covariant energy-momentum tensor component reads $\hat{T_{00}} \approx \rho_4 c^2$, where the 4-dimensional rest mass density reads $\rho_4 = m\delta(\mathbf{r}_3)/a$. However, it is not unlikely that the presence of the mass m can cause the background matter perturbation with

$$\widetilde{T}_{00} \approx \widetilde{\varepsilon}, \quad \widetilde{T}_{11} = \widetilde{T}_{22} = \widetilde{T}_{33} \approx \widetilde{p}_0 = \omega_0 \widetilde{\varepsilon}, \quad \widetilde{T}_{44} \approx \widetilde{p}_1 = \omega_1 \widetilde{\varepsilon}, \tag{17}$$

where the function $\widetilde{\varepsilon}$ also depends only on r_3; ω_0 and ω_1 are constants. Thus, this perturbation looks like a perfect fluid with different equations of state in the external and the internal spaces. The total energy-momentum tensor has the following non-zero covariant components and trace:

$$T_{00} = \hat{T}_{00} + \widetilde{T}_{00} \approx \rho_4 c^2 + \widetilde{\varepsilon}, \quad T_{11} = T_{22} = T_{33} \approx \omega_0 \widetilde{\varepsilon}, \quad T_{44} \approx \omega_1 \widetilde{\varepsilon},$$

$$T = T_{ik} g^{ik} \approx \rho_4 c^2 + (1 - 3\omega_0 - \omega_1)\widetilde{\varepsilon}. \tag{18}$$

Substituting (16) and (18) into the 5-dimensional Einstein equation

$$R_{ik} = \kappa \left(T_{ik} - \frac{1}{3} T g_{ik} \right), \quad \kappa = \frac{2 S_4 G_5}{c^4}, \quad S_4 = \frac{2\pi^2}{\Gamma(2)} = 2\pi^2, \tag{19}$$

where G_5 is the gravitational constant in the 5-dimensional space-time, we obtain

$$\frac{1}{2}\triangle_3 A_1 = \kappa \left\{ \rho_4 c^2 + \widetilde{\varepsilon} - \frac{1}{3}\left[\rho_4 c^2 + (1 - 3\omega_0 - \omega_1)\widetilde{\varepsilon}\right] \right\} =$$

$$\frac{2}{3}\kappa \rho_4 c^2 + \kappa \widetilde{\varepsilon} \frac{2 + 3\omega_0 + \omega_1}{3}, \tag{20}$$

$$\frac{1}{2}\triangle_3 B_1 = \kappa \left\{ \omega_0 \widetilde{\varepsilon} + \frac{1}{3}\left[\rho_4 c^2 + (1 - 3\omega_0 - \omega_1)\widetilde{\varepsilon}\right] \right\} = \frac{1}{3}\kappa \rho_4 c^2 + \kappa \widetilde{\varepsilon} \frac{1 - \omega_1}{3}, \tag{21}$$

$$\frac{1}{2}\triangle_3 C_1 = \kappa\left\{\omega_1\widetilde{\varepsilon} + \frac{1}{3}\left[\rho_4 c^2 + (1 - 3\omega_0 - \omega_1)\widetilde{\varepsilon}\right]\right\} =$$

$$= \frac{1}{3}\kappa\rho_4 c^2 + \kappa\widetilde{\varepsilon}\frac{1 - 3\omega_0 + 2\omega_1}{3}. \tag{22}$$

It follows from (15) and (20), (21), (22) that $\omega_0 = 0$ and, consequently,

$$\triangle_3 A_1 = \frac{4}{3}\kappa\rho_4 c^2 + 2\kappa\widetilde{\varepsilon}\frac{2 + \omega_1}{3}, \quad \triangle_3 B_1 = \frac{2}{3}\kappa\rho_4 c^2 + 2\kappa\widetilde{\varepsilon}\frac{1 - \omega_1}{3},$$

$$\triangle_3 C_1 = \frac{2}{3}\kappa\rho_4 c^2 + 2\kappa\widetilde{\varepsilon}\frac{1 + 2\omega_1}{3}. \tag{23}$$

Excluding $\widetilde{\varepsilon}$, one can show that

$$\triangle_3\left(A_1 - \frac{2 + \omega_1}{1 + 2\omega_1}C_1\right) = 2\kappa\rho_4 c^2\frac{\omega_1}{1 + 2\omega_1},$$

$$\triangle_3\left(B_1 - \frac{1 - \omega_1}{1 + 2\omega_1}C_1\right) = 2\kappa\rho_4 c^2\frac{\omega_1}{1 + 2\omega_1}. \tag{24}$$

Obviously, the inequalities $\omega_1 \neq 0$ and $\omega_1 \neq -1/2$ must hold true. From (24) we get

$$A_1 = \frac{2 + \omega_1}{1 + 2\omega_1}C_1 + \frac{2\varphi_N}{c^2}, \quad B_1 = \frac{1 - \omega_1}{1 + 2\omega_1}C_1 + \frac{2\varphi_N}{c^2}, \quad \varphi_N = -\frac{G_N m}{r_3}, \tag{25}$$

where G_N is the Newtonian gravitational constant. The non-relativistic gravitational potential φ_N satisfies the Poisson equation

$$\triangle_3\varphi_N = \kappa\rho_4 c^4\frac{\omega_1}{1 + 2\omega_1} = 4\pi G_N a\rho_4 = 4\pi G_N m\delta(\mathbf{r}_3), \tag{26}$$

where the following relationship between the multidimensional and the Newtonian gravitational constants has been established:

$$\frac{\kappa c^4}{a}\frac{\omega_1}{1 + 2\omega_1} = \frac{2S_4 G_5}{a}\frac{\omega_1}{1 + 2\omega_1} = 4\pi G_N, \quad \frac{\pi G_5}{a}\frac{\omega_1}{1 + 2\omega_1} = G_N. \tag{27}$$

Now let us turn to the case $\widetilde{\varepsilon} = \gamma C_1$, where γ is a constant. Then

$$\triangle_3 C_1 = \frac{2}{3}\kappa\rho_4 c^2 + 2\kappa\gamma\frac{1 + 2\omega_1}{3}C_1,$$

$$C_1 = \frac{2\varphi_N}{c_2}\frac{1+2\omega_1}{3\omega_1}\exp\left[-r_3\sqrt{2\kappa\gamma\frac{1+2\omega_1}{3}}\right], \tag{28}$$

where we assume that the important inequality $\gamma(1+2\omega_1) > 0$ holds true. Substituting (28) into (25), we obtain

$$A_1 = \frac{2\varphi_N}{c^2}\left\{1 + \frac{2+\omega_1}{3\omega_1}\exp\left[-r_3\sqrt{2\kappa\gamma\frac{1+2\omega_1}{3}}\right]\right\},$$

$$B_1 = \frac{2\varphi_N}{c^2}\left\{1 + \frac{1-\omega_1}{3\omega_1}\exp\left[-r_3\sqrt{2\kappa\gamma\frac{1+2\omega_1}{3}}\right]\right\}. \tag{29}$$

Obviously, if the quantity $1/\sqrt{2\kappa\gamma(1+2\omega_1)/3}$ is less than a submillimeter scale, then the second terms in braces can be neglected, and at both laboratory and astrophysical distances the important approximate equality $A_1 \approx B_1 \approx 2\varphi_N/c^2$ holds true. It means that there is no any noticeable deviation from the Newton's inverse square law as well as from predictions of the classical gravitational tests. Finally, taking into account a sharp decrease of the Yukawa potential, when r_3 increases, let us approximately replace it by the delta-function:

$$\frac{1}{r_3}\exp\left(-\frac{r_3}{\lambda}\right) \to \delta(\mathbf{r}_3)\int\frac{1}{r_3'}\exp\left(-\frac{r_3'}{\lambda}\right)dV_3' =$$

$$= 4\pi\delta(\mathbf{r}_3)\int\limits_0^{+\infty} r_3'\exp\left(-\frac{r_3'}{\lambda}\right)dr_3' = 4\pi\lambda^2\delta(\mathbf{r}_3), \tag{30}$$

where λ is a parameter, then

$$\widetilde{\varepsilon} = -\gamma\frac{2G_N m}{c^2}\frac{1+2\omega_1}{3\omega_1}\frac{1}{r_3}\exp\left[-r_3\sqrt{2\kappa\gamma\frac{1+2\omega_1}{3}}\right] \to$$

$$-\gamma\frac{2G_N m}{c^2}\frac{1+2\omega_1}{3\omega_1}\frac{4\pi}{2\kappa\gamma\frac{1+2\omega_1}{3}}\delta(\mathbf{r}_3) = -\frac{1}{1+2\omega_1}\rho_4 c^2, \tag{31}$$

where the relationship (27) has been used. This replacement means that we amass artificially the total energy of the background matter perturbation in the origin of coordinates. Substituting (31) into (18), we get

$$T_{00} \to \rho_4 c^2 - \frac{1}{1+2\omega_1}\rho_4 c^2 = \frac{2\omega_1}{1+2\omega_1}\rho_4 c^2, \quad T_{44} \to -\frac{\omega_1}{1+2\omega_1}\rho_4 c^2. \tag{32}$$

Therefore, in this limit the total pressure $p_1 \approx T_{44}$ of both a single particle and the corresponding background matter perturbation in the internal space and the total energy density $\varepsilon \approx T_{00}$ satisfy the equation of state $p_1 = -\varepsilon/2$ of a single black string. Thus, we have arrived at the possible explanation of the tension of black strings.

Conclusions

Let us enumerate briefly the main results of this work:
1. An ideal gas of ordinary non-relativistic particles has been described by the standard methods, generalized to the multidimensional case. In particular, the explicit expressions (11) and (12) for the additional free energy and pressures respectively have been derived. The pressure p_1 in the internal space is positive and temperature dependent.
2. The relativistic, negative and temperature independent tension of each black string can be explained by the corresponding background matter perturbation with the energy-momentum tensor (17), where

$$\widetilde{\varepsilon} = \gamma C_1, \quad \omega_0 = 0, \quad \omega_1 \neq; -\frac{1}{2}, \quad \gamma(1+2\omega_1) > 0.$$

Both conclusions are overwhelmingly important for further development of multidimensional theories of gravity. The first one prejudices Kaluza-Klein models with toroidal ESDs and non-dust-like matter sources of the gravitational field, while the second one gives them a chance of reprieve.
Our results can be generalized directly to the case of the multidimensional space-time with an arbitrary number of toroidal ESDs.

References

1. J. Polchinski, String Theory, Vol. 2: Superstring Theory and Beyond (Cambridge University Press, Cambridge, 1998).
2. Th. Kaluza, Zum Unitatsproblem der Physik, Sitzungsber. d. Preuss. Akad. d. Wiss., 966 (1921).
3. O. Klein, Quantentheorie and funfdimensionale Relativitatstheorie, Zeitschrift fur Physik 37, 895 (1926).

4. N. Arkani-Hamed, S. Dimopoulos, G. Dvali, Phys. Lett. B 429, 263 (1998); arXiv:hep-ph/9803315.

5. I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. B 436, 257 (1998); arXiv:hep-ph/9804398.

6. N. Arkani-Hamed, S. Dimopoulos, Phys. Rev. D 59, 086004 (1999).

7. V.A. Rubakov, Phys. Usp. 44, 871 (2001); Usp. Fiz. Nauk 171, 913 (2001); arXiv:hep-ph/0104152.

8. A.O. Barvinsky, Usp. Fiz. Nauk 175, 569 (2005).

9. M. Eingorn and A. Zhuk, Class. Quant. Grav. 27, 205014 (2010).

10. C.M. Will, Was Einstein Right? (100 Years of Relativity: Spacetime Structure - Einstein and Beyond, World Scientific, Singapore, 2005).

11. N. Straumann, General Relativity and Relativistic Astrophysics (Springer-Verlag, Berlin, Heidelberg, 1984).

12. C.M. Will, Theory and Experiment in Gravitational Physics (Cambridge University Press, Cambridge, 2000).

13. M. Eingorn and A. Zhuk, Phys. Rev. D 83, 044005 (2011).

14. D. Kramer, Acta Phys. Polon. B 2, 807 (1970).

15. D.J. Gross and M.J. Perry, Nucl. Phys. B 226, 29 (1983).

16. A. Davidson and D. Owen, Phys. Lett. 155, 247 (1985).

17. A.P. Prudnikov, Yu.A. Brychkov and O.I. Marichev, Integrals and Series, Vol. 1: Elementary Functions (Gordon and Breach, N.- Y., 1986).

18. Handbook of Mathematical Functions with formulas, graphs and mathematical tables, ed. by M. Abramowitz and I. Stegun (Washington, 1964).

19. H. Bateman and A. Erdelyi, Higher Transcendental Functions, Vol, 3 (MC Graw-Hill, N.- Y., 1955).

20. Y. Tosa, Phys. Rev. D 33, 2326 (1986).

21. L.D. Landau, E.M. Lifshitz, The Classical Theory of Fields, Vol. 2 (Pergamon Press, Oxford, 2000).

V.V. Kuzmichev, V.E. Kuzmichev

New Analytical Solutions of the Einstein Equations for Cosmological Consequences of the Redistribution of Energy between Matter Components in the very early Universe[17]

Bogolyubov Institute for Theoretical Physicsof the National Academy of Sciences of Ukraine, 14-B, Metrolohichna Str., UA-03680 Kiev, Ukraine
e-mail: vkuzmichev@bitp.kiev.ua

The evolution of matter in the expanding FRW universe during the time interval between the end of inflation and the beginning of the radiation-dominated era is studied. A constraint between the global geometry and total amount of matter in the universe as a whole, which is valid during the phase of an intensive transfer of energy to the matter degrees of freedom, is introduced. The matter is considered

[17]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

as a perfect fluid with two components between which there is energy exchange. The analytical solutions of the Einstein equations are found. The limiting cases of the the Hubble expansion rate and the total energy density, which correspond to matter production, pressure-free and radiation-dominated phases are investigated. The transition to the inflationary phase and a unidirectional evolution of matter in the universe at all phases are discussed.

1 Introduction

The standard cosmological model is based on the hot Big Bang model and the assumption about the inflationary expansion of the very early universe, when the scale factor grows quasi-exponentially, while the Hubble expansion rate remains almost constant (e.g., Refs. [1–4]). During inflation the universe is in the vacuum-like state which is usually associated with a scalar field called the inflaton. After inflation the energy density of the primordial matter (except the inflaton) which filled the universe before this stage becomes negligibly small. In order to explain the presence of conventional matter in the universe after inflation, the decay of the vacuum-like state into 'normal' particles is postulated.

The universe becomes hot as a result of interaction between particles and transits into the radiation-dominated phase. In the process of energy transfer from the inflaton to radiation (called reheating) the equation of state of matter changes. A change from a vacuum-like equation of state to the equation of relativistic matter might be gradual and an intermediate stage between these two known phases may be modeled.

In the present article, the evolution of the equation of state of the matter in the universe during the time interval between the end of inflation and the beginning of the radiation-dominated era is considered. Without rendering concrete mechanisms of decay of vacuum-like state into the conventional matter, we assume that the global geometry and total amount of matter in the universe as a whole satisfy a constraint, which is valid during some time interval, before radiation domination. This constraint is equivalent to the law of the conservation of total energy of the universe which remains equal to zero due to the gravitational mass effect, whereas the energy attributed to the particles of conventional matter increases with expansion of the universe [2,5]. In this case, at all stages of evolution the universe is described by the Einstein equations with addition of appropriate equations of state.

The paper is organized as follows. In Section 2 the basic equations which describe the homogeneous, isotropic and spatially flat universe are given. The equations of state of matter for the different phases of reheating are justified. In Section 3 a two-component perfect fluid model is introduced. The analytical solution of the non-linear equation for the Hubble expansion rate is obtained. The expressions for the deceleration parameter and the total energy density as functions of time are deduced. The limiting cases of the solutions which correspond to pressure-free and relativistic matter are considered. The Whitrow-Randall's relation [6] is rederived. In Section 4 the transition to the inflationary phase is discussed. The mechanical analogy which explains a unidirectional evolution of matter in the universe at the phases under consideration is given.

2 Equation of state parameter

Let us consider the homogeneous, isotropic and spatially flat universe in the early epoch, when its dynamics can be described by the equations of the Friedmann-Robertson-Walker (FRW) cosmology,

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3}\rho, \tag{1}$$

$$\dot{\rho} + 3H(\rho + p) = 0, \tag{2}$$

$$p = w(t)\,\rho, \tag{3}$$

where $R(t)$ is the cosmic scale factor, $\rho(t)$ is the energy density of the matter which has a form of the homogeneous perfect fluid, $p(t)$ is its pressure, $w(t)$ is the equation of state parameter, G is the Newtonian gravitational constant, an overdot denotes d/dt, t is the proper time (units $c = 1$ are used).

Let there exist an interval of time after inflation $\Delta t_{ci} = t_c - t_i$, where t_i denotes the time at which inflation ends and t_c stands for the time at which an intensive transfer of energy to the matter degrees of freedom ends. We assume that during this interval the matter is produced, so that the following condition is fulfilled, at least in good approximation,

$$M - G\frac{M^2}{R} = 0, \tag{4}$$

where $M = \frac{4}{3}\pi R^3\rho$ is total mass-energy of matter (the sum of masses of particles of conventional matter) in the equivalent flat-space volume taken without account of gravitational interaction between particles. The equation (4) can be interpreted as the law of the conservation of zero total mass-energy of the universe during its expansion with matter production [5]. According to Eq. (4), during the time interval Δt_{ci} the following relation $R = GM$ is valid. It means that the energy density

$$\rho = \frac{3}{G}\frac{1}{4\pi R^2} \tag{5}$$

decreases linearly with increasing surface area $4\pi R^2$.

From Eqs. (2) and (5), one can find the equation of state

$$p = -\frac{1}{3}\rho, \quad w(t) = -\frac{1}{3}. \tag{6}$$

After the end of this phase, the mass M remains constant on the time interval $\Delta t_{rc} = t_r - t_c$, where t_r denotes the beginning of the subsequent relativistic matter dominant era. The equation of state on the time interval Δt_{rc} takes the form

$$p = 0, \quad w(t) = 0. \tag{7}$$

In the relativistic matter dominant era, the mass attributed to relativistic matter reduces as the universe expands, $M \sim R^{-1}$, due to the cosmic redshift. For times $t > t_r$, the equations of state has a form

$$p = \frac{1}{3}\rho, \quad w(t) = \frac{1}{3}. \tag{8}$$

We will study the model of evolution of matter in the early universe, where the equation of state parameter $w(t)$ changes with time from $-\frac{1}{3}$, passing through the point $w = 0$, to $\frac{1}{3}$ taking all intermediate values. Substituting a continuous function $w(t)$ of time t,

$$w(t) = \frac{1}{3}\tanh\left(\frac{t - t_0}{\tau}\right), \tag{9}$$

for the equation of state parameter on the time interval $[t_i, t_r[$, and choosing properly a point t_0 on this interval, one can reproduce the required values of w (6)-(8). Since $t_r \gg t_i$ (in standard cosmological model, the value $t_i \sim 10^{-35}$ s is acceptable, whereas the time t_r is often evaluated as $t_r \sim 10^{-30}$ s corresponding to temperatures not exceeding 10^{12} GeV [1]), the good estimation for t_0 may be $t_0 \lesssim t_r$. The value $1/\tau$ determines the mean rate of change of the equation of state parameter $w(t)$. Such a variation of the equation of state parameter can be achieved in a system, where the matter consists of a few components between which occurs the energy transfer for some typical time τ.

3 Two-component fluid

Let us consider a two-component perfect fluid with the energy density and pressure

$$\rho = \rho_q + \rho_d, \qquad p = p_q + p_d. \tag{10}$$

These components satisfy the equations

$$\dot{\rho}_q + 3H(\rho_q + p_q) = Q, \quad \dot{\rho}_d + 3H(\rho_d + p_d) = -Q, \tag{11}$$

which represent the energy conservation law (2) rewritten for components, Q is the interaction term.

The components of the perfect fluid are imitated by scalar fields $\phi_q(t)$ and $\phi_d(t)$ with potentials $V_q(\phi_q)$ and $V_d(\phi_d)$,

$$\rho_\alpha = \frac{1}{2}\dot{\phi}_\alpha^{\,2} + V_\alpha,\ p_\alpha = \frac{1}{2}\dot{\phi}_\alpha^{\,2} - V_\alpha,\ \alpha = \{q, d\}. \tag{12}$$

The models of such a type which include a coupling between the matter components were considered in the literature, in particular, within the context of inflation and reheating and the coincidence problem of dark energy and matter in the accelerating universe (see, e.g., Refs. [7–9] and references therein). The form of the interaction term Q may be derived from different physical arguments or obtained as a solution of some dynamical equation, which describes the required properties of the matter fields ϕ_α.

Let us assume that the field ϕ_d forms the pressure-free matter component (dust),

$$\frac{1}{2}\dot{\phi}_d^{\,2} = V_d, \quad \rho_d = 2V_d, \quad p_d = 0. \tag{13}$$

Concerning the field ϕ_q, we suppose that it is described by the vacuum-type equation of state (as for the inflaton) at times $t \ll t_0$,

$$p_q \simeq -\rho_q. \tag{14}$$

From Eq. (12), it follows that at this stage the kinetic energy of the field ϕ_q can be neglected and the total energy is determined by its potential term,

$$\dot{\phi_q}^2 \simeq 0, \quad \rho_q \simeq V_q. \tag{15}$$

For times $t \simeq t_0$, the equation of state takes the form

$$p_q \simeq 0. \tag{16}$$

It means that

$$\frac{1}{2}\dot{\phi_q}^2 \simeq V_q, \quad \rho_q \simeq 2V_q. \tag{17}$$

Then, for the times $t \gg t_0$, the field ϕ_q describes the matter component with the energy density which is almost equal to its kinetic energy,

$$\rho_q \simeq \frac{1}{2}\dot{\phi_q}^2, \quad V_q \simeq 0. \tag{18}$$

This phase corresponds to the reheating of the pressure-free matter and provides the passage to relativistic matter domination. The field ϕ_q here has a form of the stiff Zel'dovich matter,

$$p_q \simeq \rho_q \quad \text{at } t \gg t_0. \tag{19}$$

The continuous transition from Eq. (14) to (16), and then from (16) to (19) can be achieved if the following condition is imposed on the field ϕ_q

$$\frac{1}{2}\dot{\phi_q}^2\, \mathrm{e}^{-2(t-t_0)/\tau} = V_q, \tag{20}$$

where $\tau < \frac{1}{2}t_0$.

Taking into account Eqs. (10), (12), (13), (15), (20), from Eq. (3) we find

$$w(t) = \frac{\mathrm{e}^{2(t-t_0)/\tau} - 1}{\mathrm{e}^{2(t-t_0)/\tau} + 1 + 2V_d/V_q}. \tag{21}$$

This relation passes into Eq. (3.2), if one introduces the following additional condition on V_d,

$$V_d = \rho_q = V_q \left[\mathrm{e}^{2(t-t_0)/\tau} + 1 \right]. \tag{22}$$

Then from Eqs. (10) and (13), we get

$$\rho = 3\rho_q, \quad p = p_q, \quad w = \frac{p_q}{3\rho_q}. \tag{23}$$

In this case, the interaction term $Q = 2Hp_q$ and the set of equations (11) reduces to one equation

$$\dot{\rho}_q + 3H\left(\rho_q + \frac{1}{3}p_q\right) = 0. \tag{24}$$

From Eqs. (1), (3.2), and (24), it follows the non-linear equation for the Hubble expansion rate,

$$\dot{H} + \frac{1}{2}\left\{3 + \tanh\left(\frac{t-t_0}{\tau}\right)\right\} H^2 = 0. \tag{25}$$

The general solution of this equation is

$$H(t) = \frac{2}{D(t)}, \tag{26}$$

where we denote

$$D(t) = Ct_0 + 3t + \tau \ln\cosh\left(\frac{t-t_0}{\tau}\right), \tag{27}$$

C is a constant of integration.

The deceleration parameter, $q = -1 - \dot{H}/H^2$, is equal to

$$q(t) = \frac{1}{2}\left\{1 + \tanh\left(\frac{t-t_0}{\tau}\right)\right\}. \tag{28}$$

The deceleration parameter changes from the value $q = 0$ for the equation of state (6), through the point $q = \frac{1}{2}$ for Eq. (7), to $q = 1$ for Eq. (8). Thus, in the model under consideration, the expansion of the universe is decelerating on the whole time interval from the end of inflation to the beginning of the radiation-dominated era.

The total energy density is

$$\rho(t) = \frac{3}{2\pi G D(t)^2}. \tag{29}$$

The limiting cases of the solutions (26) and (29) reproduce the well-known expressions for the Hubble expansion rate and the energy density. Setting $C = 0$, near the point $t = t_0$ we find for pressure-free matter [10]

$$H(t) \simeq \frac{2}{3t}, \quad \rho(t) \simeq \frac{1}{6\pi G t^2}. \tag{30}$$

Choosing the constant $C \simeq 1 + \frac{\tau}{t_0} \ln 2$, for $t \gg t_0 > 2\tau$ we obtain the relations for the relativistic matter

$$H(t) \simeq \frac{1}{2t}, \quad \rho(t) \simeq \frac{3}{32\pi G t^2}. \tag{31}$$

For times $t \ll t_0$ and $t_0 > 2\tau$, from Eq. (27) it follows

$$D(t) = 2t + (C+1)t_0 - \tau \ln 2. \tag{32}$$

Setting $C \simeq -1 + \frac{\tau}{t_0} \ln 2$, these expressions reduce to

$$H(t) \simeq \frac{1}{t}, \quad \rho(t) \simeq \frac{3}{8\pi G t^2}. \tag{33}$$

The equation for $\rho(t)$ has a form of Whitrow-Randall's relation [6].

4 Discussions

The equations (26), (27), and (29) demonstrate how the Hubble expansion rate and the energy density change with time from the inflationary phase of the universe's expansion, through the subsequent eras of an intensive energy transfer and pressure-free matter, to the beginning of the radiation domination. By choosing the constant of integration C, the solutions (26) and (29) are reduced to known 'standard' expressions (30), (31), and (33). An interesting feature of the solution (29) is that at the point $t = 0$ it is finite,

$$\rho(0) = \frac{3}{2\pi G[(C+1)t_0 - \tau \ln 2]^2}. \tag{34}$$

Thus, the two-component system does not have an initial cosmological singularity.

The equations (26), (29), and (6) can be continued into the region of extremely small values of time, $t \ll \frac{1}{2}|(C+1)t_0 - \tau \ln 2|$, where the Hubble expansion rate slightly changes with time, so that in the inflationary phase $H(t_i) \sim H(0) = \sqrt{\frac{8\pi G}{3}\rho(0)}$ and the expansion of the universe will be exponential in time, $R(t) \sim \exp\{H(t_i)t\}$.

The expression for the energy density $\rho(t_i)$ in the inflationary phase can be reduced to the 'standard' form. Setting $G = M_P^{-2}$ [3], where M_P is the Planck mass, $\tau \simeq M_P^{-1}$, and choosing $C = -1$, from Eq. (34) with a good accuracy we get $H(t_i) \simeq \sqrt{\frac{8\pi}{3}} M_P$, $\rho(t_i) \simeq M_P^{-4}$ (cf. [1]).

In the two-component model (10) with the equation of state (3) with the parameter (3.2), the evolution of the universe goes in one direction, from small times $t \ll t_0$ to large values $t \gg t_0$.

The following mechanical analogy allows to understand the reason of the origin of this 'arrow of time'. The function (3.2) can be considered as the kink solution of the equation

$$\frac{1}{2}\dot{w}^2 + [-U(w)] = 0, \quad U = \frac{9}{2\tau^2}\left(w^2 - \frac{1}{9}\right)^2, \tag{35}$$

which describes the motion of the analogue particle with zero energy in the potential $[-U(w)]$ (cf., e.g., Ref. [11]). This potential has two maxima at the points $w = \pm\frac{1}{3}$ and a local minimum at $w = 0$. The analogue particle moves along the 'trajectory' (3.2) from the value $w = -\frac{1}{3}$ in the distant past ($t = -\infty$) to the value $w = \frac{1}{3}$ reached at $t = \infty$. At the moment $t = t_0$, the particle passes through the minimum of the potential at $w = 0$. Leaving the point $w = -\frac{1}{3}$, the analogue particle can only approach the point $w = \frac{1}{3}$ at $t \to \infty$, where its velocity and acceleration vanish. It cannot return back to $w = -\frac{1}{3}$.

We note that Eq. (35) has another solution in the form of the antikink which is equal to the function (3.2) with an inverse sign. This case corresponds to the model in which the relativistic matter at $t = -\infty$ transforms into the pressure-free matter and then into a gas of low-velocity cosmic strings at $t = \infty$. It was studied in Ref. [12], where it was shown that the equation of state of matter can change with the expansion of the universe due to energy transfer between the matter components (scalar fields) allowing to reproduce the evolution of matter in the universe with non-zero cosmological constant. The description on equal footing of the universe over the total time interval from inflation through reheating to subsequent cooling and transition to the pressure-free matter using the kink and antikink solutions of Eq. (35) may indicate about their important role in the representation of the evolution of matter.

References

1. A.D. Linde, *Particle Physics and Inflationary Cosmology* (Harwood, Chur, 1990).

2. A.D. Dolgov, M.V. Sazhin, and Ya.B. Zel'dovich, *Basics of modern cosmology* (Atlantica Seguier Frontieres, Gif-sur-Yvette Cedex, 1990).

3. D.H. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999) [hep-th/9807278].

4. A. Riotto, CERN Yellow Report CERN-2010-01, 315 (2010), e-print arXiv:1010.2642 [hep-ph] (2010).

5. Ya.B. Zel'dovich, Soviet Physics - JETP **16**, 732 (1963).

6. G.J. Whitrow and D.G. Randall, MNRAS **111**, 455 (1951).

7. A.P. Billyard and A.A. Coley, Phys. Rev. **D61**, 083503 (2000) [astro-ph/9908224].

8. L. Amendola, Phys. Rev. **D62**, 043511 (2000) [astro-ph/9908023]; D. Tocchini-Valentini and L. Amendola, Phys. Rev. **D65**, 063508 (2002) [astro-ph/0108143v3].

9. W. Zimdahl, D. Pavón, and L.P. Chimento, Phys. Lett. **B521**, 133 (2001) [astro-ph/0105479v2]; W. Zimdahl and D. Pavón, Gen. Rel. Grav. **35**, 413 (2003) [astro-ph/0210484]; L.P. Chimento, A.S. Jakubi, D. Pavón, and W. Zimdahl, Phys. Rev. **D67**, 083513 (2003) [astro-ph/0303145].

10. L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields. Vol. 2* (Butterworth-Heinemann, Amsterdam, 1975).

11. R. Rajaraman, *Solitons and Instantons* (North Holland, Amsterdam, 1982).

12. V.E. Kuzmichev and V.V. Kuzmichev, *Acta Phys. Pol. B* **43**, 1850 (2012) [arXiv:1204.5313 [gr-qc]].

Alexander Makarenko

Simulation of Discontinuous Solutions in Evolutionary Equations: New Approaches for Reducing non-Physical Oscillations and Smoothing in Numerical Solutions[18]

Institute of Applied System Analysis at National Technical University of Ukraine (KPI),
37 Peremohy Avenue, 03056, Kyiv, Ukraine
e-mail: makalex@i.com.ua

It is a well-known fact that the difference schemes for approximate solutions of evolution equations have usually some errors within the interval of theoretical accuracy of the schemes. The two most known errors are the artificial smoothing of the solution and oscillations in the solutions near the places with high derivatives of the solutions (near the sharp fronts of the solution). A lot of special tools have been

[18]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

proposed to avoid such effects: artificial viscosity, artificial dispersion, anti-diffusion etc. But the problem is still open, especially in design of special difference schemes. In this paper some theoretical considerations for understanding the errors in numerical computations are proposed. It is strictly considered for some cases as extra smoothing of fronts as the origin of artificial oscillations in the solutions. It is confirmed that the smoothing is originated by dissipation in schemes and oscillations by dispersion of schemes. Some new methods of improving numerical solutions of evolution equations are proposed on the base of theoretical considerations. In the case of linear equations proposed tools can increase the order of the accuracy. The artificial viscosity and artificial dispersion for difference schemes of gas dynamics are proposed as the first examples. A new class of tools for improving numerical solutions is proposed - 'Langoliers'. 'Langoliers' are special difference operators which should be applied at each time steps after the running of original difference schemes. The design of 'Langoliers' allows to reduce the dissipative and dispersive errors of schemes. The examples are anti-diffusion, anti-dispersion and specially constructed difference schemes. Different illustrative examples of such tools are considered for gas dynamics equations and for wave equation.

Keywords: Numerical schemes; dispersion; dissipation; non-smooth solutions, anti-dispersion; 'Langoliers'; collapses.

1 Introduction

It is a well-known fact that the difference schemes for approximate solutions of evolution equations have usually some errors within the interval of theoretical accuracy of the schemes [1]- [4]. The two most known errors are the artificial smoothing of the solution and oscillations in the solutions near the places with high derivatives of the solutions (near the sharp fronts of the solution). A lot of special tools have been proposed to avoid such effects: artificial viscosity in schemes [1], artificial dispersion in schemes [3, 5, 6], anti-diffusion [7],ENO (essentially non-oscillation) schemes [8] etc. But the problem is still open, especially in design of special difference schemes. Remark the before we outline only some of known and working approaches. Some of remembered tools essentially improve the numerical solutions but the difficulties in their applications still are large (as in the theory as in the practice, especially in the modelling of 2D and 3D flows of complex media).

Because of increasing complexity of equations which should be used for modelling of evolving media and systems in hydrodynamics, gas dynamics, plasma, reology the problem of design of more accurate difference schemes is very important. For such goal it is necessary to know the peculiarities of numerical schemes behavior, the souses of 'artifacts' in the numerical solutions and better theoretical understanding of the difference schemes as the objects. So in given paper some theoretical considerations for understanding the errors in numerical computation are proposed. It is strictly considered for some cases as the extra smoothing of fronts as the origin of artificial oscillations in the solutions. It is confirmed that the smoothing is originated by dissipation in schemes and oscillations by dispersion of schemes.

On the base of theoretical considerations there are proposed some methods for improving numerical solutions of evolution equations. In the case of linear equations proposed tools can increase the order of the accuracy. The artificial viscosity and artificial dispersion for difference schemes of gas dynamics are proposed as the first examples.

A new class of tools for improving numerical solutions is proposed - 'Langoliers'. 'Langoliers' are special difference operators which should be applied at each time steps after the running of original difference schemes. The design of 'Langoliers' allows to reduce the dissipative and dispersive errors of schemes. The examples are anti-diffusion, anti-dispersion and specially constructed difference schemes. Different illustrative examples of such tools are considered for gas dynamics equations and for wave equation.

2 Dissipation and dispersion of finite-difference schemes

The terms 'dissipation' and 'dispersion' of difference schemes have a strict sense in case when the original partial differentia equations are linear and have a constant coefficients. In such case difference harmonics of difference scheme (or alternatively the harmonic of continual analogue of difference scheme) is the adequate tool for investigation of the properties of numerical schemes. Such approach is well known and is proposed in any textbook on numerical methods (for example see [1–3, 9]. The results of analysis are specific to a problem or a process. But the general chart of researches remains the same - that is the analysis of accordions and their dispersion correlation is conducted for the initial condition and for the method of its approximation. Therefore further we will illustrate a chart of methods, and also facilities of improvement on a simplest example - equation of transfer or advection. We will consider a Cauchy problem for advection equation in a region $-\infty < x < \infty$

$$Lu = \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0, a = const \tag{1}$$

with initial conditions

$$u(x, 0) = v(x). \tag{2}$$

We will consider also the general class of obvious numerical schemes for equation (1)

$$\Lambda y = \frac{y_j^{n+1} - y_j^n}{\tau} + \sum_{l=-m_1}^{m_2} a_i y_{j+l}^n, \tag{3}$$

where y denotes the numerical solution calculated on lattice $\omega_{h\tau} = \omega_h \times \omega_\tau$, $\omega_h = \{x_j = jh,\ j = 0, \pm 1, ..\}$, $\omega_\tau = \{t_n = n\tau,\ n = 0, 1, ...,\ N\}$, $a\tau/h = \gamma = const$. A scheme (3) can be rewritten in a kind

$$y^{n+1} = Ry^n, \tag{4}$$

where $(Ry^n)_j = y_j^n - \tau \sum_{l=-m_1}^{m_2} a_i y_{j+l}^n = \sum_{l=-m_1}^{m_2} b_i y_{j+l}^n$.

To conduct the analysis of accordions for chart (3) we will consider special kind schemes solutions (numerical harmonic):

$$y_j^{n+1} = q_k^n \exp(ikx_j), \tag{5}$$

where $k = 2\pi/\lambda$ is a wave number, λ is a harmonic wave-length. A value $q_k = \Re q_k + i\Im q_k$ is named the coefficient of transition of scheme. Putting (5) in (4) we get for q_k

$$q_k = \sum_{l=-m_1}^{m_2} b_l \exp(iklh).$$

A transition coefficient can be also presented as

$$q_k = \rho_k \exp(i\phi_k). \tag{6}$$

In formula (3) $\rho_k = \operatorname{mod} q_k = [(\Re q_k)^2 + (\Im q_k)^2]^{1/2}$ there is the module of coefficient of transition $\phi_k = -\arg q_k = \arctan(-\Im q_k/\Re q_k)$. We will name $\nu_k = q_k/k\tau$ phase speed of k-th harmonic. We will enter the continual analogues of the module to the transition coefficient $\rho(\zeta) = a\phi(\zeta)/\gamma\zeta$ from an argument ζ such that $\rho(\zeta_k) = \rho_k$ and also phase velocity $\nu(\zeta_k) = \nu_k$, where $\zeta_k = kh$.

We will consider numerical schemes charts for which

$$\begin{gathered}\rho(\zeta) = 1 - \omega(\zeta), 0 \leq \omega(\zeta) \leq 2, \zeta \leq 1, \\ \omega(\zeta) = c\zeta^s + O(\zeta^{s+2}), c = const, s = 2p\end{gathered} \tag{7}$$

From Rihtmayer's papers [4] it is known, that such scheme has s-th order of dissipation. By [10] a scheme has m-th order of dispersion if a dispersion function can be written down as

$$v(\zeta) = a[1 + \theta\zeta^m + O(\zeta^{m+2})], \quad \theta = const. \tag{8}$$

We now will draw some result from (Brenner & Thomee, 1970). Their character of scheme (coefficient of transition of scheme) was presented in a way where such presentations are accepted

$$q(\zeta) = \exp[-i\gamma\zeta + \Psi(\zeta)], \tag{9}$$

where

$$\Re\Psi(\zeta) = g\zeta^s[1 + o(1)], \quad \zeta \to 0, \quad \zeta > 0, \tag{10}$$

$$\Psi(\zeta) = \Psi_0\zeta^{r+1}[1 + o(1)], \quad \Psi_0 \neq 0, \quad r \geq 0. \tag{11}$$

Then s is interpreted as an order of dissipation, and r characterizes the order of approximation. For schemes with the orders of dissipation s and approximation r and initial conditions $v \in B_p^{\alpha,\infty}$, where $B_p^{\alpha,\infty}$ are spaces of Besov's functions subject to the condition $0 < \alpha < r + 1$, $\alpha \neq (r+1)(1/2 - p^{-1})$ in (Brenner & Thomee, 1970) the estimation for convergence of the scheme had been proved

$$\| y_j^n(x) - v(x, n\tau) \|_{Lp} \leq Ch^{\beta(\alpha)} \| v \|_{n_p^{\alpha,\infty}} \tag{12}$$

and the order of convergence of scheme is set by formulas

$$\beta(\alpha) = \alpha[1-(1+r)^{-1}] + \min(0,\, [\alpha-(r+1)|1/2-p^{-1}|][1/(r+1)-1/s]). \quad (13)$$

If we set the order of dissipation, it is possible to find the order of convergence. For this purpose we will rewrite the coefficient of transition of scheme as following

$$q_k = [1-\omega(\zeta)]\exp(-v_k k\tau) = \exp[-i\gamma\zeta + \Psi_1(\zeta)], \quad (14)$$

where $\Psi_1(\zeta) = [(-c\zeta^s + o(\zeta^s))] + i[\gamma\theta\zeta^{m+1} + o(\zeta^{m+1}))$. From previous formulas it is possible to get very important correlation between the order of approximation, dissipation and dispersion of scheme:

$$r = \min(s,\, m+1) - 1 \quad (15)$$

Very important conclusion is that the order of approximation can be determined by either the order of dissipation or order of dispersion.

If the initial conditions of problem belong to Sobolev's spaces $W_2^\alpha \subset B_2^{\alpha,\infty}$, α is a whole number, m is order of scheme dispersion, s is order of scheme dissipation, $v \in W_2^\alpha$, subject to the condition $0 \le \alpha \le \min(s,\, m+1)$, then speed of convergence d in space L_2 has the form $d = \beta(\alpha) = \alpha\{1 - [\min(s,\, m+1)]^{-1}\}$.

So the order of approximation may be determine either by the order of dissipation or the order of dispersion. Note that the order of approximation also determines the order of convergence (in dependence on the smoothing of the solutions). It may be found from such analysis that for example for even s the schemes have an even order of approximation and the rate of convergence is determined by dispersion effects. This implies that the large non-physical oscillations which are usually observed in schemes of even order of approximation, when computing non-smooth solutions, are precisely due to dispersion of difference harmonics. Note that in the papers [10, 13, 14] other equations and multidimensional case had been considered. The application of results on the order of dissipation and dispersion allows to understand the 'artifacts' in numerical solutions of evolution equations and to propose new tools to suppress or diminish them.

3 Some existing tools of diminishing ‘artifacts’ in calculations

Here we describe the construction of some more or less known tools for improving the quality of numerical solutions and describe their mechanisms with the help of the concepts from section 2. The diagram below shows schematically one step of running the conditional difference scheme.

$$\begin{gathered}\text{Numerical solution on time level } t = \tau k\\ \downarrow\\ \text{Basic difference scheme}\\ \downarrow\\ \text{Numerical solution on time level } t = \tau(k+1)\end{gathered}$$

3.1 Choosing new scheme with increasing accuracy

The first approach to reduce ‘artifacts’ is to take other scheme with increased accuracy. But usually it is time expensive and difficult in theoretical aspects especially for modelling by nonlinear equation in multidimensional cases. So below we discuss the methods for improving the ‘basic’ original schemes by special tools.

3.2 Artificial viscosity approach

According to this approach special terms should be added into the difference scheme for suppressing artificial oscillations by adding non-physical viscosity ([1, 3, 9] and many other papers).

$$\begin{gathered}\text{Numerical solution on time level } t = \tau k\\ \downarrow\\ \text{Basic difference scheme + Artificial viscosity}\\ \downarrow\\ \text{Numerical solution on time level } t = \tau(k+1)\end{gathered}$$

3.3 Artificial dispersion

Special terms should be added into the difference scheme for suppressing artificial oscillations by adding non-physical dispersion [3, 5, 6]

$$\begin{gathered}\text{Numerical solution on time level } t = \tau k \\ \downarrow \\ \text{Basic difference scheme} + \text{Artificial dispersion} \\ \downarrow \\ \text{Numerical solution on time level } t = \tau(k+1)\end{gathered}$$

3.4 Anti-diffusion

The idea of anti-diffusion has been developed since the works by Boris J. and Book D. [1,7,15]. In anti-diffusion the special filtration operator is applied to numerical solution after the running step of conditional scheme with the goal to reduce the oscillations by applying special rules to the solution. It had been shown that the action of such filter is equivalent to some portion of artificial smoothing viscosity. The anti-diffusion already has a lot of applications especially in gas dynamics. But the difficulties of applications lie in the nonlinear character of filter and theoretical background.

$$\begin{gathered}\text{Numerical solution on time level } t = \tau k \\ \downarrow \\ \text{Basic difference scheme} t = \tau k + \tau/2 \\ \downarrow \\ \text{Anti-diffusion} \\ \downarrow \\ \text{Numerical solution on time level } t = \tau(k+1)\end{gathered}$$

4 New tools for improving numerical schemes

In previous section we had described some new but more or less known tools for reducing the oscillations. Here we briefly describe some other tools for improving solutions which have as the background the concepts from section 2.

4.1 Composite schemes of higher order

The investigations of phase velocity and transition modules show that such functions may have either positive or negative dispersion (that is the harmonic of difference scheme may be faster of slower than the harmonic of original differential equation); also the transition modules measure the level of decreasing (increasing) of harmonic amplitude and may be less or bigger then in the harmonic of original equation. So the sequential application of two schemes with different properties for transition from one time level to the next level is equivalent to application some composite scheme with different properties. For example such combination of two different schemes may increase the order of dispersion and thus follows to the essential reduction of artificial oscillations [16]. The same trick may be used for reduction the artificial smoothing in computations.

$$
\begin{array}{c}
\text{Numerical solution on time level } t = \tau k \\
\downarrow \\
\text{Basic difference scheme I } t = \tau k + \tau/2 \\
\downarrow \\
\text{Basic difference scheme II} \\
\downarrow \\
\text{Numerical solution on time level } t = \tau(k+1)
\end{array}
$$

4.2 Anti-dispersion

The goal of anti-dispersion is to reduce the artificial dispersion of numerical schemes by application of special difference operator which dispersion is opposite to the dispersion of basic difference scheme [17]. It is useful to take as such operator the approximation of simplest differential equations with necessary dispersion. Remark that usually it is enough to take the linear part of Kortewega-de-Vreez equation but with special choice of coefficient which allows to compensate for some dispersive error of the scheme:

$$\begin{array}{c}\text{Numerical solution on time level } t = \tau k \\ \downarrow \\ \text{Basic difference scheme I } t = \tau k + \tau/2 \\ \downarrow \\ \text{Anti-dispersion} \\ \downarrow \\ \text{Numerical solution on time level } t = \tau(k+1)\end{array}$$

4.3 'Langoliers'

It is useful to introduce the special name for the class of tools which should be applied after the application of basic scheme at each time step of calculation. We named it as 'Langoliers' because such tools are applied at each point of space grid of difference scheme at given time level and the action of such 'Langoliers' consist in 'eating' 'artificial' defects of numerical solution in each point of greed. We illustrate such mechanism on the diagram below. After the application of basic scheme the solution has a lot of artificial oscillations (the true solution is step function). The application of 'Langolier' reduces the errors essentially.

Figure 1 corresponds to the case when the 'Langolier' has the 'anti-dispersive' character. The anti-diffusive filter may be considered as the 'Langolier' of 'anti-viscosity' character. Also other cases of 'Langoliers' designing may exist. We can use not a single 'Langolier' between time levels but the sequence of different 'Langoliesr'. For example as it follows from the theory of dispersion and dissipation of schemes we can for linear equations theoretically receive any order of approximation of composite 'basic schemes' + series of specially constructed 'Langoliers'. One of the construction consist in consequence 'Langoliers' of 'anti-dispersive' and 'anti-diffusive' nature (but of course of increasing order of dispersion or dissipation and thus of increasing structure). Note also that the apparatus of continual analogs of different schemes may be useful for such design.

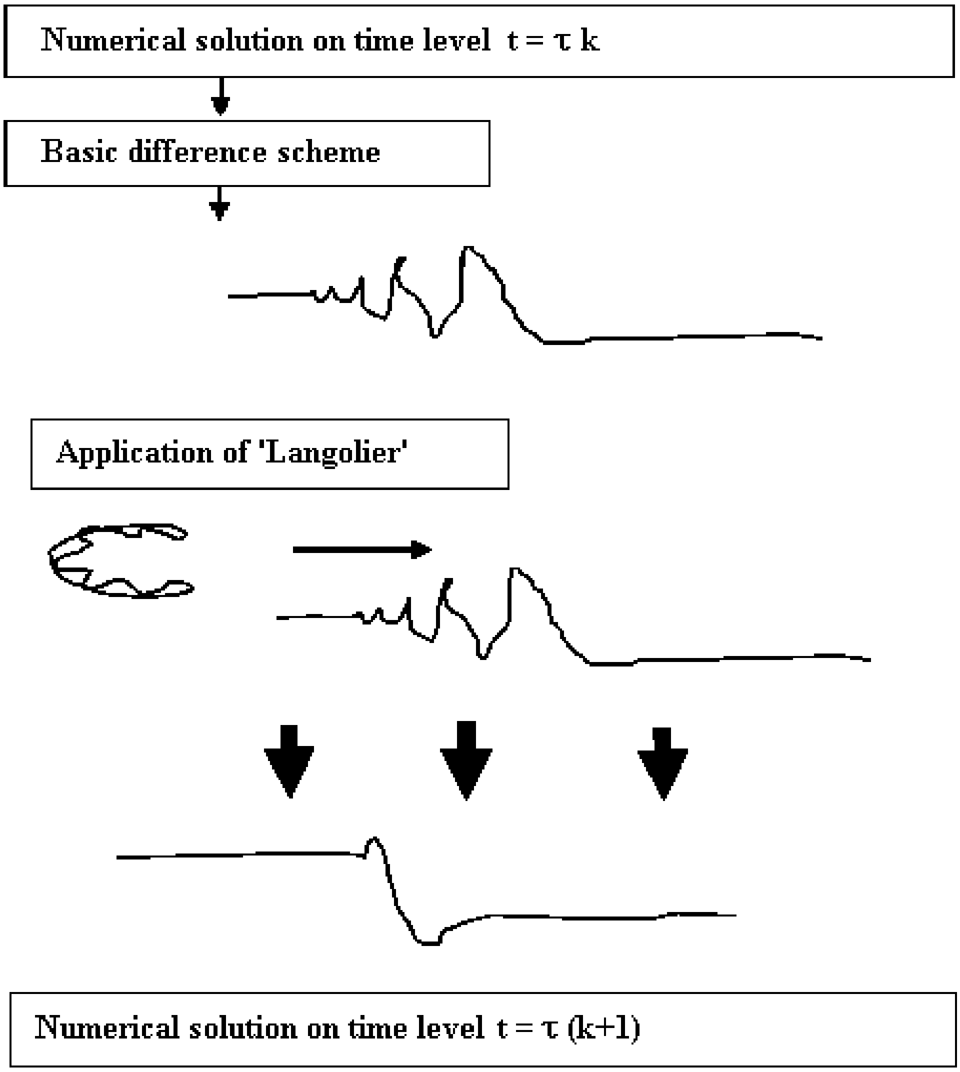


Figure 1: Mechanism of ‘Langolier’ action

5 Nonlinear case

As we already remarked the approaches above already had been developed and tested in case of some equations (the linear transport equation, the wave equation, the Kadomtcev - Petviashvily equation). The conclusion on the applicability of the above tools can be drawn from the numerical solutions of nonlinear equations. The key approach in application consists of two ideas: 1) the linearization of nonlinear equation around the 'basic' solution for original nonlinear equation and 2) the idea of 'frozen' coefficients of received linearized equation [1–3,9]. Then the analysis of harmonic should be proceeding locally. In such case the coefficients of such tools should depend on the values of the solutions at given point in given time moment. The results of such analysis for the case of nonlinear Klein-Gordon equation had been published in [13]. Other interesting example of application of proposed concept to nonlinear equations is described in [16]. The object of investigation is the numerical schemes for some system of gas dynamic equations. It was realized the scheme 4a from the section 4. We take as basic scheme I the Wendroff scheme and as the basic scheme II Lax-Wendroff scheme [4,10,16]. Remark that such schemes have opposite dispersion (positive and negative). The numerical experiments display the essential reduction of artificial oscillations. In fact such composed schemes behave as the scheme of 3d order of accuracy (the Wendroff and Lax-Wendroff schemes has the second order of accuracy).

The proposed approach also is very prospective for numerical calculations of collapses, blow-up solutions or solutions with singularities. Usually such solutions tend to infinite values by the finite time. Such increasing of solutions and their derivatives follows to the reducing of accuracy of approximate methods and to the necessity of adaptive mesh using. Such adaptation follows to decreasing of time and space steps and thus to the non-limiting grows of computational work. In described approach the accuracy of the schemes can be increased with time on the fixed space grid. Also the region of 'Langoliers' application during computation can be concentrated near the singularities points.

6 Conclusions

Thus, in this paper I described special methods and their applications for reduction of artificial errors of type "spreading" and "oscillations" in the calculations of the evolutional equations solutions. It is very important that the offered facilities also befit computations of solutions of one-sided physical processes with memory, because the evolutional equations with memory help to raise proper mathematical problems. In addition, proposed methods become especially perspective in connection with modern development of facilities for the parallel calculations (GRID computation) of solutions. This relates to the fact that that "Langoliers" can be used parallel in every knot of numerical lattice, leading here to increase the exactness in all methods of approximation.

References

1. P.J.Roache, Computational Fluid Dynamics, Moscow, Mir (1980).
2. N.N. Kalitkin, Numerical methods, Moscow, Nauka (1990).
3. A.A. Samarskii, Yu. P. Popov, Difference methods for solving of problems of gas dynamics, Moscow, Nauka (1980).
4. R.D. Richtmyer, K.W. Morton, Difference methods for initial - value problems, New York, Wiley and Sons (1967).
5. S.I. Mukhin, S.B. Popov, Yu. P. Popov, U.S.S.R. Computational Mathematics and Mathematical Physics, Springer Translation, 23 (1983) 45.
6. T. Sengupta, A. Dipankar, P. Sagaut, J. Comp. Phys., 226 (2007) 1211.
7. D.L. Book, J.P. Boris, K. Hain, J. of Comput. Phys. 18 (1975) 248.
8. A. Harten, et all., J. Comput. Physics, 71 (1987) 231.
9. Shokin Yu. I., Yanenko N.N. Differential approximation method. Novosibirsk, Nauka (1985).
10. M.N. Moskalkov, In: Numeric Analysis, Kiev, Int. of Cybern. (1978) 75.

11. Brenner P., Thomee V. Mathematica, Scandinavica, 27 (1970), 5.

12. S.M. Nikolskii, Approximation of functions of several variables and imbedding theorems [in Russian], Moscow, Nauka (1977).

13. A.S. Makarenko, Chisl. Met. Mech. Sploshn. Sred, 13 (1982) 81.

14. A.S. Makarenko, M.N. Moskalkov, Chisl. Met. Mech. Sploshn. Sredy [in Russian], 12 (1981) 64.

15. O. Bokanowski, S. Martin, R. Munos, H. Zidan, Applied Numerical Mathematics, 56 (2006) 1147.

16. A.S. Makarenko, M.N. Moskalkov, Vychisl. Prikl. Matem. [in Russian]. Kiev: Kiev Univ, issue 41 (1980) 55.

17. A. Makarenko, M. Moskalkov, U.S.S.R. Computational Mathematics and Mathematical Physics, Springer Translation, 23 (1983) 999.

S. S. Moskaliuk, N. M. Moskaliuk[19]

Category of Cayley-Klein Groups and Functor Category of VKS-trees[20]

Bogolyubov Institute for Theoretical Physics
Metrolohichna Street, 14-b, Kyiv-143, Ukraine, UA-03143
e-mail: mss@bitp.kiev.ua

Following Manin, Leinster, Markl, Aguiar and Sottile we review definitions, and basic properties of operads, and trees, and algebras over these structures in Sections 1–13. It is intended to consider categories of operads and trees as a whole, the content of each Section taking into account the contents of other ones. But also the same problem of categories of operads and trees may be discussed repeatedly, however from different points of view.

The next Sections we begin by introducing the method of Vilenkin-Kuznetsov-Smorodinsky (VKS)-trees. The Cayley-Klein groups are constructed on parameters, each of which can be real, purely imaginary and Clifford dual one. Then representations of orthogonal Cayley-

---

[19] On leave from Uzhgorod National University, Ukraine.

[20] Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

Klein groups are constructed with the method of VKS-tree. Finally, a Cayley-Klein category is defined and the functor category of VKS-trees is constructed.

1 Operads as a Generalization of Associative Algebras

1.1 Classical linear operads

In Sections 1–2 and 12–13 we fix a ground field k of characteristic zero and denote by $VECT$ the category of linear spaces over k. All tensor products are taken over k unless it is explicitly stated otherwise. The symmetric group $\mathbf{S}_n$ is defined as the group of the bijections $\underline{n} \to \underline{n}$ where $\underline{n} = \{1, \dots, n\}$.

Classical linear algebra deals with a linear space V endowed with a family $\mathcal{O}$ of linear operators $V \to V$. Usually it is convenient to close $\mathcal{O}$ by adding all operator compositions and their linear combinations to $\mathcal{O}$. In this way linear algebra becomes the study of associative k-algebras and their linear representations.

Classical linear operads arise in the same way when we start with a linear space V endowed with a family $\mathcal{P}$ of *poly*linear operators $V^{\otimes m} \to V, m = 1, 2, 3, \dots$ (for example, an associative algebra is such a space endowed with a multiplication map $V^{\otimes 2} \to V$). Closing $\mathcal{P}$ with respect to compositions (of functions with many variables) and linear combinations we get a (concrete) classical linear operad together with its linear representation in V. Axiomatizing the universal properties of such an object, we get the following notion.

DEFINITION 1.1 *A* ***classical linear operad*** $\mathcal{P}$ *consists of the data a) – d) satisfying the axioms A) – C) below.*
a) A family of linear spaces $\mathcal{P}(l)$, *for all* $l \geq 1$.
b) A left/right linear action of $\mathbf{S}_l$ *on* $\mathcal{P}(l)$, *for all* $l \geq 1 : s \in \mathbf{S}_l$ *maps* $f \in \mathcal{P}(l)$ *to* $fs = s^{-1}f$.
c) A family of composition maps $\gamma(k_1, \dots, k_l)$, *for all* $l \geq 1, k_a \geq 1$*:*

$$\gamma(k_1, \dots, k_l) : \ \mathcal{P}(l) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_l) \to \mathcal{P}(k_1 + \cdots + k_l) \,. \tag{1}$$

d) (Optional). An identity element $I \in \mathcal{P}(1)$.

We will state the axioms for these data in two forms: directly in terms of γ and in functional notation. For the latter, put $\underline{\mathcal{P}} = \oplus_{k=1}^{\infty}\mathcal{P}(k)$ and notice that (1) allows us to consider each $f \in \mathcal{P}(l)$ as a polylinear function $\underline{\mathcal{P}}^l \to \underline{\mathcal{P}}$:

$$f(g_1, \dots, g_l) := \gamma(f \otimes g_1 \otimes \cdots \otimes g_l)\ , \tag{2}$$

where $\gamma = \gamma((k_1, \dots, k_l)$ if $g_a \in \mathcal{P}(k_a)$. We will often write simply γ for such multigraded components of the operadic composition.

A) The symmetric group $\mathbf{S}_l$ *acts on the functions (represented by)* $\mathcal{P}(l)$ *by permutation of arguments:*

$$(fs)(g_1, \dots, g_l) = f(s(g_1, \dots, g_l))\ . \tag{3}$$

In γ*-notation:*

$$\gamma(fs \otimes g_1 \otimes \cdots \otimes g_l) = \gamma(f \otimes s(g_1 \otimes \cdots \otimes g_l))\ . \tag{4}$$

In addition, for $s_1 \in S_{k_1}, \dots, s_l \in S_{k_l}$*, denote by* $s_1 \times \cdots \times s_l \in S_{k_1+\dots+k_l}$ *the image of* $(s_1, \dots, s_l)$ *acting blockwise upon*

$$(1, \dots, k_1 | k_1 + 1, \dots, k_1 + k_2 | \dots | k_1 + \cdots + k_{l-1} + 1, \dots, k_1 + \cdots + k_l)\ .$$

Then

$$f(g_1 s_1, \dots, g_l s_l) = (f(g_1, \dots, g_l))(s_1 \times \cdots \times s_l)\ . \tag{5}$$

In γ*-notation:*

$$\gamma(f \otimes g_1 s_1 \otimes \cdots \otimes g_l s_l) = (\gamma(f \otimes g_1 \otimes \cdots \otimes g_l))(s_1 \times \cdots \times s_l)\ . \tag{6}$$

B) The composition maps are associative with respect to the substitution (in functional notation). That is, for any $f \in \mathcal{P}(l)$*,* $g_a \in \mathcal{P}(k_a), a = 1, \dots, l$*, and* $h_{a,b} \in \mathcal{P}(l_{a,b}), b = 1, \dots, k_a$*, we have*

$$\begin{aligned}&[f(g_1, \dots, g_l)](h_{1,1}, \dots, h_{1,k_1}; \dots; h_{l,1}, \dots, h_{l,k_l}) = \\ &\quad = f(g_1(h_{1,1}, \dots, h_{1,k_1}), \dots, g_l(h_{l,1}, \dots, h_{l,k_l}))\ .\end{aligned} \tag{7}$$

In γ*-notation:*

$$\begin{aligned}&\gamma(\gamma(f \otimes g_1 \otimes \cdots \otimes g_l) \otimes h_{1,1} \otimes \cdots \otimes h_{l,k_l}) = \\ &= \gamma(f \otimes \gamma(g_1 \otimes h_{1,1} \otimes \cdots \otimes h_{1,k_1}) \otimes \cdots \otimes \gamma(g_l \otimes h_{l,2} \otimes \cdots \otimes h_{l,k_l}))\ .\end{aligned} \tag{8}$$

C) (Optional). If $\mathcal{P}$ is endowed with identity $I \in \mathcal{P}(1)$, then I (resp. $I^{\otimes n}$) become left (resp. right) identical functions:

$$I(g) = g; \quad f(I, \ldots, I) = f \ , \tag{9}$$

$$\gamma(I \otimes g) = g; \quad \gamma(f \otimes I \otimes \cdots \otimes I) = f \ . \tag{10}$$

An operad endowed with identity which is considered as a part of its structure will be called a ***unital operad****.*

We will often call the classical linear operads simply operads until the introduction of other versions of this notion.

EXAMPLE 1.1 *Let (E, μ) be an associative algebra with multiplication $\mu : E \otimes E \to E$. Define an operad $\mathcal{P}_E$ by $\mathcal{P}_E(1) = E$, $\mathcal{P}_E(l) = \{0\}$ for $l \geq 2$, $\gamma(1) = \mu$, the rest of the data being self-explanatory. Operadic associativity of γ is clearly equivalent to the associativity of μ.*

Conversely, for any operad $\mathcal{P}, \mathcal{P}(1)$ with multiplication $\gamma(1)$ is an associative algebra. Operadic identity becomes algebra identity and vice versa.

EXAMPLE 1.2 *Let V be a linear space. Define the operad $OpEnd(V)$ by the following data:*

$$OpEnd(V)(l) = Hom_{VECT}(V^{\otimes l}, V) \ , \tag{11}$$

$\mathbf{S}_l$ acts by permuting arguments as in (3), the composition γ is defined by substitution as in the left-hand side of (2), and $I = id_V$.

DEFINITION 1.2 *A* ***morphism*** *of operads $\varphi : \mathcal{P} \to \mathcal{Q}$ is a family of linear maps $\varphi(l) : \mathcal{P}(l) \to \mathcal{Q}(l)$, $l \geq 1$, compatible with the action of symmetric groups, composition, and optionally, mapping $I_{\mathcal{P}}$ to $I_{\mathcal{Q}}$.*

Thus we have defined a *category of classical linear operads $OPER$*. In fact, we allow some ambiguity, because the existence of the identity is optional, and, even if it exists, we may decide not to consider it as a part of the structure when we define morphisms. This extends the common ambiguity in the definition of the category of associative algebras.

REMARK 1.1 *Denote by ASS one of the two categories of associative k-algebras (with or without identity). Constructions of Example 1.1 extend to the functors $ASS \to OPER$ and $OPER \to ASS$ which are adjoint to each other from both sides so that we have canonical identifications*

$$Hom_{OPER}(\mathcal{P}, \mathcal{P}_A) = Hom_{ASS}(\mathcal{P}(1), A),$$
$$Hom_{OPER}(\mathcal{P}_A, \mathcal{P}) = Hom_{ASS}(A, \mathcal{P}(1)) \ .$$

In particular, ASS is a full subcategory of $OPER$.

1.2 Operads as classifiers of algebras of different species

By species we mean here a general notion whose specializations include, e.g., associative, Cayley-Klein and Lie, commutative, and Poisson algebras; cf. Subsection 1.3 below.

DEFINITION 1.3 *Let $\mathcal{P}$ be an operad and V a linear space. A structure of $\mathcal{P}$-algebra on V, or equivalently, a* ***linear representation*** *of $\mathcal{P}$ in V, is a morphism of operads $\rho : \mathcal{P} \to OpEnd(V)$ sending I to id_V if $\mathcal{P}$ is unital.*

As Definition 1.1 shows, $\underline{\mathcal{P}} = \bigoplus_{l \geq 1} \mathcal{P}(l)$ has a canonical structure of $\mathcal{P}$-algebra (regular representation).

Generally, to define a structure of $\mathcal{P}$-algebra on V is the same as to define for every element $f \in \mathcal{P}(l)$ a l-ary multiplication map $m_f : V^{\otimes l} \to V$ linearly depending on f, translating γ-composition to substitution and the action of the symmetric groups to the permutation of the arguments.

DEFINITION 1.4 *Let V, W be two $\mathcal{P}$-algebras. A morphism between them is a linear map $\varphi : V \to W$ such that for every $f \in \mathcal{P}(l)$ we have*

$$\varphi(m_l^V(v_1 \otimes \cdots \otimes v_l)) = m_f^W(\varphi(v_1) \otimes \cdots \otimes \varphi(v_l)) \ . \tag{12}$$

We will show that for certain species C of k-algebras which we may call "*operadic*" one can find a unital operad COp such that COp-algebras and morphisms between them "are" algebras of the species C and their morphisms.

REMARK 1.2 *Let us start with an example, taking again associative algebras without unit, this time considered as a species. Besides the identity map* $V \to V$*, any associative algebra is commonly given by one generating bilinear multiplication* $m : V \otimes V \to V$*, but the transposition of arguments transforms it into another multiplication* m^{op}*. Therefore we must put* $AssOp(1) = \langle I \rangle$ *(brackets denoting the linear span),* $AssOp(2) = \langle m, m^{op} \rangle$*, the regular representation of* S_2*. In* $AssOp(3)$ *we have then twelve ternary operations that can be constructed from* I, m, m^{op}*: in the functional notation they are* $m(m, I)$*,* $m(I, m^{op})$*, etc. In plain words, each such operation applied to* $v_1 \otimes v_2 \otimes v_3 \in V^{\otimes 3}$ *picks two* v_i*'s, multiplies them in some order, and then multiplies the result by the remaining* v_l*.*

These twelve ternary operations are related by identities expressing the associativity of m *and its consequence, that of* $m^{op} : m(m, I) = m(I, m)$*, etc. As a result,* $AssOp(3)$ *is isomorphic to the regular representation of* S_3 *generated by, say,* $m(m, I)$*.*

The general pattern is as follows. Pick an infinite sequence of independent noncommuting but associative variables $x_1, x_2, x_3, \ldots$*. Instead of* $m, m^{op}, m(I, m^{op})$*, etc., write the value of the respective operation applied to the initial segment of this sequence, getting respectively* $x_1x_2, x_2x_1, x_1x_3x_2$*, etc. A contemplation shows that one can thus identify* $AssOp(n)$ *with the linear space generated by all associative monomials* $x_{s(1)} \ldots x_{s(n)}$ *where* $s \in \mathbf{S}_n$*, with the evident action of* $\mathbf{S}_n$*.*
Namely, $m(\ldots(m(m, I), I)\ldots)$ *produces the monomial* $(\ldots((x_1x_2)x_3)\ldots)x_n$ $= x_1 \ldots x_n$*, and the application of* $\mathbf{S}_n$ *furnishes the rest. It remains to describe the* γ*-composition of a monomial* $x_{s(1)} \ldots x_{s(n)}$ *with* $g_1 \otimes \cdots \otimes g_n \in \bigoplus_{a=1}^n AssOp(l_a)$*. We first replace arguments* $x_1, \ldots, x_{l_a}$ *in* g_a *by adding* $l_1 + \cdots + l_{a-1}$ *to all subscripts thus getting* $\widetilde{g}_a$*, and then put*

$$\gamma(x_{s(1)} \otimes \cdots \otimes x_{s(n)} \otimes g_1 \otimes \cdots \otimes g_n) := \widetilde{g}_{s(1)} \ldots \widetilde{g}_{s(n)} \ .$$

Now let us try to construct a functor from AssOp-algebras to associative algebras.

A structure of an AssOp-algebra on V*, clearly, is uniquely determined by the restriction of the operadic morphism*

$$\rho(2) : AssOp(2) \to Hom_{VECT}(V^{\otimes 2}, V) \ .$$

However, the image of $\rho(2)$ *is a two-dimensional space of multiplications* $\{am + bm^{op}\}$ *whereas classically we need just one associative multiplication. Let us write* *the associativity equation* $\mu(\mu, I) = \mu(I, \mu)$ *for* $\mu = am + am^{op}$ *in the functional notation with free arguments* x, y, z*:*

$$\mu(\mu, I) = a[(axy + byx)z] + b[z(axy + byx)] \ ,$$
$$\mu(I, \mu) = a[x(ayz + bzy) + b[(ayz + bzy)x \ .$$

Comparing coefficients, one sees that the universal *associativity (in any linear representation) is equivalent to* $ab = 0$*. Hence the best we can do is to pinpoint in any AssOp-algebra* V two lines of associative multiplications: $\langle\rho(m))\rangle$ *and* $\langle\rho(m^{op})\rangle$*. An additional choice of unit would reduce each line to one (non-zero) element, however there is nothing in the structure of AssOp that would help us to do this. In fact, we encounter here a general problem: how to account for eventual structural special elements, i.e. 0-ary operations. In principle, we could have extended the definition of an operad* $\mathcal{P}$ *by including* $\mathcal{P}(0)$ *and extending correspondingly (1). In particular, we can put* $AssOp(0) =$ *ground field,* $OpEnd(V)(0) = V$*, and define the identity in* V *as the image of 1. In other cases this might not work.*

We will now summarize the preceding discussion in a deliberately vague "metatheorem" (for more precise statements, see below).

1.3 Species of algebras and operads

Let C be a category of algebras which is defined by a family of multilinear operations $\{m_i | i \in I\}$ and a family of universal identities between them constructed of compositions and linear combinations. Morphisms in C are linear maps compatible with m_i's. Examples: associative algebras without identity (multiplication; associativity); Cayley-Klein and Lie algebras (bracket; skew-symmetry; Jacobi identity); Poisson algebras without identity (multiplication, bracket; associativity, commutativity, Jacobi, Leibniz); commutative rings with an m-dimensional linear space of pairwise commuting derivations, etc.

Then one can construct a classical linear operad COp with the following properties.

a) $COp(l)$ as a representation space of $\mathbf{S}_l$ is isomorphic to a subspace of the free algebra $F_C(x_1, \dots, x_l)$ of the species C freely generated by l independent variables $x_1, \dots, x_l$. This subspace consists of forms of total degree l linear in each x_a, upon which $\mathbf{S}_l$ acts by permuting arguments.

b) Compositions γ are induced by substitution.

c) To give a structure of a COp-algebra on a space V is the same as giving a set of structures of species C on V. Various elements of this set are obtained by choosing in COp various generating families of solutions $\{m_i'\}$ of the universal identities defining C. The group $Aut(COp)$ acts transitively on this set.

d) The category of COp-algebras is equivalent to the category of algebras of the species C. Every choice of generators $\{m_i\}$, as above fixes one equivalence functor. However, two different choices may lead to non-isomorphic functors.
This happens, e.g., with $AssOp$ and functors corresponding to m and m^{op}.

To give a precise statement and proof of this theorem, we would have to explain in more detail the two different notions of “freeness” and “defining an object by generators and relations”: separately for operads and algebras over a given operad. Above we used them on an intuitive level.

Before proceeding further, we want to list the limitations of the operadic approach to species, some of which can be overcome by modifying the notion of the classical operad.

- We cannot account for the structure constants, partly because of the lack of $\mathcal{P}(0)$.
- We cannot account for the use of dual spaces in the definitions of some species, e.g., algebras with invariant scalar products interpreted as $V \to V^*$. (In this case, a remedy is the introduction of the cyclic operads).
- We cannot account for the structure morphisms like comultiplication $V \to V \otimes V$, and generally tensors of various co- and contravariant degrees.
- We cannot account for non-linear and not everywhere defined operations like inversion in the multiplicative group of a field.

1.4 Operads as analogs of associative algebras

In Example 1.1 and Subsection 1.1 we have shown that the associative algebras naturally form a part of the classical operads (with $\mathcal{P}(l) = 0$ for $l \geq 2$). We will now demonstrate that the total classical operad $\mathcal{P}$ is in a very definite sense an analog of associative algebra.

To do this convincingly, we must start with a definition of an associative algebra as a couple (V, m) where V is an object of the monoidal category $VECT$, and m is an associative morphism $V \otimes V \to V$, eventually endowed with identity which is a morphism $\mathbf{1} \to V$ where $\mathbf{1}$ is the ground field considered as an identity object in $VECT$. The two categories $(VECT, \otimes)$ and ASS obtained in this way are connected by the two adjoint functors

$$\text{forget } m : ASS \to VECT\ ,$$
$$\text{free tensor algebra} : VECT \to ASS\ .$$

In order to present the classical linear operads in the same way we have to start with specifying an analog of the functor "forget m". This can be done in several ways because we can choose to forget any subset of the data given in Definition 1.1. Here we will decide that m corresponds to all γ's. What is left then is the following category $SMOD$ of $\mathbf{S}$-modules:

Definition 1.5 *An object of $SMOD$ is a family of linear spaces $V(l), l \geq 1$, endowed with an action of $\mathbf{S}_l$.*

A morphism in $SMOD$ is a family of linear maps $V(l) \to W(l)$ compatible with the $\mathbf{S}_l$-action.

We will sometimes say that $V(l)$ is the part of V of degree l.

Lemma 1.1 *a) The category $SMOD$ possesses a bifunctorial product $*$ which can be defined on the objects by the following formula:*

$$V * W(n) = \bigoplus_{l=1}^{n} V(l) \otimes_{\mathbf{S}_l} \left(\bigoplus_{\pi:\underline{n}\to\underline{l}} \bigotimes_{i=1}^{l} W(|\pi^{-1}(i)|) \right)\ . \tag{13}$$

Here $\underline{n} = \{1, \ldots, n\}$ and π runs over all surjective maps. The action of $\mathbf{S}_l$ must be self-explanatory, and the tensor product is taken over the group ring of $\mathbf{S}_l$.

This product is functorially associative but not commutative so that $(SMOD, *)$ is a monoidal category. It possesses a two-sided identity object $\mathbf{1}$: the ground field placed in degree 1, zero elsewhere.

b) The map $V \mapsto (V, 0, 0, \ldots)$ extends to a functor identifying $(VECT, \otimes, \mathbf{1})$ with a full monoidal subcategory of $(SMOD, *, \mathbf{1})$.

Now consider an associative algebra (V,μ), $\mu : V * V \to V$ *in the monoidal category of* $\mathbf{S}$*-modules.* From (13) we see that μ is a family of maps

$$\mu(n) : \bigoplus_{l=1}^{n} V(l) \otimes \mathbf{S}_l \left(\bigoplus_{\pi:\underline{n}\to\underline{l}} \bigotimes_{i=1}^{l} V(|\pi^{-1}(i)|) \right) \to V(n),\ n \geq 1\ . \tag{14}$$

For given $(l; k_1,\dots,k_l)$, $k_1+\cdots+k_l = n$, consider the component of (14) corresponding to the $\mathbf{S}_l$-orbit of the map sending $\{1,\dots,k\}$ to $1, k_1+1,\dots,k_1+k_2$ to 2, etc. We can identify this part of the source with $V(l)\otimes V(k_1)\otimes\cdots\otimes V(k_l)$ so that μ generates a family of maps

$$\gamma(k_1,\dots,k_l) : V(l)\otimes V(k_1)\otimes\cdots\otimes V(k_l) \to V(k_1+\ldots+k_l)\ . \tag{15}$$

PROPOSITION 1.1 *a) The associativity of μ translates into the associativity of γ's in the sense of (7).*

b) The fact that μ is a morphism in $SMOD$ translates into the compatibility axioms (3–6)

c) In this way we get a functor

$$\text{Associative algebras in } (SMOD, *) \to OPER$$

which is an equivalence of categories.
There exists a similar equivalence between associative algebras with identity and unital operads.

Proof. We will now sketch a proof of the main statements in Lemma 1.1 and Proposition 1.1. In order to understand the main formula (13), we will show that it expresses the substitution law of "formal series in $VECT$".

To be more precise, denote by $FSETS$ the category of finite non-empty sets and bijections. Let $F[\cdot] : FSETS \to VECT$ be a functor.

$FSETS$ is equivalent to its full subcategory whose objects are $\underline{n}$. Restricting F to this subcategory we get an $\mathbf{S}$-module $V_F : V_F(n) := F[\underline{n}]$, the action of $\mathbf{S}_n$ being induced by the bijections of $\underline{n}$.

Now consider $V_F(n)$ as coefficients of the formal series defining the functor $F(\cdot) : VECT \to VECT$:

$$F(X) := \bigoplus_{n\geq 1} V_F(n) \otimes_{\mathbf{S}_n} X^{\otimes n}\ .$$

Such functors will be called *analytic ones.*

We will show that these constructions establish an equivalence of the three categories involved: functors $F[\cdot]$ and their morphisms, $SMOD$, analytic functors. Moreover, the composition of analytic functors is again analytic, and it induces on the coefficients exactly the *-product:

$$V_{F\circ G}(n) = (V_F * V_G)(n) \ .$$

The equivalence of the category of functors $F[\cdot]$ and $SMOD$ is a part of general nonsense because $FSETS$ is equivalent to its subcategory of natural numbers. The only point deserving explication is the possibility to lift every $\mathbf{S}$-module to an $F[\cdot]$ canonically without using the axiom of choice. Namely, for a finite set M with $|M| = m$ put

$$\widetilde{F}[M] := F[\underline{m}] \otimes_{\mathbf{S}_m} \langle Iso(\underline{m}, M)\rangle \ .$$

Here $\langle Iso(\underline{m}, M)\rangle$ is the linear space freely generated by the bijections $\underline{m} \to M$. Strictly speaking, now $\widetilde{F}[\underline{m}]$ is not $F[\underline{m}]$, but these $\mathbf{S}_m$-modules are canonically isomorphic, and we forget about this subtlety and say, for example, that the $\mathbf{S}$-module $F[\underline{m}] = X^{\otimes m}$ extends to the functor $F[M] = X^{\otimes M}$ on the category of finite sets.

The equivalence of $SMOD$ and the category of analytic functors $F(\cdot)$ also becomes a formal fact once we learn how to reconstruct functorially the coefficients $V_F(n)$. Let $F(\cdot)$ be given. Multiplication by any element λ of the ground field is an endomorphism of the identical functor of $VECT$. Hence it acts functorially on each $F(X)$, and the λ^n-eigenspace of $F(X)$ is exactly $F_n(X) := V_F(n) \otimes_{\mathbf{S}_n} X^{\otimes n}$, at least when λ is not 0 or a root of unity. Now consider the space $X_n = \langle \underline{n}\rangle$ freely generated by the vectors $e_1, \dots, e_n$. Then $e_1 \otimes \dots \otimes e_n$ generates the regular $\mathbf{S}_n$-submodule R_n which is the image of the projector $p_n : X_n^{\otimes n} \to X_n^{\otimes n}$. Since F_n is a functor, we can define $Im(F_n(p_n)) = V_F(n) \otimes_{\mathbf{S}_n} R_n = V_F(n)$, both equalities denoting canonical isomorphisms.

We will apply this prescription to the calculation of the coefficients of the composition of analytic functors:

$$(F \circ G)(X) = \bigoplus_{l=1}^{\infty} V_F(X) \otimes_{\mathbf{S}_l} \left(\bigoplus_{k=1}^{\infty} V_G(k) \otimes_{S_k} X^{\otimes k} \right)^{\otimes l} =$$

$$= \bigoplus_{l=1}^{\infty} V_F(X) \otimes_{\mathbf{S}_l} \left[\bigoplus_{k_1,\dots,k_l=1}^{\infty} (V_G(k_1) \otimes_{S_{k_1}} X^{\otimes k_1}) \otimes \cdots \otimes (V_G(k_l) \otimes_{S_{k_l}} X^{\otimes k_l} \right] .$$

It follows that

$$(F \circ G)_n(X) = \bigoplus_{l=1}^{\infty} V_F(X) \otimes_{\mathbf{S}_l} \left[\bigoplus_{k_1+\cdots+k_l=n} \bigotimes_{a=1}^{l} (V_G(k_a) \otimes_{S_{k_a}} X^{\otimes k_a}) \right] .$$

Now we must put $X = X_n$ as above and look at the image of $(F \circ G)_n(p_n)$ or, more intuitively, at the tensor coefficients of the vectors $e_{s(1)} \otimes \cdots \otimes e_{s(n)}$. Clearly, for a given l, such terms in square brackets correspond to the partitions of $\underline{n}$ into l blocks indexed by $1, \dots, l$, i.e. to the surjections $\underline{n} \to \underline{l}$ as in (13).

To finish the Proof, it remains to establish that the functor $F \circ G(\cdot)$ is isomorphic to the sum of $\mathrm{Im}(F \circ G)_n(p_n)$. We leave this to the reader. ■

DEFINITION 1.6 *Let V be an object of $SMOD$. Put*

$$F(V) := \sum_{n=1}^{\infty} V^{*n} .$$

There is an obvious multiplication map $V^{*m} * V^{*n} \to V^{*m+n}$ which makes $F(V)$ an associative algebra, or an operad. It is called the ***free operad*** *generated by V (without identity).*

As in the classical linear algebra, F is adjoint to the forgetful functor $OPER \to SMOD$. This completes the analogy sketched at the beginning of Subsection 1.4.

1.5 Operads and topology: homology of moduli spaces

We will now introduce the basic operad of the quantum cohomology. Denote by $H_*(\overline{M}_{0,n+1})$ the homology space of the moduli space of stable curves of genus zero (with coefficients in the ground field for $VECT$). We will define the classical linear operad $H_*\overline{M}_0$ by the following data:

a) $H_*\overline{M}_0(n) = H_*(\overline{M}_{0,n+1})$ for $n \geq 2$, the first component being the ground field.

In the following, it will be convenient to assume that the structure sections of $\overline{C}_{n+1} \to \overline{M}_{0,n+1}$ are labeled by $\{0, \dots, n\}$.

b) $\mathbf{S}_n$ acts upon $H_*\overline{M}_0(n)$ by renumbering the sections $x_1, \dots, x_n$.

c) The structure map

$$\begin{aligned}&\gamma(k_1, \dots, k_l) :\\ &H_*(\overline{M}_{0,l+1}) \otimes H_*(\overline{M}_{0,k_1+1}) \otimes \dots \otimes H_*(\overline{M}_{0,k_l+1}) \to H_*(\overline{M}_{0,k_1+\dots+k_l+1})\end{aligned} \tag{16}$$

is induced by the embedding of the boundary stratum

$$b(k_1, \dots, k_l) : \overline{M}_{0,l+1} \times \overline{M}_{0,k_1+1} \times \dots \times \overline{M}_{0,k_l+1} \to \overline{M}_{0,k_1+\dots+k_l+1} \ . \tag{17}$$

On the level of geometric points, given $l+1$ stable labeled curves of genus zero,

$$(C; x_0, x_1, \dots, x_l); \quad (D_a; y_{0,a}, \dots, y_{k_a,a}), \quad a = 1, \dots, l \ .$$

$b(k_1, \dots, k_l)$ produces from them the stable curve

$$\left(C \coprod \left(\coprod_{a=1}^{l} D_a \right) / (\sim); \quad z_0, \dots, z_{k_1} + \dots + k_l \right) \ ,$$

where $(\sim)$ is the equivalence relation gluing x_a and $y_{0,a}$ for all $a = 1, \dots, l$, and furthermore

$$z_0 = x_0, (z_1, \dots, z_{k_1+\dots+k_l}) = (y_{1,1}, \dots, y_{k_1,1}; \dots; y_{1,a}, \dots, y_{k_a,a}).$$

Operadic axioms for $H_*\overline{M}_0$ follow from their evident versions for the spaces $\overline{M}_{0n}$.

Remark 1.3 *What we are actually saying here is that we can define the more general notion of operad by replacing the basic category $VECT$ by any symmetric monoidal category, eventually with the identity object, and that the moduli spaces form such an operad. The homology functor (with respect to the pushforward maps)* *from the monoidal category of manifolds to* $(VECT, \otimes)$ *then produces from a geometric operad the classical linear operad. This viewpoint will be discussed in more detail in the following Section.*

Notice in conclusion that $H_*\overline{M}_0$ *is endowed with important additional structures. Namely, the components of this operad are in fact coalgebras (pushforward with respect to the diagonal map), and compositions (16) as well as representations of* $\mathbf{S}_n$ *are coalgebra morphisms. This is the intrinsic reason for the existence of the operation of the tensor product on the category of* $H_*\overline{M}_0$*-algebras.*

2 Operads and Trees

In this Section we sketch in their natural generality several themes which have already emerged in the previous Section. Briefly speaking, there are many useful types of operads, and each type is determined by the choice of two categories:

1) Basic symmetric monoidal category $(\mathcal{C}, \boxtimes)$ replacing $(VECT, \otimes)$ which supports the classical linear operads.

2) A category of (labeled) graphs Γ reflecting the combinatorics of the operadic data and axioms.

A concrete operad from this viewpoint is a functor $\Gamma \to \mathcal{C}$.

To clarify the role of $\mathcal{C}$, we first explain how to extend Definition 1.1.

2.1 May's operads in a monoidal category

Let us recall (see [5]) that a *symmetric monoidal category* $(\mathcal{C}, \boxtimes)$ is a category endowed with the bifunctor $\boxtimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ together with an involutive commutativity constraint and an associativity constraint. Taken together, they define a family of compatible and functorial isomorphisms $s_* : X_1 \boxtimes \cdots \boxtimes X_n \widetilde{\to} X_{s^{-1}(1)} \boxtimes \cdots \boxtimes X_{s^{-1}(n)}$, for any objects $X_1, \dots, X_n$ of $\mathcal{C}$ and all $s \in \mathbf{S}_n$.

Most of our monoidal categories will have an identity object $\mathbf{1}_{\mathcal{C}} = \mathbf{1}$. The functors $\mathbf{1}\boxtimes$ and $\boxtimes\mathbf{1} : \mathcal{C} \to \mathcal{C}$ are canonically isomorphic to the identity functor.

In order to be able to extend the constructions of Subsection 1.4, we will assume that $\mathcal{C}$ has small colimits preserved by any functor $X\boxtimes$. In particular, $\mathcal{C}$ must have an initial object 0.

We can now define *a classical operad* $\mathcal{P}$ in $\mathcal{C}$ by closely following Definition

1.1. Components $\mathcal{P}(n)$ will be objects of $\mathcal{C}$ endowed with the action of $\mathbf{S}_n$, $\otimes$ will be replaced by $\boxtimes$, and operadic multiplications γ will be morphisms in $\mathcal{C}$. Axioms $A) - C)$ must be written down as commutative diagrams, involving in particular permutation isomorphisms of tensor products in $\mathcal{C}$.

A neater version of the definition is again obtained by passing to the category SC every object which is a family of $\mathbf{S}_n$-objects $\mathcal{P}(n)$ in $\mathcal{C}$ given for $n \geq 1$. To be able to write it as a sum of its components, we will require that $\mathcal{C}$ has small limits. The category $\mathcal{SC}$ admits a *non-symmetric* monoidal structure *, furnished by the formula (13). It has the unit object $\mathbf{1}_{\mathcal{SC}}$ with $\mathbf{1}$ as the first component, 0 elsewhere. An associative monoid in $\mathcal{SC}$ is a pair $(\mathcal{P}, \mu)$ where $\mu : \mathcal{P} * \mathcal{P} \to \mathcal{P}$ is an associative multiplication. Giving an additional morphism $\mathbf{1} \to \mathcal{P}$ with the usual properties defines unital monoids. An analog of Proposition 1.1 holds true, establishing the equivalence of the category of associative monoids in $(\mathcal{SC}, *)$ and the category of classical operads in $\mathcal{C}$. However, the proof of Proposition 1.1 must be changed, because we have used in it not only the monoidal structure of $VECT$ but the linear structure and the language of elements as well. This can be avoided in different ways. Here we will take this fact for granted, and we leave to the reader the transposition of other constructions of Section 1 to the present context.

REMARK 2.1 *Let us consider the main classes of monoidal categories. Sets with direct product and linear spaces with tensor product form two archetypal classes of symmetric monoidal categories.*

Variations include imposing additional structure on the objects. Sets more often appear endowed with a topology or manifold structure (in smooth or analytic category). Linear spaces come equipped with grading and/or differential. In this way we get classical topological operads, classical operads in the category of complexes, and so on. Monoidal functors between symmetric monoidal categories extend to the respective categories of operads.

2.2 Oriented trees as substitution schemes

Let T be a *tree* with at least two flags at each vertex. Orient T by choosing one tail as *root* and declaring that direction to the root is positive. Then every vertex has at least one incoming flag and exactly one outgoing flag. Label each vertex of T by a symbol of the function whose arguments are labeled by the incoming flags of this vertex, and whose value labels the outgoing flag f and also its l-image, that is, the other half of the edge if f belongs to an edge. Then the whole tree symbolizes a computation, or substitution scheme. The input values are assigned to the incoming tails of T, and the output value is assigned to the root. For example, one vertex tree with n incoming tails symbolizes $f(x_1,\dots,x_n)$ and the $(m+1)$-vertex tree with the appropriate distribution of flags symbolizes $f(g_1(x_1^{(1)},\dots,x_{n_1}^{(1)}),\dots,g_m(x_1^{(m)},\dots,x_{n_m}^{(m)}))$.

If we label flags by objects of a symmetric monoidal category and label each vertex v by a morphism mapping the $\boxtimes$-product of the labels of incoming flags to the label of the outgoing flag, the tree will describe the respective composite morphism from the $\boxtimes$-product of input objects to the output object.

If $\boxtimes$ is not supposed to be symmetric, we must assume that all sets of incoming flags of each vertex are totally ordered. For a symmetric $\boxtimes$-product, the respective actions of symmetric groups on arguments of various levels can be succinctly described by saying that this construction is functorial on the category of oriented trees with isomorphisms compatible with orientation.

2.3 Trees and and *-product

Let $(\mathcal{C},\boxtimes)$ be a *symmetric monoidal category*, V an object of *non-symmetric monoidal category* $(\mathcal{SC},*)$, and $F(V)=\coprod_{n=1}^{\infty}V^{*n}$ the free operad generated by V, as in Definition 1.6. We can define $\boxtimes$-products indexed by arbitrary finite sets and extend $n\mapsto V(n)$ to a functor $T\mapsto V(T)$ on the category of non-empty finite sets and their bijections. For an oriented tree as above put

$$V(T):=\boxtimes_{v\in V_T}V(F_T^{\mathrm{in}}(v))\ . \tag{18}$$

Then we have functorial isomorphisms

$$F(V)(n)=\coprod_{\{n-\mathrm{trees}\}T/(iso)}V(T)\ . \tag{19}$$

Here n-trees are oriented trees with the set $\{1,\dots,n\}$ of incoming tails.

This statement summarizes in a more conceptual way the bookkeeping scheme described above. It can be deduced with some pain from the formalism of analytic functors as in Subsection 1.4. We will also reproduce the relevant combinatorics in the context of formal series below, in Section 12 dedicated to sums over trees.

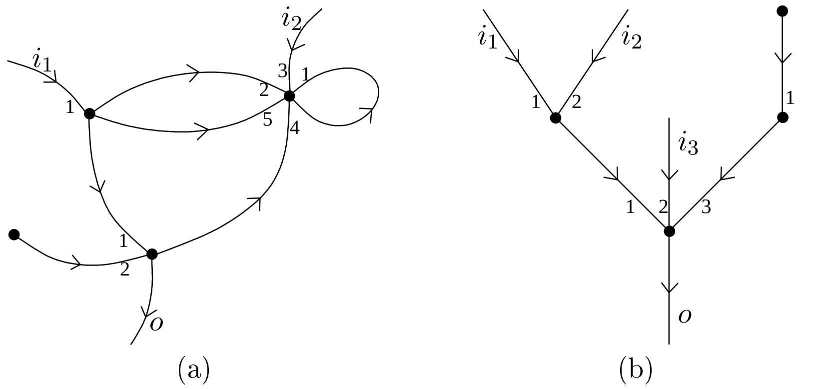


Figure 2: (a) Input-output graph with 4 vertices and 2 input edges i_1, i_2, (b) combinatorial tree with 4 vertices and 3 input edges i_1, i_2, i_3. In both, the numbers indicate the order on the edges arriving at each vertex.

2.4 Combinatorial trees

In Subsections 2.2–2.3 trees were defined in a purely abstract way: T is the free plain operad on the terminal object of $\mathtt{Set}_f^{\mathbb{N}}$, and an n-leafed tree is an element of T_n. But we give here a graph-theoretic definition of (finite, rooted, planar) tree.

The main subtlety is that the trees we use are not quite finite graphs in the usual sense: some of the edges have a vertex at only one of their ends. This suggests the following definitions.

Definition 2.1

A (planar) ***input − output graph*** *(Fig. 2(a)) consists of*

- *a finite set V (the* ***vertices****)*
- *a finite set E (the* ***edges****), a subset $I \subseteq E$ (the* ***input edges****), and an element $o \in E$ (the* ***output edge****)*
- *a function $s : E \backslash I \longrightarrow V$ (****source****) and a function $t : E \backslash \{o\} \longrightarrow V$ (****target****)*
- *for each $v \in V$, a total order $\leq$ on $t^{-1}\{v\}$.*

We write $v \xrightarrow{e}$ to mean that e is a non-input edge with $s(e) = v$, and similarly $\xrightarrow{e} v'$ to mean that e is a non-output edge with $t(e) = v'$, and of course $v \xrightarrow{e} v'$ to mean that e is a non-input, non-output edge with $s(e) = v$ and $t(e) = v'$.

A tree is roughly speaking a connected, simply connected graph, and the following notion of path allows us to express this.

DEFINITION 2.2 *A **path** from a vertex v to an edge e in an input-output graph is a diagram*

$$v = v_1 \xrightarrow{e_1} v_2 \xrightarrow{e_2} \cdots \xrightarrow{e_{l-1}} v_l \xrightarrow{e_l = e}$$

in the graph. That is, a path from v to e consists of

- *an integer $l \geq 1$*
- *a sequence $(v_1, v_2, \ldots, v_l)$ of vertices with $v_1 = v$*
- *a sequence $(e_1, \ldots, e_{l-1}, e_l)$ of edges with $e_l = e$*

such that

$$v_1 = s(e_1),\; t(e_1) = v_2 = s(e_2),\; \ldots,\; t(e_{l-1}) = v_l = s(e_l)$$

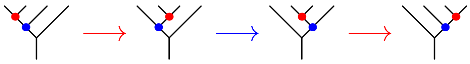
and all of these sources and targets are defined.

DEFINITION 2.3 *A **combinatorial tree** is an input-output graph such that for every vertex v, there is precisely one path from v to the output edge.*

Fig. 2(b) shows a combinatorial tree. The ordering of the edges arriving at each vertex encodes the planar embedding. 'Tree' is an abbreviation for 'finite, rooted, planar tree'. If we were doing symmetric operads then we would use non-planar trees, if we were doing cyclic operads then we would use non-rooted trees, and so on.

2.5 Geometric interpretation of trees

Let T_n be the set of rooted, planar binary trees with n interior nodes (and thus $n+1$ leaves). The *Tamari order* (see [2]) on T_n is the partial order whose cover relations are obtained by moving a child node directly above a given node from the left to the right branch above the given node. Thus



is an increasing chain in T_3 (the moving vertices are marked with dots). Only basic properties of the Tamari order are needed in this Subsection; their proofs will be provided. For more properties, see Chapters 3–7. Figure 3 shows the Tamari order on T_3 and T_4.

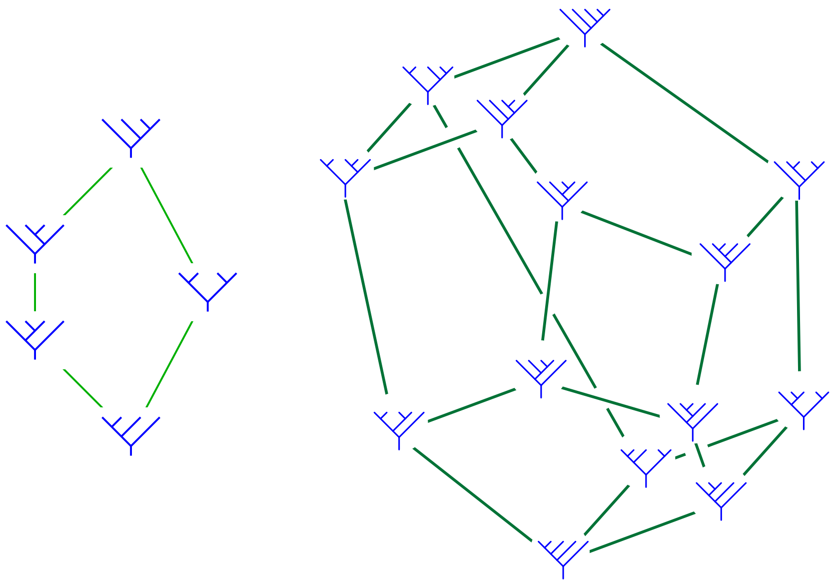
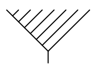
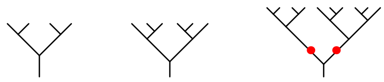


Figure 3: The Tamari order on T_3 and T_4.

Let 1_n be the minimum tree in T_n. It is called a *right comb* as all of its leaves are right pointing:

1_4 =  , 1_7 =  .

Given trees $s \in T_p$ and $t \in T_q$, the tree $s \vee t \in T_{p+q+1}$ is obtained by grafting the root of s onto the left leaf of the tree  and the root of t onto its right leaf. Below we display trees s, t, and $s \vee t$, indicating the position of the grafts with dots.



For $n > 0$, every tree $t \in T_n$ has a unique decomposition $t = t_l \vee t_r$ with $t_l \in T_p$, $t_r \in T_q$, and $n = p+q+1$. Thus T_n is in bijection with $\coprod_{p+q=n-1} T_p \times T_q$, and since $T_0 = \{|\}$ and $T_1 = \{\curlyvee\}$, we shall see in Section 3.5 that T_n contains the Catalan number $\frac{(2n)!}{n!(n+1)!}$ of trees.

The Hasse diagram of T_n is isomorphic to the 1-skeleton of the *associahedron* $\mathcal{A}_n$, an $(n-1)$-dimensional polytope. (See [6] and [7].) The faces of $\mathcal{A}_n$ are in one-to-one correspondence with collections of non-intersecting diagonals of a polygon with $n+2$ sides (an $(n+2)$-gon). Equivalently, the faces of $\mathcal{A}_n$ correspond to polygonal subdivisions of an $n+2$-gon with facets corresponding to diagonals and vertices to triangulations. The dual graph of a polygonal subdivision is a planar tree and the dual graph of a triangulation is a planar binary tree. If we distinguish one edge to be the root edge, the trees are rooted, and this furnishes a bijection between the vertices of $\mathcal{A}_n$ and T_n. Figure 4 shows two views of the associahedron $\mathcal{A}_3$, the first as polygonal subdivisions of the pentagon, and the second as the corresponding dual graphs (planar trees). The root is at the bottom.

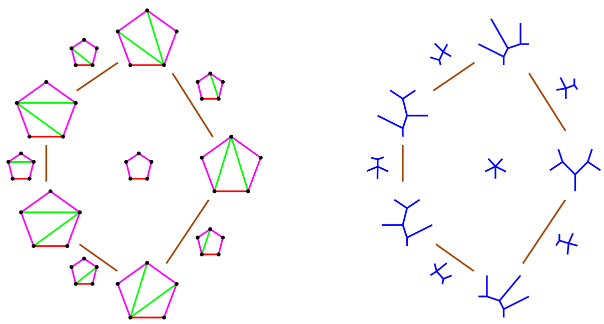


Figure 4: Two views of the associahedron $\mathcal{A}_3$

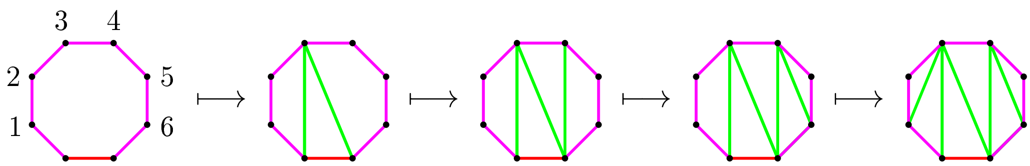
Let $\mathfrak{S}_n$ be the group of permutations of $[n]$ which denotes the set $\{1, 2, \ldots, n\}$. We describe the map $\lambda\colon \mathfrak{S}_n \to T_n$ in terms of triangulations of the $(n+2)$-gon where we label the vertices with $0, 1, \ldots, n, n+1$ beginning with the left vertex of the root edge and proceeding clockwise. Let $\sigma \in \mathfrak{S}_n$ and set $w_i := \sigma^{-1}(n+1-i)$, for $i = 1, \ldots, n$. This records the positions of the values of σ taken in decreasing order. We inductively construct the triangulation, beginning with the empty triangulation consisting of the root edge, and after i steps we have a triangulation τ_i of the polygon

$$P_i \ := \ \mathit{Conv}\{0, n{+}1, w_1, \ldots, w_i\}\,.$$

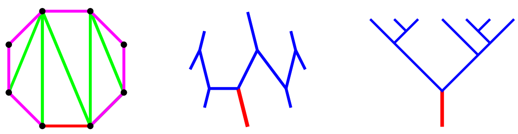
Some edges of P_i will be edges of the original $(n{+}2)$-gon and others will be diagonals. Each diagonal cuts the $(n{+}2)$-gon into two pieces, one containing P_i and the other a polygon which is not yet triangulated and whose root edge we take to be that diagonal. Subsequent steps add to the triangulation τ_i and its support P_i.

First set $\tau_1 := \mathit{Conv}\{0, n{+}1, w_1\}$, the triangle with base the root edge and apex the vertex $w_1 = \sigma^{-1}(n)$. Set $P_1 := \tau_1$ and continue. After i steps we have constructed τ_i and P_i in such a way that the vertex w_{i+1} is not in P_i. Hence it must lie in some untriangulated polygon consisting of some consecutive edges of the $(n{+}2)$-gon and a diagonal that is an edge of P_i. Add the join of the vertex w_{i+1} and the diagonal to the triangulation to obtain a triangulation τ_{i+1} of the polygon P_{i+1}. The process terminates when $i = n$.

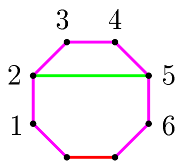
For example, we display this process for the permutation $\sigma = 316524$, where we label the vertices of the first octagon:

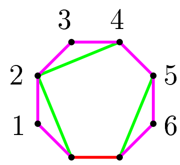


The last two steps are supressed as they add no new diagonals. The dual graph to the triangulation τ_n is the planar binary tree $\lambda(\sigma)$. Here is the triangulation, its dual graph, and a 'straightened' version, which we recognize as the tree $\lambda(316524)$.



A subset S of $[n]$ determines a face Φ_{S} of the associahedron $\mathcal{A}_n$ as follows. Suppose that we label the vertices of the $(n{+}2)$-gon as above. Then the vertices labeled $0, n+1$ and those labeled by S form a $(\#\mathsf{S}+2)$-gon whose edges include a set E of non-crossing diagonals of the original $(n+2)$-gon. These diagonals determine the face Φ_{S} of $\mathcal{A}_n$ corresponding to S. We give two examples of this association when $n = 6$ below.

$\{1,2,5,6\} \longleftrightarrow$ 

$\{2,4,5\} \longleftrightarrow$ 

We determine the image of f_ζ using the above description of the map $\lambda\colon \mathfrak{S}_n \to T_n$. We say that a face of $\mathcal{A}_{p+q}$ of the form Φ_{S} with $\#\mathsf{S} = q$ has *type* (p,q). If a face has a type, this type is unique. A permutation $\zeta \in \mathfrak{S}^{(p,q)}$ is uniquely determined by the set $\zeta\{p+1,\ldots,p+q\}$. Therefore, a face of type (p,q) is the image of f_ζ for a unique permutation $\zeta \in \mathfrak{S}^{(p,q)}$. This allows us to speak of the vertex of the face corresponding to a pair $(s,t) \in T_p \times T_q$ (under f_ζ).

2.6 Classical operads as functors

Denote by $\mathtt{Tree}_{clas}$ as the category whose objects are finite rooted trees with the following properties: a) the multiplicity of each vertex is at least two; b) at each vertex either all incoming flags are halves of edges, or all incoming flags are tails. Morphisms are generated by the following two classes of maps:

a) Isomorphisms compatible with orientation.

b) Contraction of all edges having a common vertex with some outgoing flag and keeping orientation.

More formally, a morphism $\varphi : \sigma \to \tau$ consists of two maps $\varphi_V : V_\sigma \to V_\tau$ and $\varphi^F : F_\tau \to F_\sigma$ compatible with boundaries and involutions and such that φ^F sends tails to tails. Composition of the morphisms corresponds to the composition of the induced maps on vertices and flags. A morphism contracts an edge e if φ_V glues its vertices, and both flags of this edge do not belong to the image of φ^F.

Contractions of different edges commute in an evident sense.

Let v be a vertex of a rooted tree T. Its *star* T_v is a one-vertex tree with vertex v, tails $F_T(v)$, and the outcoming flag as a root.

PROPOSITION 2.1 *The category of classical linear operads (without identity) in a symmetric monoidal category* $(\mathcal{C}, \boxtimes)$ *is equivalent to the category of functors* $\mathcal{P} : \mathtt{Tree}_{class} \to \mathcal{C}$ *isomorphic to a functor satisfying the following condition:*

$$\mathcal{P}(T) = \boxtimes_{v \in V_T} \mathcal{P}(T_v) \ . \tag{20}$$

Sketch of Proof. *a) From functors to operads.* Given such a functor $\mathcal{P}$, we construct the data of Definition 1.1 in the following way: $\mathcal{P}(l) := \mathcal{P}(T_l)$ where T_l is the one-vertex tree with tails $0, 1, 2, \dots, l$ and root 0. The action of $\mathbf{S}_l$ corresponds to the automorphisms of T_l permuting the tails $1, \dots, l$. The multiplication map $\gamma(k_1, \dots, k_l)$ corresponds to the morphism contracting all edges $\sigma \to \tau_{k_1 + \cdots + k_l}$, where σ has $l+1$ vertices and l edges and the tails are distributed in an obvious way. The relations A) and B) follow from the functoriality.

b) *From operads to functors.* Given an operad $(\mathcal{P}(n), \gamma)$, we first extend it to the functor from finite sets to $\mathcal{C}$, then define $\mathcal{P}(T)$ by (18), and finally use γ in order to define $\mathcal{P}$ on morphisms contracting all edges having a common vertex with some outgoing flag. ■

DEFINITION 2.4 *From the graph-theoretic viewpoint it would be more natural to allow all rooted trees with* $|v| \geq 2$ *as objects, and contractions of any subset of edges as morphisms. The functors from this category* $\mathtt{Tree}_M$ *to* $\mathcal{C}$ *satisfying (20) (up to functor isomorphism) are called* ***Markl's operads***.

REMARK 2.2 *Consider now the category* $\mathtt{Tree}_{cyc}$ *of finite non-rooted trees with* $|v| \geq 2$, *with morphisms generated by contraction of edges and isomorphisms. Neither root nor orientation is a part of the structure. Functors* $\mathtt{Tree}_{cyc} \to \mathcal{C}$ *satisfying (20) are essentially* cyclic operads *in the sense of [8]. The most essential new feature of cyclic operads is the action of* $\mathbf{S}_{l+1}$ *upon* $\mathcal{P}(l)$.

2.7 Classifying space of the category of stable trees

Let us consider a graphical definition of a category of trees. By Definition 2.3, $\mathbf{tr}$ is the free plain operad on the terminal object of $\mathtt{Set}_f^{\mathbb{N}}$, and an n-leafed tree is an element of $\mathbf{tr}_n$. As we saw, the sets $\mathbf{tr}_n$ also admit the following recursive description:

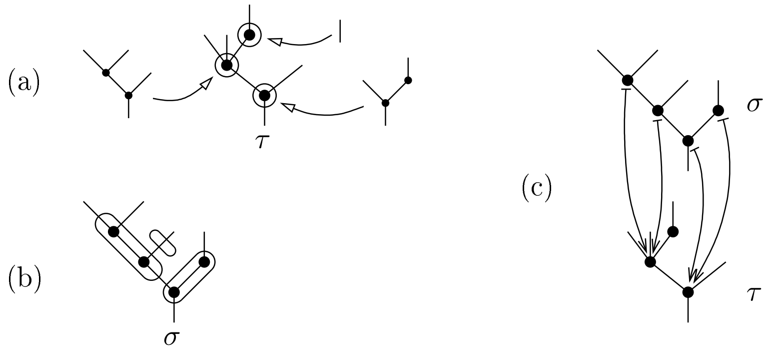


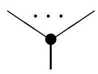
Figure 5: Three pictures of a map in $\texttt{Tree}_4$

- $| \in \mathbf{tr}_1$
- if $n, k_1, \ldots, k_n \in \mathbb{N}$ and $\tau_1 \in \mathbf{tr}_{k_1}, \ldots, \tau_n \in \mathbf{tr}_{k_n}$ then $(\tau_1, \ldots, \tau_n) \in \mathbf{tr}_{k_1+\cdots+k_n}$.

A category of trees $\texttt{Tree}$ is the disjoint union $\coprod_{n\in\mathbb{N}} \texttt{Tree}_n$. An object of $\texttt{Tree}_n$ is an n-leafed tree. The set of maps in $\texttt{Tree}_n$ is

$$(T_2^2 1)(n) = (T_2(\mathbf{tr}))(n),$$

that is, a map is an n-leafed tree τ in which each k-ary vertex v has assigned to it a k-leafed tree σ_v; the domain of the map is the tree obtained by gluing the σ_v's together in the way dictated by the shape of τ, and the codomain is τ itself. Put another way, what a map does is to take a tree σ (the domain), partition it into a finite number of (possibly trivial) subtrees, and replace each of these subtrees by the corolla



with the same number of leaves, to give the codomain τ. Fig. 5 depicts a certain map $\sigma \longrightarrow \tau$ in $\texttt{Tree}_4$ in three different ways: in (a) as a 4-leafed tree τ with a k-leafed tree σ_v assigned to each k-ary vertex v, in (b) as a 4-leafed tree σ partitioned into subtrees σ_v, and in (c) as something looking more like a function. We will return to the third point of view later; for now, just observe that there is an induced function from the vertices of σ to the vertices of τ, in which the

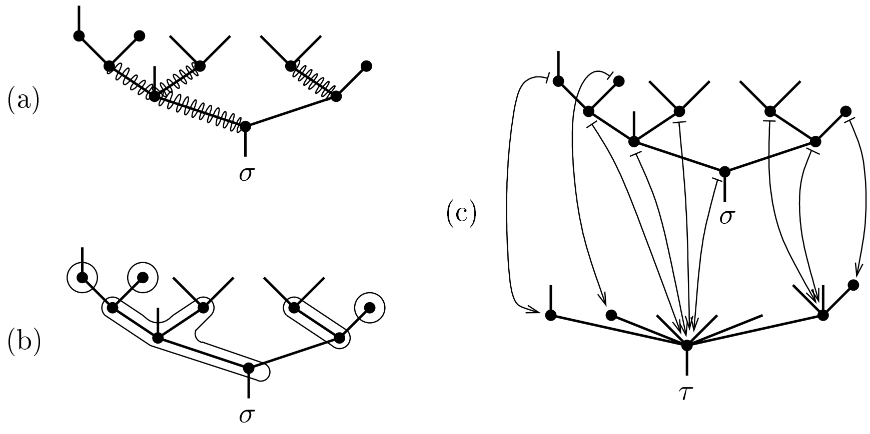


Figure 6: Three pictures of an epic in $\mathtt{Tree}_6$

inverse image of a vertex v of τ is the set of vertices of σ_v. In some texts a map of trees is described as something that 'contracts some internal edges'. (Here an *internal edge* is an edge that is not the root or a leaf; maps of trees keep the root and leaves fixed. To 'contract' an internal edge means to shrink it down to a vertex.) With one important caveat, this is what our maps of trees do: for in a map $\sigma \longrightarrow \tau$, the replacement of each partitioning subtree σ_v by the corolla with the same number of leaves amounts to the contraction of all the internal edges of σ_v. For example, Fig. 6(a) shows a tree σ with some of its edges marked for contraction, and Figs. 6(b) and 6(c) show the corresponding maps $\sigma \longrightarrow \tau$ in two different styles (as in Figs. 5(b) and (c)); so τ is the tree obtained by contracting the marked edges of σ.The caveat is that some of the σ_v's may be the trivial tree, and these are replaced by the 1-leafed corolla . This does *not* amount to the contraction of internal edges: it is, rather, the addition of a vertex to the middle of a (possibly external) edge. Any map of trees can be viewed as a combination of contractions of internal edges and additions of vertices to existing edges. For example, the map illustrated in Fig. 5 contracts two internal edges and adds a vertex to one edge.

Some further understanding of the category of trees can be gained by considering just those trees in which each vertex has at least two branches coming up out of it. We shall call these 'stable trees', following Kontsevich and Manin [9].

(a) (b)

Figure 7: (a) The category of 3-leafed stable trees, and (b) its classifying space

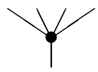
Formally, $\mathbf{StTree}_n$ is the full subcategory of $\mathtt{Tree}_n$ with objects defined by the recursive clauses

- $| \in \mathbf{StTree}_1$
- if $n \geq 2$, $k_1, \ldots, k_n \in \mathbb{N}$, and $T_1 \in \mathbf{StTree}_{k_1}, \ldots, T_n \in \mathbf{StTree}_{k_n}$ then $(T_1, \ldots, T_n) \in \mathbf{StTree}_{k_1+\cdots+k_n}$,

and an *n-leafed stable tree* is an object of $\mathbf{StTree}_n$. Since a stable tree can contain no subtree of the form , all maps between stable trees are 'surjections', that is, consist of just contractions of internal edges, without insertions of new vertices.

The first few categories $\mathbf{StTree}_n$ are trivial:

$$\begin{aligned} \mathbf{StTree}_0 &= \emptyset, \\ \mathbf{StTree}_1 &= \{ | \}, \\ \mathbf{StTree}_2 &= \{ \quad \} \end{aligned}$$



where in each case there are no arrows except for identities. The cases $n = 3$, 4, and 5 are illustrated in Figs. 7(a), 8(a), and 9(a). Identity arrows are not shown, and the categories $\mathbf{StTree}_n$ are ordered sets: all diagrams commute. Vertices are also omitted; since the trees are stable, this does not cause ambiguity. Parts (b) of the figures show the classifying spaces of these categories, solid polytopes of dimensions 1, 2 and 3. In the case of 5-leafed trees (Fig. 9) only about half of the category is shown, corresponding to the front faces of the polytope; the back faces and the terminal object of the category (the 5-leafed corolla), which sits at the centre of the polytope, are hidden. The whole polytope has 6 pentagonal faces, 3 square faces, and 3-fold rotational symmetry about the central vertical axis.

For $n \leq 5$, the classifying space $B(\mathbf{StTree}_n)$ is homeomorphic to the associahedron $\mathcal{A}_n$ (see [10] and Figure 4 above), and it seems very likely that this persists for all $n \in \mathbb{N}$. Indeed, the family of categories $(\mathbf{StTree}_n)_{n \in \mathbb{N}}$ forms a

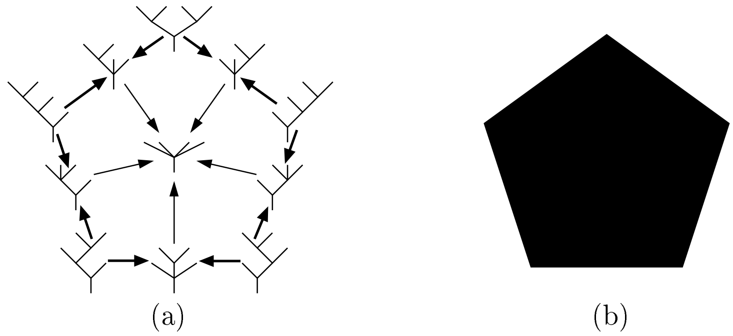


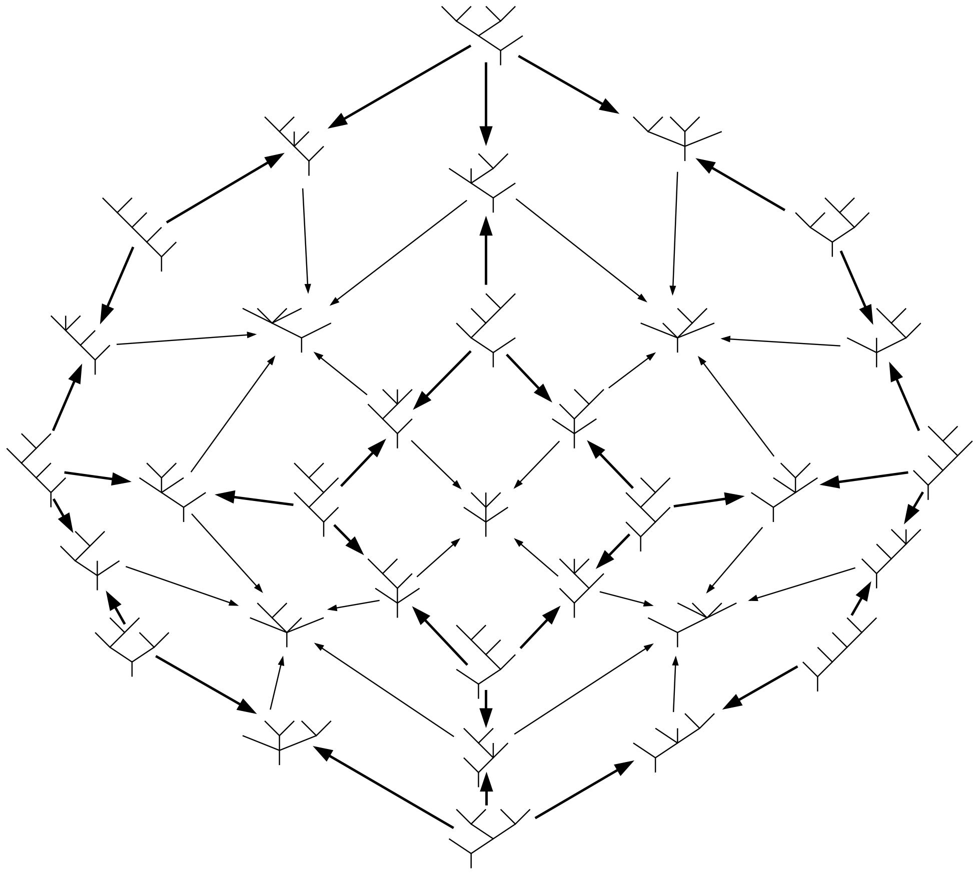
Figure 8: (a) The category of 4-leafed stable trees, and (b) its classifying space

sub-**Cat**-operad **STTR** of **Cat**-operad **TR**, and the classifying space functor $B : \mathbf{Cat} \longrightarrow \mathbf{Top}$ preserves finite products, so there is a (non-symmetric) topological operad $B(\mathbf{STTR})$ whose nth part is the classifying space of $\mathbf{StTree}_n$. (To make B preserve finite products we must interpret **Top** as the category of compactly generated or Kelley spaces: see [11] and [12].) This operad $B(\mathbf{STTR})$ is presumably isomorphic to Stasheff's operad $K = (K_n)_{n\in\mathbb{N}}$. A K-algebra is called an *A_∞-space*, and should be thought of as an up-to-homotopy version of a topological semigroup; the basic example is a loop space.

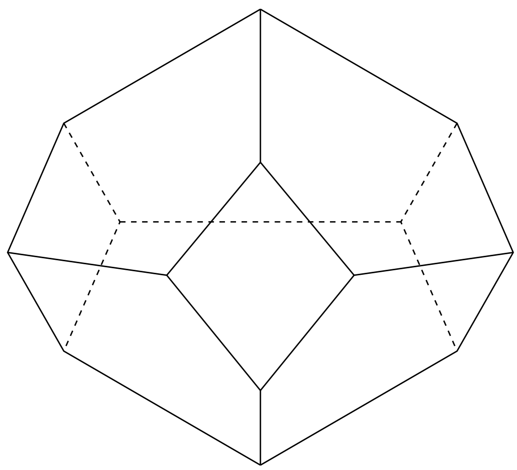
The categories $\mathbf{StTree}_n$ also give rise to the notion of an A_∞-algebra (see [10]). For each $n \in \mathbb{N}$, there is a chain complex $P(n)$ whose degree k part is the free abelian group on the set of n-leafed stable trees with $(n - k - 1)$ vertices.

When the signs are chosen appropriately this defines an operad P of chain complexes. A P-algebra is called an *A_∞-algebra*, to be thought of as an up-to-homotopy differential graded non-unital algebra; the usual example is the singular chain complex of an A_∞-space. A P-category is called an A_∞-category (see [13]), and consists of a collection of objects, a chain complex $\mathrm{Hom}(a, b)$ for each pair (a, b) of objects, maps defining binary composition, chain homotopies witnessing that this composition is associative up to homotopy, further homotopies witnessing that the previous homotopies obey the pentagon law up to homotopy, and so on.

Finally, since the polytopes $K_n = B(\mathbf{StTree}_n)$ describe higher associativity conditions, they also arise in definitions of higher-dimensional category. For example, the pentagon K_4 occurs in the classical definition of bicategory [13], and the polyhedron K_5 occurs as the 'non-abelian 4-cocycle condition' in Gordon, Power



(a)



(b)

Figure 9: (a) About half of the category of 5-leafed stable trees, and (b) the classifying space of the whole category

and Street's definition of tricategory [14].

We have already described operad of trees as the set $\mathbf{tr}_n$ of n-leafed trees. Maps $\sigma \longrightarrow \tau$ between trees are described by induction on the structure of τ:

- if $\tau = |$ then there is only one map into τ; it has domain $|$ and we write it as $1_| : | \longrightarrow |$
- if $\tau = (\tau_1, \ldots, \tau_n)$ for $\tau_1 \in \mathbf{tr}_{k_1}, \ldots, \tau_n \in \mathbf{tr}_{k_n}$ then a map $\sigma \longrightarrow \tau$ consists of trees $\rho \in \mathbf{tr}_n), \rho_1 \in \mathbf{tr}_{k_1}, \ldots, \rho_n \in \mathbf{tr}_{k_n}$ such that $\sigma = \rho\circ(\rho_1, \ldots, \rho_n)$, together with maps
$$\rho_1 \xrightarrow{\theta_1} \tau_1, \quad \ldots, \quad \rho_n \xrightarrow{\theta_n} \tau_n,$$
and we write this map as
$$\sigma = \rho\circ(\rho_1, \ldots, \rho_n) \overset{!_\rho * (\theta_1, \ldots, \theta_n)}{\longrightarrow} (\tau_1, \ldots, \tau_n) = \tau. \tag{21}$$

It follows easily that the n-leafed corolla $\nu_n = (\,|, \ldots, |\,)$ is the terminal object of $\mathtt{Tree}_n$: the unique map from $\sigma \in \mathbf{tr}_n$ to ν_n is written as $!_\sigma * (1_|, \ldots, 1_|)$. The rest of the structure of the **Cat**-operad **TR** can be described in a similarly explicit recursive fashion.

To make precise the intuition that a map of trees is a function of some sort, functors
$$V : \mathtt{Tree} \longrightarrow \mathtt{Set}_f, \qquad E : \mathtt{Tree}^{\mathrm{op}} \longrightarrow \mathtt{Set}_f$$
can be defined, encoding what happens on vertices and edges respectively. Both functors turn out to be faithful, which means that a map of trees is completely determined by its effect on either vertices or edges. The following account of V and E is just a sketch.

The more obvious of the two is the vertex functor V, defined on objects by

- $V(\,|\,) = \emptyset$
- $V((\tau_1, \ldots, \tau_n)) = 1 + V(\tau_1) + \cdots + V(\tau_n)$.

The edge functor E can be defined by first defining a functor $E_n : \mathtt{Tree}(n)^{\mathrm{op}} \longrightarrow (n+1)/\mathtt{Set}_f$ for each $n \in \mathbb{N}$, where $(n+1)/\mathtt{Set}_f$ is the category of sets equipped with $(n+1)$ ordered marked points. This definition is again by induction, the

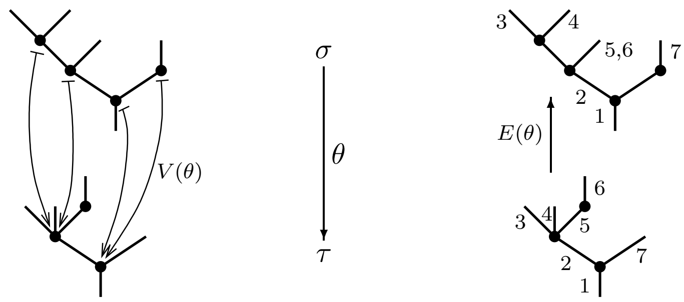


Figure 10: The effect on (a) vertices and (b) edges of a certain map of 4-leafed trees

idea being that E_n associates to a tree its edge-set with the n input edges and the one output edge (root) distinguished. Fig. 10 illustrates a map $\theta : \sigma \longrightarrow \tau$ in $\mathtt{Tree}(4)$; part (a) (= Fig. 5(c)) shows its effect $V(\theta)$ on vertices; part (b) shows $E(\theta)$, taking $E(\tau) = \{1, \ldots, 7\}$ and labelling the image of $i \in \{1, \ldots, 7\}$ under $E(\theta)$ by an i on the edge $(E(\theta))(i)$ of σ.

A map of trees will be called surjective if it is built up from contractions of internal edges. Formally, the *surjective* maps in `Tree` are defined by:

- $1_| : | \longrightarrow |$ is surjective
- with notation as in (21), $!_\rho * (\theta_1, \ldots, \theta_n)$ is surjective if and only if each θ_i is surjective and $\rho \neq |$.

The crucial part is the last: the unique map $!_\rho$ from $\rho \in \mathbf{tr}_n$ to the corolla ν_n is made up of edge-contractions just as long as ρ is not the unit tree $|$.

Dually, a map of trees is *injective* if, informally, it is built up from adding vertices to the middle of edges. Formally,

- $1_| : | \longrightarrow |$ is injective
- with notation as above, $!_\rho * (\theta_1, \ldots, \theta_n)$ is injective if and only if each θ_i is injective and ρ is either ν_n or $|$ (the latter only being possible if $n = 1$).

3 Unital Operads

Although operads and the most of related structures were defined in Sections 1 and 2 as an arbitrary symmetric monoidal category with countable coproducts, in Sections 3–11 we decided to follow the choice of [15] and formulate precise definitions only for the category $\mathtt{Mod}_{\mathbf{k}} = (\mathtt{Mod}_{\mathbf{k}}, \otimes)$ of modules over a commutative unital ring $\mathbf{k}$, with the monoidal structure given by the tensor product $\otimes = \otimes_{\mathbf{k}}$ over $\mathbf{k}$. The reason for such a decision was to give, in Section 6, a clean construction of free operads. In a general monoidal category, this construction involves the unordered $\odot$-product [16] so the free operad is then a double colimit, see [16]. Our choice also allows us to write formulas involving maps in terms of elements, which is sometimes a welcome simplification. We believe that the reader can easily reformulate next definitions and notations into other monoidal categories used in Sections 1–2 and 12–13 (see, also [16, 17]).

Let $\mathbf{k}[\Sigma_n]$ denote the $\mathbf{k}$-group ring of the symmetric group Σ_n.

Definition 3.1 (May's operad) *An* operad *in the category of* $\mathbf{k}$*-modules is a collection* $\mathcal{P} = \{\mathcal{P}(n)\}_{n \geq 0}$ *of right* $\mathbf{k}[\Sigma_n]$*-modules, together with* $\mathbf{k}$*-linear maps (operadic compositions)*

$$\gamma : \mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \to \mathcal{P}(k_1 + \cdots + k_n), \tag{22}$$

for $n \geq 1$ *and* $k_1, \ldots, k_n \geq 0$*, and a unit map* $\eta : \mathbf{k} \to \mathcal{P}(1)$*. These data fulfill the following axioms.*

Associativity. *Let* $n \geq 1$ *and let* $m_1, \ldots, m_n$ *and* $k_1, \ldots, k_m$*, where* $m := m_1 + \cdots + m_n$*, be non-negative integers. Then the following diagram, in which* $g_s := m_1 + \cdots + m_{s-1}$ *and* $h_s = k_{g_s+1} \cdots + k_{g_{s+1}}$*, for* $1 \leq s \leq n$*, commutes.*

$$\begin{tikzcd}
\left(\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \mathcal{P}(m_s)\right) \otimes \bigotimes_{r=1}^{m} \mathcal{P}(k_r) \arrow[r, "\gamma \otimes id"] \arrow[dd, "shuffle"'] & \mathcal{P}(m) \otimes \bigotimes_{r=1}^{m} \mathcal{P}(k_r) \arrow[d, "\gamma"] \\
 & \mathcal{P}(k_1 + \cdots + k_m) \\
\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \left(\mathcal{P}(m_s) \otimes \bigotimes_{q=1}^{m_s} \mathcal{P}(k_{g_s+q})\right) \arrow[r, "id \otimes (\bigotimes_{s=1}^{n} \gamma)"] & \mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \mathcal{P}(h_s) \arrow[u, "\gamma"']
\end{tikzcd}$$

Equivariance. *Let $n \geq 1$, let $k_1, \ldots, k_n$ be non-negative integers and $\sigma \in \Sigma_n$, $\tau_1 \in \Sigma_{k_1}, \ldots, \tau_n \in \Sigma_{k_n}$ permutations. Let $\sigma(k_1, \ldots, k_n) \in \Sigma_{k_1+\cdots+k_n}$ denote the permutation that permutes n blocks $(1, \ldots, k_1), \ldots, (k_{n-1}+1, \ldots, k_n)$ as σ permutes $(1, \ldots, n)$ and let $\tau_1 \oplus \cdots \oplus \tau_n \in \Sigma_{k_1+\cdots+k_n}$ be the block sum of permutations. Then the following diagrams commute.*

$$\begin{tikzcd}
\mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \arrow[r, "\sigma \otimes \sigma^{-1}"] \arrow[d, "\gamma"'] & \mathcal{P}(n) \otimes \mathcal{P}(k_{\sigma(1)}) \otimes \cdots \otimes \mathcal{P}(k_{\sigma(n)}) \arrow[d, "\gamma"] \\
\mathcal{P}(k_1 + \cdots + k_n) \arrow[r, "\sigma(k_{\sigma(1)}, \ldots, k_{\sigma(n)})"] & \mathcal{P}(k_{\sigma(1)} + \cdots + k_{\sigma(n)})
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \arrow[r, "id \otimes \tau_1 \otimes \cdots \otimes \tau_n"] \arrow[d, "\gamma"'] & \mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \arrow[d, "\gamma"] \\
\mathcal{P}(k_1 + \cdots + k_n) \arrow[r, "\tau_1 \oplus \cdots \oplus \tau_n"] & \mathcal{P}(k_1 + \cdots + k_n)
\end{tikzcd}$$

Unitality. *For each $n \geq 1$, the following diagrams commute.*

$$\begin{tikzcd}
\mathcal{P}(n) \otimes \mathbf{k}^{\otimes n} \arrow[r, "\cong"] \arrow[d, "id \otimes \eta^{\otimes n}"'] & \mathcal{P}(n) \\
\mathcal{P}(n) \otimes \mathcal{P}(1)^{\otimes n} \arrow[ur, "\gamma"']
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{k} \otimes \mathcal{P}(n) \arrow[r, "\cong"] \arrow[d, "\eta \otimes id"'] & \mathcal{P}(n) \\
\mathcal{P}(1) \otimes \mathcal{P}(n) \arrow[ur, "\gamma"']
\end{tikzcd}$$

A straightforward modification of the above definition makes sense in any symmetric monoidal category $(\mathcal{M}, \odot, \mathbf{1})$ such as the category of differential graded modules, simplicial sets, topological spaces, etc, see [16] or [17]. We then speak about *differential graded* operads, *simplicial* operads, *topological* operads, etc.

EXAMPLE 3.1 *All properties axiomatized by Definition 3.1 can be read from the* ***endomorphism operad*** $\mathcal{E}nd_V = \{\mathcal{E}nd_V(n)\}_{n\geq 0}$ *of a* $\mathbf{k}$*-module* V*. It is defined by setting* $\mathcal{E}nd_V(n)$ *to be the space of* $\mathbf{k}$*-linear maps* $V^{\otimes n} \to V$*. The operadic composition of* $f \in \mathcal{E}nd_V(n)$ *with* $g_1 \in \mathcal{E}nd_V(k_1), \ldots, g_n \in \mathcal{E}nd_V(k_n)$ *is given by the usual composition of multilinear maps as*

$$\gamma(f, g_1, \ldots, g_n) := f(g_1 \otimes \cdots \otimes g_n),$$

the symmetric group acts by

$$\gamma\sigma(f, g_1, \ldots, g_n) := f(g_{\sigma^{-1}(1)} \otimes \cdots \otimes g_{\sigma^{-1}(n)}),\ \sigma \in \Sigma_n,$$

and the unit map $\eta : \mathbf{k} \to \mathcal{E}nd_V(1)$ *is given by* $\eta(1) := id_V : V \to V$*. The endomorphism operad can be constructed over an object of an arbitrary symmetric monoidal category with an internal hom-functor, as it was done in [16].*

One often considers operads $\mathcal{A}$ such that $\mathcal{A}(0) = 0$ (the trivial $\mathbf{k}$-module). We will indicate that $\mathcal{A}$ is of this type by writing $\mathcal{A} = \{\mathcal{A}(n)\}_{n\geq 1}$.

EXAMPLE 3.2 *Let us denote by* $\mathcal{A}ss = \{\mathcal{A}ss(n)\}_{n\geq 1}$ *the operad with* $\mathcal{A}ss(n) := \mathbf{k}[\Sigma_n]$*,* $n \geq 1$*, and the operadic composition defined as follows. Let* $id_n \in \Sigma_n$*,* $id_{k_1} \in \Sigma_{k_1}, \ldots, id_{k_n} \in \Sigma_{k_n}$ *be the identity permutations. Then*

$$\gamma(id_n, id_{k_1}, \ldots, id_{k_n}) := id_{k_1+\cdots+k_n} \in \Sigma_{k_1+\cdots+k_n}.$$

The above formula determines $\gamma(\sigma, \tau_1, \ldots \tau_n)$ *for general* $\sigma \in \Sigma_n$*,* $\tau_1 \in \Sigma_{k_1}, \ldots, \tau_n \in \Sigma_{k_n}$ *by the equivariance axiom. The unit map* $\eta : \mathbf{k} \to \mathcal{A}ss(1)$ *is given by* $\eta(1) := id_1$*.*

EXAMPLE 3.3 *Let us give an example of a topological operad. For* $k \geq 1$*, the* ***little*** k***–discs operad*** $\mathcal{D}_k = \{\mathcal{D}_k(n)\}_{n\geq 0}$ *is defined as follows [16]. Let*

$$\mathbb{D}^k := \{(x_1, \ldots, x_k) \in \mathbb{R}^k;\ x_1^2 + \cdots + x_k^2 \leq 1\}$$

be the standard closed disc in $\mathbb{R}^k$. *A little* k*-disc is then a linear embedding* $d : \mathbb{D}^k \hookrightarrow \mathbb{D}^k$ *which is the restriction of a linear map* $\mathbb{R}^k \to \mathbb{R}^k$ *with parallel axes. The* n*-th space* $\mathcal{D}_k(n)$ *of the little* k*-disc operad is the space of all* n*-tuples* $(d_1, \ldots, d_n)$ *of little* k*-discs such that the images of* $d_1, \ldots, d_n$ *have mutually disjoint interiors. The operad structure is obvious – the symmetric group* Σ_n *acts on* $\mathcal{D}_k(n)$ *by permuting the labels of the little discs and the structure map* γ *is given by composition of embeddings. The unit is the identity embedding* $id : \mathbb{D}^k \hookrightarrow \mathbb{D}^k$.

EXAMPLE 3.4 *The collection of normalized singular chains* $C_*(\mathcal{T}) = \{C_*(\mathcal{T}(n))\}_{n\geq 0}$ *of a topological operad* $\mathcal{T} = \{\mathcal{T}(n)\}_{n\geq 0}$ *is an operad in the category of differential graded* $\mathbb{Z}$*-modules. For a ring* R*, the singular homology* $H_*(\mathcal{T}(n); R) = H_*(C_*(\mathcal{T}(n)) \otimes_{\mathbb{Z}} R)$ *forms an operad* $H_*(\mathcal{T}; R)$ *in the category of graded* R*-modules, see [15] for details.*

DEFINITION 3.2 *Let* $\mathcal{P} = \{\mathcal{P}(n)\}_{n\geq 0}$ *and* $\mathcal{Q} = \{\mathcal{Q}(n)\}_{n\geq 0}$ *be two operads. A* homomorphism $f : \mathcal{P} \to \mathcal{Q}$ *is a sequence* $f = \{f(n) : \mathcal{P}(n) \to \mathcal{Q}(n)\}_{n\geq 0}$ *of equivariant maps which commute with the operadic compositions and preserve the units.*

An operad $\mathcal{R} = \{\mathcal{R}(n)\}_{n\geq 0}$ *is a* ***suboperad*** *of* $\mathcal{P}$ *if* $\mathcal{R}(n)$ *is, for each* $n \geq 0$*, a* Σ_n*-submodule of* $\mathcal{P}(n)$ *and if all structure operations of* $\mathcal{R}$ *are the restrictions of those of* $\mathcal{P}$*. Finally, an* ideal *in the operad* $\mathcal{P}$ *is the collection* $\mathcal{I} = \{\mathcal{I}(n)\}_{n\geq 0}$ *of* Σ_n*-invariant subspaces* $\mathcal{I}(n) \subset \mathcal{P}(n)$ *such that*

$$\gamma_{\mathcal{P}}(f, g_1, \ldots, g_n) \in \mathcal{I}(k_1 + \cdots + k_n)$$

if either $f \in \mathcal{I}(n)$ *or* $g_i \in \mathcal{I}(k_i)$ *for some* $1 \leq i \leq n$.

EXAMPLE 3.5 *Given an operad* $\mathcal{P} = \{\mathcal{P}(n)\}_{n\geq 0}$*, let* $\widehat{\mathcal{P}} = \{\widehat{\mathcal{P}}(n)\}_{n\geq 0}$ *be the collection defined by* $\widehat{\mathcal{P}}(n) := \mathcal{P}(n)$ *for* $n \geq 1$ *and* $\widehat{\mathcal{P}}(0) := 0$*. Then* $\widehat{\mathcal{P}}$ *is a suboperad of* $\mathcal{P}$*. The correspondence* $\mathcal{P} \mapsto \widehat{\mathcal{P}}$ *is a full embedding of the category of operads* $\mathcal{P}$ *with* $\mathcal{P}(0) \cong \mathbf{k}$ *into the category of operads* $\mathcal{A}$ *with* $\mathcal{A}(0) = 0$*. Operads satisfying* $\mathcal{P}(0) \cong \mathbf{k}$ *have been called* ***unital*** *while operads with* $\mathcal{A}(0) = 0$ ***non − unital*** *operads. We will not use this terminology because non-unital operads will mean something different in this book, see Section 4.*

An example of an operad $\mathcal{A}$ *which is not of the form* $\widehat{\mathcal{P}}$ *for some operad* $\mathcal{P}$ *with* $\mathcal{P}(0) \cong \mathbf{k}$ *can be constructed as follows. Observe first that operads* $\mathcal{P}$ *with the property that*

$$\mathcal{P}(0) \cong \mathbf{k} \text{ and } \mathcal{P}(n) = 0 \text{ for } n \geq 2$$

are the same as augmented associative algebras. Indeed, the space $\mathcal{P}(1)$ with the operation $\circ_1 : \mathcal{P}(1)\otimes\mathcal{P}(1) \to \mathcal{P}(1)$ is clearly a unital associative algebra, augmented by the composition

$$\mathcal{P}(1) \xrightarrow{\cong} \mathcal{P}(1) \otimes \mathbf{k} \xrightarrow{\cong} \mathcal{P}(1) \otimes \mathcal{P}(0) \xrightarrow{\circ_1} \mathcal{P}(0) \cong \mathbf{k}.$$

Now take an arbitrary unital associative algebra A and define the operad $\mathcal{A} = \{\mathcal{A}(n)\}_{n\geq 1}$ by

$$\mathcal{A}(n) := \begin{cases} A, & \text{for} \quad n = 1 \quad \text{and} \\ 0, & \text{for} \quad n \neq 1, \end{cases}$$

with $\circ_1 : \mathcal{A}(1) \otimes \mathcal{A}(1) \to \mathcal{A}(1)$ the multiplication of A. It follows from the above considerations that $\mathcal{A} = \widehat{\mathcal{P}}$ for some operad $\mathcal{P}$ with $\mathcal{P}(0) \cong \mathbf{k}$ if and only if A admits an augmentation. Therefore any unital associative algebra that does not admit an augmentation produces the desired example.

EXAMPLE 3.6 *Kernels, images, etc., of homomorphisms between operads in the category of $\mathbf{k}$-modules are defined componentwise. For example, if $f : \mathcal{P} \to \mathcal{Q}$ is such homomorphism, then $Ker(f) = \{Ker(f)(n)\}_{n\geq 0}$ is the collection with*

$$Ker(f)(n) := Ker\left(f : \mathcal{P}(n) \to \mathcal{Q}(n)\right), \ n \geq 0.$$

It is clear that $Ker(f)$ is an ideal in $\mathcal{P}$.

Also quotients are defined componentwise. If $\mathcal{I}$ is an ideal in $\mathcal{P}$, then the collection $\mathcal{P}/\mathcal{I} = \{(\mathcal{P}/\mathcal{I})(n)\}_{n\geq 0}$ with $(\mathcal{P}/\mathcal{I})(n) := \mathcal{P}(n)/\mathcal{I}(n)$ for $n \geq 0$, has a natural operad structure induced by the structure of $\mathcal{P}$. The canonical projection $\mathcal{P} \to \mathcal{P}/\mathcal{I}$ has the expected universal property. The kernel of this projection equals $\mathcal{I}$.

Sometimes it suffices to consider operads without the symmetric group action. This notion is formalized by:

DEFINITION 3.3 (MAY'S NON-Σ OPERAD) *A* ***non*$-\Sigma$ *operad*** *in the category of $\mathbf{k}$-modules is a collection $\underline{\mathcal{P}} = \{\underline{\mathcal{P}}(n)\}_{n\geq 0}$ of $\mathbf{k}$-modules, together with operadic compositions*

$$\underline{\gamma} : \underline{\mathcal{P}}(n) \otimes \underline{\mathcal{P}}(k_1) \otimes \cdots \otimes \underline{\mathcal{P}}(k_n) \to \underline{\mathcal{P}}(k_1 + \cdots + k_n),$$

for $n \geq 1$ and $k_1, \ldots, k_n \geq 0$, and a unit map $\underline{\eta} : \mathbf{k} \to \underline{\mathcal{P}}(1)$ that fulfill the associativity and unitality axioms of Definition 3.1.

Each operad can be considered as a non-Σ operad by forgetting the Σ_n-actions. On the other hand, given a non-Σ operad $\underline{\mathcal{P}}$, there is an associated operad $\Sigma[\underline{\mathcal{P}}]$ with $\Sigma[\underline{\mathcal{P}}](n) := \underline{\mathcal{P}}(n) \otimes \mathbf{k}[\Sigma_n]$, $n \geq 0$, with the structure operations induced by the structure operations of $\underline{\mathcal{P}}$. Operads of this form are sometimes called *regular* operads.

EXAMPLE 3.7 *Consider the operad* $\mathcal{C}om = \{\mathcal{C}om(n)\}_{n\geq 1}$ *such that* $\mathcal{C}om(n)$ $:= \mathbf{k}$ *with the trivial* Σ_n*-action,* $n \geq 1$*, and the operadic compositions (22) given by the canonical identifications*

$$\mathcal{C}om(n) \otimes \mathcal{C}om(k_1) \otimes \cdots \otimes \mathcal{C}om(k_n) \cong \mathbf{k}^{\otimes(n+1)} \xrightarrow{\cong} \mathbf{k} \cong \mathcal{C}om(k_1 + \cdots + k_n).$$

The operad $\mathcal{C}om$ *is obviously not regular. Observe also that* $\mathcal{C}om \cong \widehat{\mathcal{E}nd}_{\mathbf{k}}$*, where* $\widehat{\mathcal{E}nd}_{\mathbf{k}}$ *is the endomorphism operad of the ground ring without the initial component, see Example 3.5 for the notation.*

Let $\underline{\mathcal{A}ss}$ *denote the operad* $\mathcal{C}om$ *considered as a non-*Σ *operad. Its symmetrization* $\Sigma[\underline{\mathcal{A}ss}]$ *then equals the operad* $\mathcal{A}ss$ *introduced in Example 3.2.*

As we already observed in Sections 1 and 2, there is an alternative approach to operads. For the purposes of comparison, in the rest of this Section and in the following Section we will refer to operads viewed from this alternative perspective as to Markl's operads.

DEFINITION 3.4 *A* ***Markl's operad*** *in the category of* $\mathbf{k}$*-modules is a collection* $\mathcal{S} = \{\mathcal{S}(n)\}_{n\geq 0}$ *of right* $\mathbf{k}[\Sigma_n]$*-modules, together with* $\mathbf{k}$*-linear maps (*$\circ_i$*-compositions)*

$$\circ_i : \mathcal{S}(m) \otimes \mathcal{S}(n) \to \mathcal{S}(m+n-1),$$

for $1 \leq i \leq m$ *and* $n \geq 0$*. These data fulfill the following axioms.*
Associativity. *For each* $1 \leq j \leq a$*,* $b, c \geq 0$*,* $f \in \mathcal{S}(a)$*,* $g \in \mathcal{S}(b)$ *and* $h \in \mathcal{S}(c)$*,*

$$(f \circ_j g) \circ_i h = \begin{cases} (f \circ_i h) \circ_{j+c-1} g, & \text{for } 1 \leq i < j, \\ f \circ_j (g \circ_{i-j+1} h), & \text{for } j \leq i < b+j, \text{ and} \\ (f \circ_{i-b+1} h) \circ_j g, & \text{for } j+b \leq i \leq a+b-1, \end{cases}$$

see Figure 11.

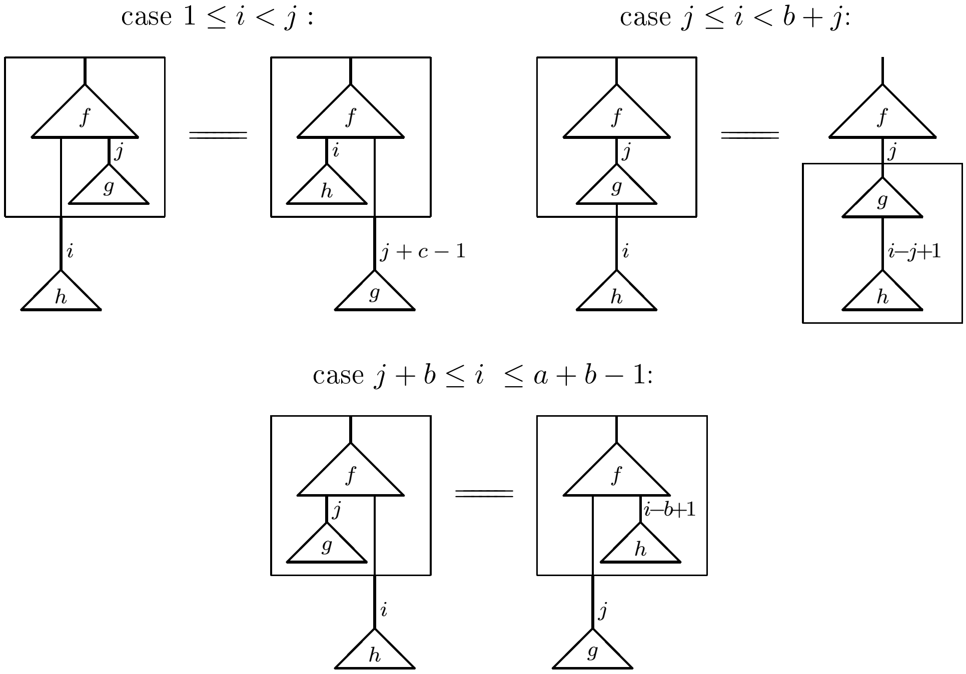


Figure 11: Flow charts explaining the associativity in Markl's operads.

Equivariance. *For each* $1 \le i \le m$, $n \ge 0$, $\tau \in \Sigma_m$ *and* $\sigma \in \Sigma_n$, *let* $\tau \circ_i \sigma \in \Sigma_{m+n-1}$ *be given by inserting the permutation* σ *at the ith place in* τ. *Let* $f \in \mathcal{S}(m)$ *and* $g \in \mathcal{S}(n)$. *Then*

$$(f\tau) \circ_i (g\sigma) = (f \circ_{\tau(i)} g)(\tau \circ_i \sigma).$$

Unitality. *There exists* $e \in \mathcal{S}(1)$ *such that*

$$f \circ_i e = e \quad \textit{and} \quad e \circ_1 g = g \tag{23}$$

for each $1 \le i \le m$, $n \ge 0$, $f \in \mathcal{S}(m)$ *and* $g \in \mathcal{S}(n)$.

Example 3.8 *All axioms in Definition 3.4 can be read from the endomorphism operad* $\mathcal{E}nd_V = \{\mathcal{E}nd_V(n)\}_{n \ge 0}$ *of a* $\mathbf{k}$*-module* V *reviewed in Example 3.1, with* $\circ_i$*-operations given by*

$$f \circ_i g := f(id_V^{\otimes i-1} \otimes g \otimes id_V^{\otimes m-1}),$$

for $f \in \mathcal{E}nd_V(m)$, $g \in \mathcal{E}nd_V(n)$, $1 \le i \le m$ *and* $n \ge 0$.

The following proposition shows that Definition 3.1 describes the same objects as Definition 3.4.

PROPOSITION 3.1 *The category of May's operads is isomorphic to the category of Markl's operads.*

Proof. Given a Markl's operad $\mathcal{S} = \{\mathcal{S}(n)\}_{n\geq 0}$ as in Definition 3.4, define a May's operad $\mathcal{P} = \mathrm{May}(\mathcal{S})$ by $\mathcal{P}(n) := \mathcal{S}(n)$ for $n \geq 0$, with the γ-operations given by

$$\gamma(f, g_1, \ldots, g_n) := (\cdots((f \circ_n g_n) \circ_{n-1} g_{n-1}) \cdots) \circ_1 g_1 \tag{24}$$

where $f \in \mathcal{P}(n)$, $g_i \in \mathcal{P}(k_i)$, $1 \leq i \leq n$, $k_1, \ldots, k_n \geq 0$. The unit morphism $\eta : \mathbf{k} \to \mathcal{P}(1)$ is defined by $\eta(1) := e$. It is easy to verify that $\mathrm{May}(-)$ extends to a functor from the category of Markl's operads the category of May's operads.

On the other hand, given a May's operad $\mathcal{P}$, one can define a Markl's operad $\mathcal{S} = \mathrm{Mar}(\mathcal{P})$ by $\mathcal{S}(n) := \mathcal{P}(n)$ for $n \geq 0$, with the $\circ_i$-operations:

$$f \circ_i g := \gamma(f, \underbrace{e, \ldots, e}_{i-1}, g, \underbrace{e, \ldots, e}_{m-i}), \tag{25}$$

for $f \in \mathcal{S}(m)$, $g \in \mathcal{S}(n)$, $m \geq 1$, $n \geq 0$, where $e := \eta(1) \in \mathcal{P}(1)$. It is again obvious that $\mathrm{Mar}(-)$ extends to a functor that the functors $\mathrm{May}(-)$ and $\mathrm{Mar}(-)$ are mutually inverse isomorphisms between the category of Markl's operads and the category of May's operads. ■

The equivalence between May's and Markl's operads implies that an operad can be defined by specifying $\circ_i$-operations and a unit. This is sometimes simpler that to define the γ-operations directly, as illustrated by:

EXAMPLE 3.9 *Let Σ be a Riemann sphere, that is, a nonsingular complex projective curve of genus 0. By a puncture or a parametrized hole we mean a point p of Σ together with a holomorphic embedding of the standard closed disc $U = \{z \in \mathbb{C};\ |z| \leq 1\}$ to Σ centered at the point. Thus a puncture is a holomorphic embedding $u : \tilde{U} \to \Sigma$, where $\tilde{U} \subset \mathbb{C}$ is an open neighborhood of U and $u(0) = p$. We say that two punctures $u_1 : \tilde{U}_1 \to \Sigma$ and $u_2 : \tilde{U}_2 \to \Sigma$ are disjoint, if*

$$u_1(\overset{\circ}{U}) \cap u_2(\overset{\circ}{U}) = \emptyset,$$

where $\overset{\circ}{U} := \{z \in \mathbb{C}\,;\ |z| < 1\}$ *is the interior of* U.

Let $\widehat{\mathfrak{M}}_0(n)$ *be the moduli space of Riemann spheres* Σ *with* $n+1$ *disjoint punctures* $u_i : \tilde{U}_i \to \Sigma$, $0 \leq i \leq n$, *modulo the action of complex projective automorphisms. The topology of* $\widehat{\mathfrak{M}}_0(n)$ *is a very subtle thing and we are not going to discuss this issue here; see [18]. The constructions below will be made only 'up to topology.'*

Renumbering the holes $u_1, \ldots, u_n$ *defines on each* $\widehat{\mathfrak{M}}_0(n)$ *a natural right* Σ_n-*action and the* Σ-*module* $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n\geq 0}$ *forms a topological operad under sewing Riemannian spheres at punctures. Let us describe this operadic structure using the* $\circ_i$-*formalism. Thus, let* Σ *represent an element* $x \in \widehat{\mathfrak{M}}_0(m)$ *and* Δ *represent an element* $y \in \widehat{\mathfrak{M}}_0(n)$. *For* $1 \leq i \leq m$, *let* $u_i : \tilde{U}_i \to \Sigma$ *be the* i*th puncture of* Σ *and let* $u_0 : \tilde{U}_0 \to \Delta$ *be the* 0*th puncture of* Δ.

There certainly exists some $0 < r < 1$ *such that both* $\tilde{U}_0$ *and* $\tilde{U}_i$ *contain the disc* $U_{1/r} := \{z \in \mathbb{C}\,;\ |z| < 1/r\}$. *Let now* $\Sigma_r := \Sigma \setminus u_i(U_r)$ *and* $\Delta_r := \Delta \setminus u_0(U_r)$. *Define finally*

$$\Xi := (\Sigma_r \bigsqcup \Delta_r)/\sim,$$

where the relation $\sim$ *is given by*

$$\Sigma_r \ni u_i(\xi) \sim u_0(1/\xi) \in \Delta_r,$$

for $r < |\xi| < 1/r$. *It is immediate to see that* Ξ *is a well-defined punctured Riemannian sphere, with* $n+m-1$ *punctures induced in the obvious manner from those of* Σ *and* Δ, *and that the class of the punctured surface* Ξ *in the moduli space* $\widehat{\mathfrak{M}}_0(m+n-1)$ *does not depend on the representatives* Σ, Δ *and on* r. *We define* $x \circ_i y$ *to be the class of* Ξ.

The unit $e \in \widehat{\mathfrak{M}}_0(1)$ *can be defined as follows. Let* $\mathbb{CP}^1$ *be the complex projective line with homogeneous coordinates* $[z,w]$, $z, w \in \mathbb{C}$, *[19, Example I.1.6]. Let* $0 := [0,1] \in \mathbb{CP}^1$ *and* $\infty := [1,0] \in \mathbb{CP}^1$. *Recall that we have two canonical isomorphisms* $p_\infty : \mathbb{CP}^1 \setminus \infty \to \mathbb{C}$ *and* $p_0 : \mathbb{CP}^1 \setminus 0 \to \mathbb{C}$ *given by*

$$p_\infty([z,w]) := z/w \text{ and } p_0([z,w]) := w/z.$$

Then $p_\infty^{-1} : \mathbb{C} \to \mathbb{CP}^1$ *(respectively* $p_0^{-1} : \mathbb{C} \to \mathbb{CP}^1$*) is a puncture at* 0 *(respectively at* ∞*). We define* $e \in \widehat{\mathfrak{M}}_0(1)$ *to be the class of* $(\mathbb{CP}^1, p_0^{-1}, p_\infty^{-1})$.

It is not hard to verify that the above constructions make the collection $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n\geq 0}$ *a Markl's operad. By Proposition 3.1,* $\widehat{\mathfrak{M}}_0$ *is a also May's operad.*

In the rest of this book, we will consider May's and Markl's operads as two versions of the same object which we will call simply a *unital operad.*

4 Non-unital Operads

It turns out that the combinatorial structure of the moduli space of stable genus zero curves is captured by a certain non-unital version of operad. Let $\mathcal{M}_{0,n+1}$ be the moduli space of $(n+1)$-tuples $(x_0, \ldots, x_n)$ of distinct numbered points on the complex projective line $\mathbb{CP}^1$ modulo projective automorphisms, that is, transformations of the form

$$\mathbb{CP}^1 \ni [\xi_1, \xi_2] \mapsto [a\xi_1 + b\xi_2, c\xi_1 + d\xi_2] \in \mathbb{CP}^1,$$

where $a, b, c, d \in \mathbb{C}$ with $ad - bc \neq 0$.

The moduli space $\mathcal{M}_{0,n+1}$ has, for $n \geq 2$, a canonical compactification $\overline{\mathcal{M}}_0(n) \supset \mathcal{M}_{0,n+1}$ introduced by A. Grothendieck and F.F. Knudsen [20, 21]. The space $\overline{\mathcal{M}}_0(n)$ is the moduli space of stable $(n+1)$-pointed curves of genus 0:

Definition 4.1 *A* ***stable*** $(n+1)$-***pointed curve of genus*** 0 *is an object*

$$(C; x_0, \ldots, x_n),$$

where C is a (possibly reducible) algebraic curve with at most nodal singularities and $x_0, \ldots, x_n \in C$ are distinct smooth points such that

(i) each component of C is isomorphic to $\mathbb{CP}^1$,

(ii) the graph of intersections of components of C (i.e. the graph whose vertices correspond to the components of C and edges to the intersection points of the components) is a tree, and

(iii) each component of C has at least three special points, where a special point means either one of the x_i, $0 \leq i \leq n$, or a singular point of C (the stability).

It can be seen that a stable curve $(C; x_0, \ldots, x_n)$ admits no infinitesimal automorphisms that fix marked points $x_0, \ldots, x_n$, therefore $(C; x_0, \ldots, x_n)$ is 'stable' in the usual sense. Observe also that $\overline{\mathcal{M}}_0(0) = \overline{\mathcal{M}}_0(1) = \emptyset$ (there are no stable curves with less than three marked points) and that $\overline{\mathcal{M}}_0(2)$ = the point corresponding to the three-pointed stable curve $(\mathbb{CP}^1; \infty, 1, 0)$. The space $\mathcal{M}_{0,n+1}$ forms an open dense part of $\overline{\mathcal{M}}_0(n)$ consisting of marked curves $(C; x_0, \ldots, x_n)$ such that C is isomorphic to $\mathbb{CP}^1$.

Let us try to equip the collection $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n \geq 2}$ with an operad structure as in Definition 3.1. For $C = (C, x_1, \ldots, x_n) \in \overline{\mathcal{M}}_0(n)$ and $C_i = (C_i, y_1^i, \ldots, y_{k_i}^i) \in \overline{\mathcal{M}}_0(k_i)$, $1 \leq i \leq n$, let

$$\gamma(C, C_1, \ldots, C_n) \in \overline{\mathcal{M}}_0(k_1 + \cdots + k_n) \tag{26}$$

be the stable marked curve obtained from the disjoint union $C \sqcup C^1 \sqcup \cdots \sqcup C^n$ by identifying, for each $1 \leq i \leq n$, the point $x_i \in C$ with the point $y_0^i \in C_i$, introducing a nodal singularity, and relabeling the remaining marked points accordingly. The symmetric group acts on $\overline{\mathcal{M}}_0(n)$ by

$$(C, x_0, x_1, \ldots, x_n) \longmapsto (C, x_0, x_{\sigma(1)}, \ldots, x_{\sigma(n)}),\ \sigma \in \Sigma_n.$$

We have defined the γ-compositions and the symmetric group action, but there is no room for the identity, because $\overline{\mathcal{M}}_0(1)$ is empty! The above structure is, therefore, a non-unital operad in the sense of the following definition (which is formulated, as all definitions in Sections 1.3–1.11, for the monoidal category of $\mathbf{k}$-modules).

DEFINITION 4.2 *A* ***May's non − unital operad*** *in the category of* $\mathbf{k}$*-modules is a collection* $\mathcal{P} = \{\mathcal{P}(n)\}_{n \geq 0}$ *of* $\mathbf{k}[\Sigma_n]$*-modules, together with operadic compositions*

$$\gamma : \mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \to \mathcal{P}(k_1 + \cdots + k_n),$$

for $n \geq 1$ *and* $k_1, \ldots, k_n \geq 0$*, that fulfill the associativity and equivariance axioms of Definition 3.1.*

We may as well define on the collection $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n \geq 2}$ operations

$$\circ_i : \overline{\mathcal{M}}_0(m) \times \overline{\mathcal{M}}_0(n) \to \overline{\mathcal{M}}_0(m + n - 1) \tag{27}$$

for $m, n \geq 2$, $1 \leq i \leq m$, by

$$(C^1; y_0, \ldots, y_m) \times (C^2; x_0, \ldots, x_n) \longmapsto (C; y_0, \ldots, y_{i-1}, x_0, \ldots, x_n, y_{i+1}, \ldots, y_m)$$

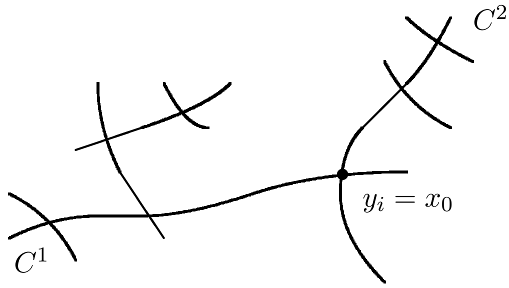


Figure 12: The $\circ_i$-compositions in $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geq 2}$.

where C is the quotient of the disjoint union $C^1 \bigsqcup C^2$ given by identifying x_0 with y_i at a nodal singularity, see Figure 12. The collection $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geq 2}$ with $\circ_i$-operations (27) is an example of another version of non-unital operads, recalled in:

Definition 4.3 *A* ***non − unital Markl's operad*** *in the category of* $\mathbf{k}$*-modules is a collection* $\mathcal{P} = \{\mathcal{P}(n)\}_{n\geq 0}$ *of* $\mathbf{k}[\Sigma_n]$*-modules, together with operadic compositions*

$$\circ_i : \mathcal{S}(m) \otimes \mathcal{S}(n) \to \mathcal{S}(m+n-1),$$

for $1 \leq i \leq m$ *and* $n \geq 0$*, that fulfill the associativity and equivariance axioms of Definition 3.4.*

As we saw in Proposition 3.1, in the presence of operadic units, May's operads are the same as Markl's operads. Surprisingly, the non-unital versions of these structures are *radically different* – Markl's operads capture more information than May's operads! This is made precise in the following:

Proposition 4.1 *The category of non-unital Markl's operads is a subcategory of the category of non-unital May's operads.*

Proof. It is easy to see that (24) defines, as in the proof of Proposition 3.1, a functor ψMay$(-)$ which is an embedding of the category of non-unital Markl's operads into the category of non-unital May's operads. ■

Observe that formula (25), inverse to (24), does not make sense without units. The relation between various versions of operads discussed so far is summarized in the following diagram of categories and their inclusions:

$$\begin{tikzcd}
\text{May's operads} \arrow[d, hook] \arrow[rr, bend left, "\mathrm{Mar}"] & & \text{Markl's operads} \arrow[ll, "\mathrm{May}"'] \arrow[d, hook] \\
\text{non-unital May's operads} & & \text{non-unital Markl's operads} \arrow[ll, hook', "\psi\mathrm{May}"']
\end{tikzcd}$$

The following example shows that non-unital Markl's operads form a proper sub-category of the category of non-unital May's operads.

EXAMPLE 4.1 *We describe a non-unital May's operad $\mathcal{V} = \{\mathcal{V}(n)\}_{n\geq 0}$ which is not of the form $\psi\mathrm{May}(\mathcal{S})$ for some non-unital Markl's operad $\mathcal{S}$. Let*

$$\mathcal{V}(n) := \begin{cases} \mathbf{k}, & \text{for} \quad n = 2 \quad \text{or} \quad 4, \quad \text{and} \\ 0, & \text{otherwise.} \end{cases}$$

The only non-trivial γ-composition is $\gamma : \mathcal{V}(2) \otimes \mathcal{V}(2) \otimes \mathcal{V}(2) \to \mathcal{V}(4)$, given as the canonical isomorphism

$$\mathcal{V}(2) \otimes \mathcal{V}(2) \otimes \mathcal{V}(2) \cong \mathbf{k}^{\otimes 3} \xrightarrow{\cong} \mathbf{k} \cong \mathcal{V}(4).$$

Suppose that $\mathcal{V} = \mathrm{May}(\mathcal{S})$ for some non-unital Markl's operad $\mathcal{S}$. Then, according to (24), for $f, g_1, g_2 \in \mathcal{V}(2)$,

$$\gamma(f, g_1, g_2) = (f \circ_2 g_2) \circ_1 g_1.$$

Since $(f \circ_2 g) \in \mathcal{V}(3) = 0$, this would imply that γ is trivial, which is not true.

Proposition 4.2 below shows that Markl's rather than May's non-unital operads are true non-unital versions of operads. We will need the following definition in which $\mathcal{K} = \{\mathcal{K}(n)\}_{n\geq 1}$ is the trivial (unital) operad with $\mathcal{K}(1) := \mathbf{k}$ and $\mathcal{K}(n) = 0$, for $n \neq 1$.

DEFINITION 4.4 *An **augmentation** of an operad $\mathcal{P}$ in the category of* $\mathbf{k}$*-modules is a homomorphism $\epsilon : \mathcal{P} \to \mathcal{K}$. Operads with an augmentation are called **augmented** operads. The kernel*

$$\overline{\mathcal{P}} := Ker\,(\epsilon : \mathcal{P} \to \mathcal{K})$$

*is called the **augmentation ideal**.*

The following proposition was proved in [22].

PROPOSITION 4.2 *The correspondence $\mathcal{P} \mapsto \overline{\mathcal{P}}$ is an isomorphism between the category of augmented operads and the category of Markl's non-unital operads.*

Proof. The $\circ_i$-operations of $\mathcal{P}$ obviously restrict to $\overline{\mathcal{P}}$, making it a non-unital Markl's operad. It is simple to describe a functorial inverse $\mathcal{S} \mapsto \widetilde{\mathcal{S}}$ of the correspondence $\mathcal{P} \mapsto \overline{\mathcal{P}}$. For a Markl's non-unital operad $\mathcal{S}$, denote by $\widetilde{\mathcal{S}}$ the collection

$$\widetilde{\mathcal{S}}(n) := \begin{cases} \mathcal{S}(n), & \text{for} \quad n \neq 1, \quad \text{and} \\ \mathcal{S}(1) \oplus \mathbf{k} & \text{for} \quad n = 1. \end{cases} \tag{28}$$

The $\circ_i$-operations of $\widetilde{\mathcal{S}}$ are uniquely determined by requiring that they extend the $\circ_i$-operations of $\mathcal{S}$ and satisfy (23), with the unit $e := 0 \oplus 1_{\mathbf{k}} \in \mathcal{S}(1) \oplus \mathbf{k} = \widetilde{\mathcal{S}}(1)$. Informally, $\widetilde{\mathcal{S}}$ is obtained from the Markl's non-unital operad $\mathcal{S}$ by adjoining a unit. ■

Observe that if $\mathcal{S}$ were a May's, not Markl's, non-unital operad, the construction of $\widetilde{\mathcal{S}}$ described in the above proof would not make sense, because we would not know how to define

$$\gamma(f, \underbrace{e, \ldots, e}_{i-1}, g, \underbrace{e, \ldots, e}_{m-i})$$

for $f \in \mathcal{S}(m)$, $g \in \mathcal{S}(n)$, $m \geq 2$, $n \geq 0$, $1 \leq i \leq m$. Proposition 4.2 should be compared to the obvious statement that the category of augmented unital associative algebras is isomorphic to the category of (non-unital) associative algebras. In the following proposition, `Oper` denotes the category of $\mathbf{k}$-linear operads and ψ`Oper` the category of $\mathbf{k}$-linear Markl's non-unital operads.

PROPOSITION 4.3 *Let $\mathcal{P}$ be an augmented operad and $\mathcal{Q}$ an arbitrary operad in the category of* $\mathbf{k}$*-modules. Then there exists a natural isomorphism*

$$Mor_{\mathtt{Oper}}(\mathcal{P},\mathcal{Q}) \cong Mor_{\psi\mathtt{Oper}}(\overline{\mathcal{P}}, \psi\mathrm{May}(\mathcal{Q})). \tag{29}$$

The proof is simple and we leave it the reader. Combining (29) with the isomorphism of Proposition 4.2 one obtains a natural isomorphism

$$Mor_{\mathtt{Oper}}(\widetilde{\mathcal{S}},\mathcal{Q}) \cong Mor_{\psi\mathtt{Oper}}(\mathcal{S}, \psi\mathrm{May}(\mathcal{Q})) \tag{30}$$

which holds for each Markl's non-unital operad $\mathcal{S}$ and operad $\mathcal{Q}$. Isomorphism (30) means that $\widetilde{\ }: \psi\mathtt{Oper} \to \mathtt{Oper}$ and $\psi\mathrm{May} : \mathtt{Oper} \to \psi\mathtt{Oper}$ are adjoint functors. This adjunction will be used in the construction of free operads in Section 6.

In the rest of this book, non-unital Markl's operads will be called simply non-unital operads. This will not lead to confusion, since all non-unital operads referred to in the rest of this book will be Markl's.

5 Operad Algebras

As we already remarked, operads are important through their representations called operad algebras or simply algebras.

DEFINITION 5.1 *Let V be a* $\mathbf{k}$*-module and $\mathcal{E}nd_V$ the endomorphism operad of V recalled in Example 3.1. A* $\mathcal{P}$-algebra *is a homomorphism of operads $\rho : \mathcal{P} \to \mathcal{E}nd_V$.*

The above definition admits an obvious generalization into an arbitrary symmetric monoidal category with an internal hom-functor. The last assumption is necessary for the existence of the 'internal' endomorphism operad, see [16]. Definition 5.1 can be however unwrapped into the form given in [15] that makes sense in an arbitrary symmetric monoidal category without the internal hom-functor assumption:

PROPOSITION 5.1 *Let $\mathcal{P}$ be an operad. A $\mathcal{P}$-algebra is the same as a* $\mathbf{k}$*-module V together with maps*

$$\alpha : \mathcal{P}(n) \otimes V^{\otimes n} \to V,\ \ n \geq 0, \tag{31}$$

that satisfy the following axioms.

Associativity. *For each $n \geq 1$ and non-negative integers $k_1, \ldots, k_n$, the following diagram commutes.*

$$\begin{tikzcd}
\left(\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \mathcal{P}(k_s)\right) \otimes \bigotimes_{s=1}^{n} V^{\otimes k_s} \arrow[r, "\gamma \otimes id"] \arrow[dd, "shuffle"'] & \mathcal{P}(k_1 + \cdots + k_n) \otimes V^{\otimes(k_1+\cdots+k_n)} \arrow[d, "\alpha"] \\
& V \\
\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \left(\mathcal{P}(k_s) \otimes V^{\otimes k_s}\right) \arrow[r, "id \otimes (\bigotimes_{s=1}^{n} \alpha)"] & \mathcal{P}(n) \otimes V^{\otimes n} \arrow[u, "\alpha"']
\end{tikzcd}$$

Equivariance. *For each $n \geq 1$ and $\sigma \in \Sigma_n$, the following diagram commutes.*

$$\begin{tikzcd}
\mathcal{P}(n) \otimes V^{\otimes n} \arrow[rr, "\sigma \otimes \sigma^{-1}"] \arrow[dr, "\alpha"'] & & \mathcal{P}(n) \otimes V^{\otimes n} \arrow[dl, "\alpha"] \\
& V &
\end{tikzcd}$$

Unitality. *For each $n \geq 1$, the following diagram commutes.*

$$\begin{tikzcd}
\mathbf{k} \otimes V \arrow[r, "\cong"] \arrow[d, "\eta \otimes id"'] & V \\
\mathcal{P}(1) \otimes V \arrow[ur, "\alpha"] &
\end{tikzcd}$$

We leave as an exercise to formulate a version of Proposition 5.1 that would use $\circ_i$-operations instead of γ-operations.

EXAMPLE 5.1 *In this example we verify, using Proposition 5.1, that algebras over the operad $\mathcal{C}om = \{\mathcal{C}om(n)\}_{n \geq 1}$ recalled in Example 3.7 are ordinary commutative associative algebras. To simplify the exposition, let us agree that v's with various subscripts denote elements of V. Since $\mathcal{C}om(n) = \mathbf{k}$ for $n \geq 1$, the structure map (31) determines, for each $n \geq 1$, a linear map $\mu_n : V^{\otimes n} \to V$ by*

$$\mu_n(v_1, \ldots, v_n) := \alpha(1_n, v_1, \ldots, v_n),$$

where 1_n denotes in this example the unit $1_n \in \mathbf{k} = \mathcal{C}om(n)$. The associativity of Proposition 5.1 says that

$$\mu_n\left(\mu_{k_1}(v_1,\ldots,v_{k_1}),\ldots,\mu_{k_n}(v_{k_1+\cdots+k_{n-1}+1},\ldots,v_{k_1+\cdots+k_n})\right) = \mu_{k_1+\cdots+k_n}(v_1,\ldots,v_{k_1+\cdots+k_n}), \tag{32}$$

for each $n, k_1, \ldots, k_n \geq 1$. The equivariance of Proposition 5.1 means that each μ_n is fully symmetric

$$\mu_n(v_1,\ldots,v_n) = \mu_n(v_{\sigma(1)},\ldots,v_{\sigma(n)}),\ \sigma \in \Sigma_n, \tag{33}$$

and the unitality implies that μ_1 is the identity map,

$$\mu_1(v) = v. \tag{34}$$

The above structure can be identified with a commutative associative multiplication on V. Indeed, the bilinear map $\cdot := \mu_2 : V \otimes V \to V$ is clearly associative:

$$(v_1 \cdot v_2) \cdot v_3 = v_1 \cdot (v_2 \cdot v_3) \tag{35}$$

and commutative:

$$v_1 \cdot v_2 = v_2 \cdot v_1. \tag{36}$$

On the other hand, $\mu_1(v) := v$ and

$$\mu_n(v_1,\ldots,v_n) := (\cdots(v_1 \cdot v_2)\ \cdots\ v_{n-1}) \cdot v_n \quad \text{for } n \geq 2$$

defines multilinear maps $\{\mu_n : V^{\otimes n} \to V\}$ satisfying (32)–(34). It is equally easy to verify that algebras over the operad $\mathcal{A}ss$ introduced in Example 3.2 are ordinary associative algebras.

Following Leinster [13], one could say that (32)–(34) is an *unbiased* definition of associative commutative algebras, while (35)–(36) is a definition of the same object *biased* towards bilinear operations. Operads therefore provide unbiased definitions of algebras.

EXAMPLE 5.2 *Let us denote by $U\mathcal{C}om$ the endomorphism operad $\mathcal{E}nd_{\mathbf{k}}$ of the ground ring $\mathbf{k}$. It is easy to verify that $U\mathcal{C}om$-algebras are* **unital** *commutative associative algebras. We leave it to the reader to describe the operad $U\mathcal{A}ss$ governing unital associative operads.*

Algebras over a non-Σ operad $\underline{\mathcal{P}}$ are defined as algebras, in the sense of Definition 5.1, over the symmetrization $\Sigma[\underline{\mathcal{P}}]$ of $\underline{\mathcal{P}}$. Algebras over non-unital operads discussed in Section 4 are defined by appropriate obvious modifications of Definition 5.1.

EXAMPLE 5.3 *Let Y be a topological space with a base point $*$ and $\mathbb{S}^k$ the k-dimensional sphere, $k \geq 1$. The k-fold loop space $\Omega^k Y$ is the space of all continuous maps $\mathbb{S}^k \to Y$ that send the south pole of $\mathbb{S}^k$ to the base point of Y. Equivalently, $\Omega^k Y$ is the space of all continuous maps $\lambda : (\mathbb{D}^k, \mathbb{S}^{k-1}) \to (Y, *)$ from the standard closed k-dimensional disc $\mathbb{D}^k$ to Y that map the boundary $\mathbb{S}^{k-1}$ of $\mathbb{D}^k$ to the base point of Y. Let us show, following Boardman and Vogt [23], that $\Omega^k Y$ is a natural topological algebra over the little k-discs operad $\mathcal{D}_k = \{\mathcal{D}_k(n)\}_{n \geq 0}$ recalled in Example 3.3.*

The action $\alpha : \mathcal{D}_k(n) \times (\Omega^k Y)^{\times n} \to \Omega^k Y$ is, for $n \geq 0$, defined as follows. Given $\lambda_i : (\mathbb{D}^k, \mathbb{S}^{k-1}) \to (Y, *) \in \Omega^k Y$, $1 \leq i \leq n$, and little k-discs $d = (d_1, \ldots, d_n) \in \mathcal{D}_k(n)$ as in Example 3.3, then

$$\alpha(d, \lambda_1, \ldots, \lambda_n) : (\mathbb{D}^k, \mathbb{S}^{k-1}) \to (Y, *) \in \Omega^k Y$$

is the map defined to be $\lambda_i : \mathbb{D}^k \to Y$ (suitably rescaled) on the image of d_i, and to be $*$ on the complement of the images of the maps d_i, $1 \leq i \leq n$.

Therefore each k-fold loop space is a $\mathcal{D}_k$-space. The following classical theorem is a certain form of the inverse statement.

THEOREM 5.1 *(Boardman-Vogt [23], May [24]) A path-connected $\mathcal{D}_k$-algebra X has the weak homotopy type of a k-fold loop space.*

The connectedness assumption in the above theorem can be weakened by assuming that the $\mathcal{D}_k$-action makes the set $\pi_0(X)$ of path components of X a group.

EXAMPLE 5.4 *The non-unital operad $\overline{\mathcal{M}}_0$ of stable pointed curves of genus 0 (also called the $\boldsymbol{configuration\ (non-unital)\ operad}$) recalled on page 141 is a non-unital operad in the category of smooth complex projective varieties. It therefore makes sense, as explained in Example 3.4, to consider its homology operad $H_*(\overline{\mathcal{M}}_0, \mathbf{k}) = \{H_*(\overline{\mathcal{M}}_0(n), \mathbf{k})\}_{n \geq 2}$.*

An algebra over this non-unital operad is called a (tree level) ***cohomological conformal field theory*** *or a* ***hyper − commutative algebra*** *[9]. It consist of a family of linear operations* $\{(-,\ldots,-) : V^{\otimes n} \to V\}_{n\geq 2}$ *which are totally symmetric, that is*

$$(v_{\sigma(1)},\ldots,v_{\sigma(n)}) = (v_1,\ldots,v_n),$$

for each permutation $\sigma \in \Sigma_n$*. Moreover, we require the following form of associativity:*

$$\sum_{(S,T)} ((u,v,x_i;\ i\ \in S),w,x_j;\ j\in T) = \sum_{(S,T)} (u,(v,w,x_i;\ i\ \in S),x_j; j\in T), \quad (37)$$

where $u,v,w,x_1,\ldots,x_n \in V$ *and* (S,T) *runs over disjoint decompositions* $S \sqcup T = \{1,\ldots,n\}$*. For* $n = 0$*, (37) means the (usual) associativity of the bilinear operation* $(-,-)$*, i.e.* $((u,v),w) = (u,(v,w))$*. For* $n = 1$ *we get*

$$((u,v),w,x) + ((u,v,x),w) = (u,(v,w,x)) + (u,(v,w),x).$$

EXAMPLE 5.5 *In this example,* $\mathbf{k}$ *is a field of characteristic* 0*. The non-unital operad* $\overline{\mathcal{M}}_0(\mathbb{R}) = \{\overline{\mathcal{M}}_0(\mathbb{R})(n)\}_{n\geq 2}$ *of real points in the configuration operad* $\overline{\mathcal{M}}_0$ *is called the* ***mosaic non − unital operad*** *[25]. Algebras over the homology* $H_*(\overline{\mathcal{M}}_0(\mathbb{R}),\mathbf{k}) = \{H_*(\overline{\mathcal{M}}_0(\mathbb{R})(n),\mathbf{k})\}_{n\geq 2}$ *of this operad were recently identified [26] with* 2*-**Gerstenhaber algebras**, which are structures* (V,μ,τ) *consisting of a commutative associative product* $\mu : V\otimes V \to V$ *and an anti-symmetric degree* $+1$ *ternary operation* $\tau : V\otimes V\otimes V \to V$ *which satisfies the generalized Jacobi identity*

$$\sum_{\sigma} sgn(\sigma)\cdot\tau(\tau(x_{\sigma(1)},x_{\sigma(2)},x_{\sigma(3)}),x_{\sigma(4)},x_{\sigma(5)}) = 0,$$

where the summation runs over all $(3,2)$*-unshuffles* $\sigma(1) < \sigma(2) < \sigma(3)$*,* $\sigma(4) < \sigma(5)$*. Moreover, the ternary operation* τ *is tied to the multiplication* μ *by the distributive law*

$$\tau(\mu(s,t),u,v) = \mu(\tau(s,u,v),t) + (-1)^{(1+|u|+|v|)|s|}\cdot\mu(s,\tau(t,u,v)),\ \ s,t,u,v\in V,$$

saying that the assignment $s \mapsto \tau(s,u,v)$ *is a degree* $(1+|u|+|v|)$*-derivation of the associative commutative algebra* (V,μ)*, for each* $u,v\in V$*.*

6 Free Operads and a Category of rooted Trees

The purpose of this Section is three-fold. First, we want to study free operads because each operad is a quotient of a free one. The second reason why we are interested in free operads is that their construction involves trees. Indeed, it turns out that rooted trees provide 'pasting schemes' for operads and that, replacing trees by other types of graphs, one can introduce several important generalizations of operads, such as cyclic operads, modular operads, and PROPs. The last reason is that the free operad functor defines a monad which provides an unbiased definition of operads as algebras over this monad. Everything in this Section is written for $\mathbf{k}$-linear operads, but the constructions can be generalized into an arbitrary symmetric monoidal category with countable coproducts $(\mathcal{M}, \odot, \mathbf{1})$ whose monoidal product $\odot$ is distributive over coproducts, see [16].

Recall that a Σ-*module* is a collection $E = \{E(n)\}_{n \geq 0}$ in which each $E(n)$ is a right $\mathbf{k}[\Sigma_n]$-module. There is an obvious forgetful functor $\square : \mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$ from the category $\mathtt{Oper}$ of $\mathbf{k}$-linear operads to the category $\Sigma\text{-}\mathtt{mod}$ of Σ-modules.

DEFINITION 6.1 *The* **free operad functor** *is a left adjoint [27]* $\Gamma : \Sigma\text{-}\mathtt{mod} \to \mathtt{Oper}$ *to the forgetful functor* $\square : \mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$. *This means that there exists a functorial isomorphism*

$$Mor_{\mathtt{Oper}}(\Gamma(E), \mathcal{P}) \cong Mor_{\Sigma\text{-}\mathtt{mod}}(E, \square(\mathcal{P}))$$

for an arbitrary Σ*-module* E *and operad* $\mathcal{P}$. *The operad* $\Gamma(E)$ *is the* free operad *generated by the* Σ*-module* E. *Similarly, the* **free non-unital operad functor** *is a left adjoint* $\Psi : \Sigma\text{-}\mathtt{mod} \to \psi\mathtt{Oper}$ *of the obvious forgetful functor* $\square_\psi : \psi\mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$, *that is*

$$Mor_{\psi\mathtt{Oper}}(\Psi(E), \mathcal{S}) \cong Mor_{\Sigma\text{-}\mathtt{mod}}(E, \square_\psi(\mathcal{S})),$$

where E *is a* Σ*-module and* $\mathcal{S}$ *a non-unital operad. The non-unital operad* $\Psi(E)$ *is the* **free non-unital operad** *generated by the* Σ*-module* E.

Let $\widetilde{\ }: \psi\mathtt{Oper} \to \mathtt{Oper}$ be the functor of 'adjoining the unit' considered in the proof of Proposition 4.2 on page 144. Functorial isomorphism (30) implies that one may take

$$\Gamma := \widetilde{\Psi}, \tag{38}$$

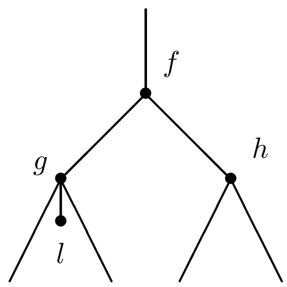
which means that the free operad $\Gamma(E)$ can be obtained from the free non-unital operad $\Psi(E)$ by formally adjoining the unit.

Let us indicate how to construct the free non-unital operad $\Psi(E)$, a precise description will be given later in this Section. The free non-unital operad $\Psi(E)$ must be built up from all formal $\circ_i$-compositions of elements of E modulo the axioms listed in Definition 3.4. For instance, given $f \in E(2)$, $g \in E(3)$, $h \in E(2)$ and $l \in E(0)$, the component $\Psi(E)(5)$ must contain the following five compositions

$$\begin{gathered}(f \circ_1 (g \circ_2 l)) \circ_3 h,\ \ (f \circ_2 h) \circ_1 (g \circ_2 l),\ \ ((f \circ_2 h) \circ_1 g) \circ_2 l,\\ ((f \circ_1 g) \circ_2 l) \circ_3 h \ \text{ and } \ ((f \circ_1 g) \circ_4 h) \circ_2 l.\end{gathered} \tag{39}$$

The elements in (39) can be depicted by the 'flow diagrams' of Figure 13. Nodes of these diagrams are decorated by elements f, g, h and l of E in such a way that an element of $E(n)$ decorates a node with n input lines, $n \geq 0$. Thin 'amoebas' indicate the nesting which specifies the order in which the $\circ_i$-operations are performed.

The associativity of Definition 3.4 however says that the result of the composition does not depend on the order, therefore the amoebas can be erased and the common value of the compositions represented by



(40)

Let us look more closely how diagram (40) determines an element of the (still hypothetical) free non-unital operad $\Psi(E)$. The crucial fact is that the underlying graph of (40) is a planar rooted tree. Recall (see Section 2) that a *tree* is a finite connected simply connected graph without loops and multiple edges. For a tree T we denote, as usual, by $\mathit{Vert}(T)$ the set of vertices and $\mathit{Edg}(T)$ the set of edges of T. The number of edges adjacent to a vertex $v \in \mathit{Vert}(T)$ is called the *valence* of v and denoted $\mathit{val}(v)$. We assume that one is given a subset

$$\mathit{ext}(T) \subset \{v \in \mathit{Vert}(T);\ \mathit{val}(v) = 1\}$$

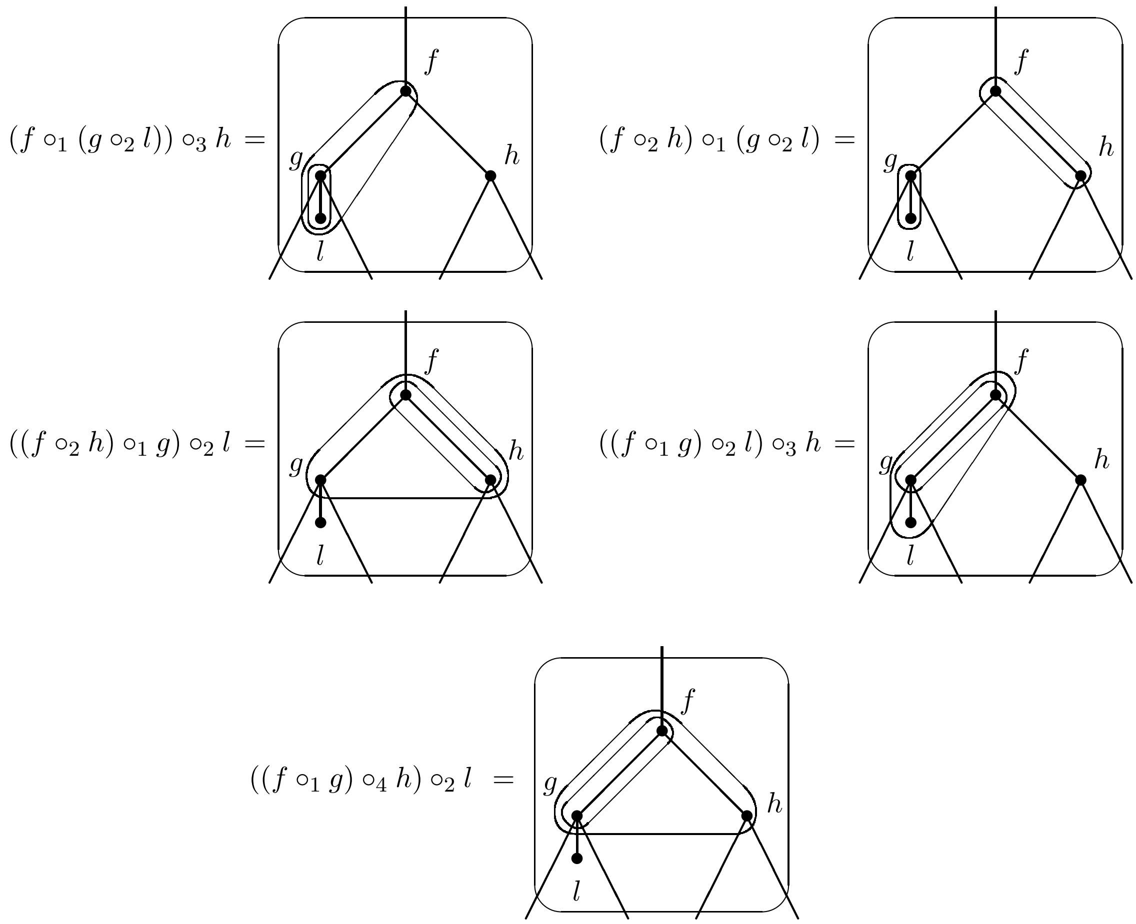


Figure 13: Flow diagrams in non-unital operads.

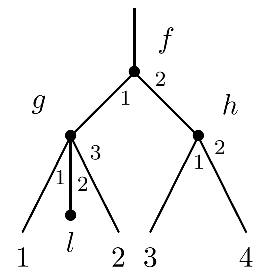
of *external* vertices, the remaining vertices are *internal*. Let us denote

$$vert(T) := Vert(T) \setminus ext(T)$$

the set of all internal vertices. Henceforth, we will assume that our trees have at least one internal vertex. This excludes at this stage the *exceptional tree* consisting of two external vertices connected by an edge.

Edges adjacent to external vertices are the *legs* of T. A tree is *rooted* if one of its legs, called the *root*, is marked and all other edges are oriented, pointing to the root. The legs different from the root are the *leaves* of T. For example, the tree in (40) has 4 internal vertices decorated f, g, h and l, and 4 leaves. Finally, the *planarity* means that an embeddings of T into the plane is specified. In Sections 6–11 for all pictures, the root will always be placed on the top. By a vertex we will always mean an internal one.

The planarity and a choice of the root of the underlying tree of (40) specifies a total order of the set $in(v)$ of input edges of each vertex $v \in vert(T)$ as well as a total order of the set $Leaf(T)$ of the leaves of T, by numbering from the left to the right:



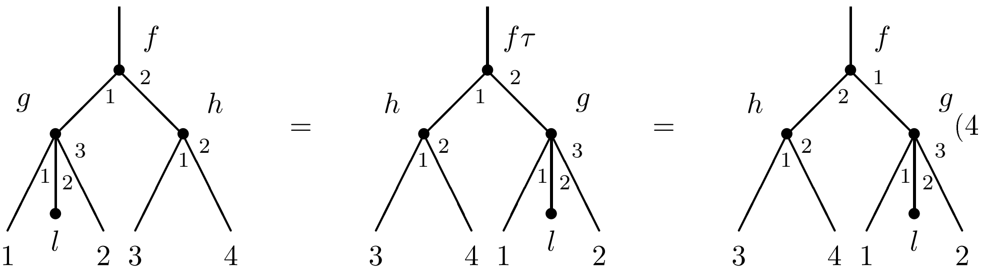
(41)

This tells us that l should be inserted into the second input of g, g into the first input of f and h into the second input of f. Using 'abstract variables' v_1, v_2, v_3 and v_4, the element represented by (41) can also be written as the 'composition' $f(g(v_1, l, v_2), h(v_3, v_4))$.

Now we need to take into account also the symmetric group action. If τ is the generator of Σ_2, then the obvious equality

$$f(g(v_1, l, v_2), h(v_3, v_4)) = f\tau(h(v_3, v_4), g(v_1, l, v_2))$$

of 'abstract compositions' coming from the equivariance of Definition 3.4 translates into the following equality of flow diagrams:



(42)

Relation (42) shows that the equivariance of Definition 3.4 violates the linear orders induced by the planar embedding of T. This leads us to the conclusion that the flow diagrams describing elements of free non-unital operads are (abstract, non-planar) rooted, leaf-labeled decorated trees.

Let us describe, after these motivations, a precise construction of $\Psi(E)$. The first subtlety one needs to understand is how to decorate vertices of non-planar trees. In Sections 6–11, we need to explain how each Σ-module $E = \{E(n)\}_{n\geq 0}$ naturally extends into a functor (denoted again E) from the category $\mathtt{Set}_f$ of finite sets and their bijections to the category of $\mathbf{k}$-modules. If X and Y are finite sets, denote by

$$Bij(Y, X) := \{\vartheta : X \xrightarrow{\cong} Y\} \tag{43}$$

the set of all isomorphisms between X and Y (notice the unexpected direction of the arrow!). It is clear that $Bij(Y, X)$ is a natural left Aut_Y- right Aut_X-bimodule, where $Aut_X := Bij(X, X)$ and $Aut_Y := Bij(Y, Y)$ are the sets of automorphisms with group structure given by composition. For a finite set $S \in \mathtt{Set}_f$ of cardinality n and a Σ-module $E = \{E(n)\}_{n\geq 0}$ define $E(S)$ to be

$$E(S) := E(n) \times_{\Sigma_n} Bij([n], S) \tag{44}$$

where, as usual, $[n] := \{1, \ldots, n\}$ and, of course, $\Sigma_n = Aut_{[n]}$.

Let us recall that a *(leaf-) labeled rooted n-tree* is a rooted tree T together with a specified bijection $\ell : Leaf(T) \xrightarrow{\sim} [n]$. Let $\mathtt{Tree}_n$ be the category of labeled rooted n-trees and their bijections. For $T \in \mathtt{Tree}_n$ define

$$E(T) := \bigotimes_{v \in vert(T)} E(in(v)) \tag{45}$$

where $in(v)$ is, as before, the set of all input edges of a vertex $v \in vert(T)$. It is easy to verify that $E \mapsto E(T)$ defines a functor from the category $\mathtt{Tree}_n$ to the category of $\mathbf{k}$-modules.

Recall that the colimit of a covariant functor $F : \mathcal{D} \to \mathtt{Mod}_{\mathbf{k}}$ is the quotient

$$\operatorname*{colim}_{x \,\in\, \mathcal{D}} F(x) = \bigoplus_{x \in \mathcal{D}} F(x)/\sim,$$

where $\sim$ is the equivalence generated by

$$F(y) \ni a \sim F(f)(a) \in F(z),$$

for each $a \in F(y)$, $y, z \in \mathcal{D}$ and $f \in Mor_{\mathcal{D}}(y, z)$. Define finally

$$\Psi(E)(n) := \operatorname*{colim}_{T \,\in\, \mathtt{Tree}_n} E(T),\ n \geq 0. \tag{46}$$

The following theorem was proved in [16].

Theorem 6.1 *There exists a natural non-unital operad structure on the Σ-module*

$$\Psi(E) = \{\Psi(E)(n)\}_{n \geq 0},$$

with the $\circ_i$-operations given by the grafting of trees and the symmetric group relabeling the leaves, such that $\Psi(E)$ is the free non-unital operad generated by the Σ-module E.

One could simplify (46) by introducing $\mathcal{T}ree(n)$ as the set of *isomorphism* classes of n-trees from $\mathtt{Tree}_n$ and defining $\Psi(E)$ by the formula

$$\Psi(E)(n) = \bigoplus_{[T] \in \mathcal{T}ree(n)} E(T),\ n \geq 0, \tag{47}$$

which does not involve the colimit. The drawback of (47) is that it assumes a choice of a representative $[T]$ of each isomorphism class in $\mathcal{T}ree(n)$, while (46) is functorial and admits simple generalizations to other types of operads and PROPs. See [16] for other representations of the free non-unital operad functor.

Having constructed the free non-unital operad $\Psi(E)$, we may use (38) to define the free operad $\Gamma(E)$. This is obviously equivalent to enlarging, in (46) for $n = 1$, the category $\mathtt{Tree}_n$ by the *exceptional rooted tree* $|$ with one leg and no internal vertex. If we denote this enlarged category of trees and their isomorphisms (which however differs from $\mathtt{Tree}_n$ only at $n = 1$) by $\mathtt{UTree}_n$, we may represent the free operad as

$$\Gamma(E)(n) := \operatorname*{colim}_{T \in \mathtt{UTree}_n} E(T),\ n \geq 0. \tag{48}$$

If E is a Σ-module such that $E(0) = E(1) = 0$, then (47) reduces to a summation over *reduced trees*, that is trees whose all vertices have at least two input edges. By simple combinatorics, the number of isomorphism classes of reduced trees in $\mathtt{Tree}_n$ is finite for each $n \geq 0$. This implies the following proposition that says that operads are relatively small objects.

PROPOSITION 6.1 *Let $E = \{E(n)\}_{n\geq 0}$ be a Σ-module such that*

$$E(0) = E(1) = 0$$

and that $E(n)$ are finite-dimensional for $n \geq 2$. Then the spaces $\Psi(E)(n)$ and $\Gamma(E)(n)$ are finite-dimensional for each $n \geq 0$.

We close this Section by showing how the free operad functor can be used to define operads. It follows from general principles that any operad $\mathcal{P}$ is a quotient $\mathcal{P} = \Gamma(E)/(R)$, where E and R are Σ-modules and (R) is the operadic ideal (see Definition 3.2) generated by R in $\Gamma(E)$.

EXAMPLE 6.1 *The commutative associative operad $\mathcal{C}om$ recalled in Example 3.7 is generated by the Σ-module*

$$E_{\mathcal{C}om}(n) := \begin{cases} \mathbf{k}\cdot\mu, & \text{if} \quad n = 2 \quad \text{and} \\ 0, & \text{if} \quad n \neq 2. \end{cases}$$

where $\mathbf{k}\cdot\mu$ is the trivial representation of Σ_2. The ideal of relations is generated by

$$R_{\mathcal{C}om} := Span_{\mathbf{k}}\{\mu(\mu\otimes id) - \mu(id\otimes\mu)\} \subset \Gamma(E_{\mathcal{C}om})(3),$$

where $\mu(\mu\otimes id) - \mu(id\otimes\mu)$ is the obvious shorthand for $\gamma(\mu,\mu,e) - \gamma(\mu,e,\mu)$, with e the unit of $\Gamma(E_{\mathcal{C}om})$.

Similarly, the operad $\mathcal{A}ss$ for associative algebras reviewed in Example 3.2 is generated by the Σ-module $E_{\mathcal{A}ss}$ such that

$$E_{\mathcal{A}ss}(n) := \begin{cases} \mathbf{k}[\Sigma_2], & \text{if} \quad n = 2 \quad \text{and} \\ 0, & \text{if} \quad n \neq 2. \end{cases}$$

The ideal of relations is generated by the $\mathbf{k}[\Sigma_3]$-closure $R_{\mathcal{A}ss}$ of the associativity

$$\alpha(\alpha \otimes id) - \alpha(id \otimes \alpha) \in \Gamma(E_{\mathcal{A}ss})(3), \tag{49}$$

where α is a generator of the regular representation $E_{\mathcal{A}ss}(2) = \mathbf{k}[\Sigma_2]$.

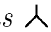
Example 6.2 *The operad $\mathcal{L}ie$ governing Lie algebras is the quotient $\mathcal{L}ie := \Gamma(E_{\mathcal{L}ie})/(R_{\mathcal{L}ie})$, where $E_{\mathcal{L}ie}$ is the Σ-module*

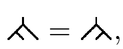
$$E_{\mathcal{L}ie}(n) := \begin{cases} \mathbf{k} \cdot \beta, & \textit{if } n = 2 \textit{ and} \\ 0, & \textit{if } n \neq 2, \end{cases}$$

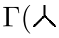
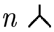
with $\mathbf{k} \cdot \beta$ is the signum representation of Σ_2. The ideal of relations $(R_{\mathcal{L}ie})$ is generated by the Jacobi identity:

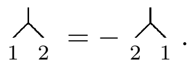
$$\beta(\beta \otimes id) + \beta(\beta \otimes id)c + \beta(\beta \otimes id)c^2 = 0, \tag{50}$$

in which $c \in \Sigma_3$ is the cyclic permutation $(1, 2, 3) \mapsto (2, 3, 1)$.

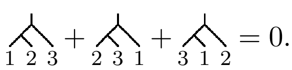
Example 6.3 *We show how to describe the presentations of the operads $\mathcal{A}ss$ and $\mathcal{L}ie$ given in Examples 6.1 and 6.2 in a simple graphical language. The generator α of $E_{\mathcal{A}ss}$ is an operation with two inputs and one output, so we depict it as* *. The associativity (49) then reads as*



therefore $\mathcal{A}ss = \Gamma($*$)/($* *$=$* *$)$. Also the operad for $\mathcal{L}ie$ algebras is generated by one bilinear operation* *, but this time the operation is anti-symmetric*



The Jacobi identity (50) reads



The kind of description used in the above examples is 'tautological' in the sense that it just says that the operad $\mathcal{P}$ governing a certain type of algebras is generated by operations of these algebras, with an appropriate symmetry, modulo the axioms satisfied by these operations. It does not say directly anything about the properties of the individual spaces $\mathcal{P}(n)$, $n \geq 0$. Describing these individual components may be a very nontrivial task, see for example the formula for the Σ_n-modules $\mathcal{L}ie(n)$ given in [16]. Operads in Examples 6.1 and 6.2 are quadratic in the sense of the following:

DEFINITION 6.2 *An operad $\mathcal{P}$ is* ***quadratic*** *if it has a presentation $\mathcal{P} = \Gamma(E)/(R)$, where $E = \mathcal{P}(2)$ and $R \subset \Gamma(E)(3)$.*

Quadratic operads form a very important class of operads. Each quadratic operad $\mathcal{P}$ has a *quadratic dual* $\mathcal{P}^!$ [28], [16] which is a quadratic operad defined, roughly speaking, by dualizing the generators of $\mathcal{P}$ and replacing the relations of $\mathcal{P}$ by their annihilator in the dual space. For example, $\mathcal{A}ss^! = \mathcal{A}ss$, $\mathcal{C}om^! = \mathcal{L}ie$ and $\mathcal{L}ie^! = \mathcal{C}om$. A quadratic operad $\mathcal{P}$ is *Koszul* if it has the homotopy type of the bar construction of its quadratic dual [28], [16]. For quadratic Koszul operads, there is a deep understanding of the derived category of the corresponding algebras. Operads $\mathcal{A}ss$, $\mathcal{C}om$ and $\mathcal{L}ie$ above, as well as most quadratic operads one encounters in everyday life, are Koszul.

7 Category of May's Trees

In this Section, we review the definition of a triple (monad) and give, in Theorem 7.1, a description of unital and non-unital operads in terms of algebras over a triple. The relevant triples come from the endofunctors Ψ and Γ recalled in Section 6. Let $End(\mathcal{C})$ be the strict symmetric monoidal category of endofunctors on a category $\mathcal{C}$ where multiplication is the composition of functors.

DEFINITION 7.1 *A* ***triple*** *(also called a* ***monad****) T on a category $\mathcal{C}$ is an associative and unital monoid (T, μ, υ) in $End(\mathcal{C})$. The multiplication $\mu : TT \to T$ and unit morphism $\upsilon : id \to T$ satisfy the axioms given by commutativity of the diagrams in Figure 14.*

$$
\begin{tikzcd}
TTT \arrow[r, "T\mu"] \arrow[d, "\mu T"'] & TT \arrow[d, "\mu"] \\
TT \arrow[r, "\mu"] & T
\end{tikzcd}
\qquad
\begin{tikzcd}
T \arrow[rr, "T\upsilon"] \arrow[dr, "id"'] & & TT \arrow[dl, "\mu"] \\
& T &
\end{tikzcd}
\qquad
\begin{tikzcd}
T \arrow[rr, "\upsilon T"] \arrow[dr, "id"'] & & TT \arrow[dl, "\mu"] \\
& T &
\end{tikzcd}
$$

Figure 14: Associativity and unit axioms for a triple.

Triples arise naturally from pairs of adjoint functors. Given an adjoint pair [27, II.7]

$$
\begin{tikzcd}
\mathtt{A} \arrow[r, bend left, "G"] & \mathtt{B}, \arrow[l, bend left, "F"]
\end{tikzcd}
$$

with associated functorial isomorphism

$$Mor_{\mathtt{A}}(F(X), Y) \cong Mor_{\mathtt{B}}(X, G(Y)),\ \ X \in \mathtt{B},\ Y \in \mathtt{A},$$

there is a triple in $\mathtt{B}$ defined by $T := GF$. The unit of the adjunction $id \to GF$ defines the unit υ of the triple and the counit of the adjunction $FG \to id$ induces a natural transformation $GFGF \to GF$ which defines the multiplication μ. In fact, it is a theorem of Eilenberg and Moore [29] that all triples arise in this way from adjoint pairs. This is exactly the situation with the free operad and free non-unital operad functors that were described in Section 6. We will show how operads and non-unital operads can actually be defined using the concept of an algebra over a triple:

DEFINITION 7.2 *A T-*algebra *or* ***algebra over the triple*** *T is an object A of $\mathcal{C}$ together with a structure morphism $\alpha : T(A) \to A$ satisfying*

$$\alpha(T(\alpha)) = \alpha(\mu_A) \ \textit{and}\ \alpha\upsilon_A = id_A,$$

see Figure 15.

The category of T-algebras in $\mathcal{C}$ will be denoted $\mathtt{Alg}_T(\mathcal{C})$. Since the free non-unital operad functor Ψ and the free operad functor Γ described in Section 6 are left adjoints to $\square_\psi : \psi\mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$ and $\square : \mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$, respectively, the functors $\square_\psi\Psi$ (denoted simply Ψ) and $\square\Gamma$ (denoted Γ) define triples on $\Sigma\text{-}\mathtt{mod}$.

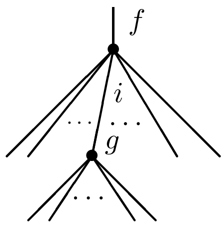
$$
\begin{tikzcd}
T(T(A)) \arrow[r, "T(\alpha)"] \arrow[d, "\mu"'] & T(A) \arrow[d, "\alpha"] \\
T(A) \arrow[r, "\alpha"] & A
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[rr, "\upsilon_A"] \arrow[dr, "id"'] & & T(A) \arrow[dl, "\alpha"] \\
& A &
\end{tikzcd}
$$

Figure 15: T-algebra structure.

THEOREM 7.1 *A Σ-module $\mathcal{S}$ is a Ψ-algebra if and only if it is a non-unital operad and it is a Γ-algebra if and only if it is an operad. In shorthand:*

$$\mathtt{Alg}_\Psi(\Sigma\text{-}\mathtt{mod}) \cong \psi\mathtt{Oper} \quad \textit{and} \quad \mathtt{Alg}_\Gamma(\Sigma\text{-}\mathtt{mod}) \cong \mathtt{Oper}.$$

Proof. We outline first the proof of the implication in the direction from algebra to non-unital operad. Let $\mathcal{S}$ be a Ψ-algebra. The restriction of the structure morphism $\alpha : \Psi(\mathcal{S}) \longrightarrow \mathcal{S}$ to the components of $\Psi(\mathcal{S})$ supported on trees with one internal edge defines the non-unital operad composition maps $\circ_i$, as indicated by:



$$\xmapsto{\alpha} \quad f \circ_i g.$$

In the opposite direction, for a non-unital operad $\mathcal{S}$, the Ψ-algebra structure $\alpha : \Psi(\mathcal{S}) \to \mathcal{S}$ is the contraction along the edges of underlying trees, using the $\circ_i$-operations. The proof that Γ-algebras are operads is similar. ■

Let us change our perspective and consider formula (46) as defining an endofunctor $\Psi : \Sigma\text{-}\mathtt{mod} \to \Sigma\text{-}\mathtt{mod}$, ignoring that we already know that it represents free non-unital operads. We are going to construct maps

$$\mu : \Psi\Psi \to \Psi \text{ and } \upsilon : id \to \Psi$$

making Ψ a triple on the category $\Sigma\text{-}\mathtt{mod}$. Let us start with the triple multiplication μ. It follows from (46) that, for each Σ-module E,

$$\Psi\Psi(E)(n) := \operatorname*{colim}_{T \in \mathtt{Tree}_n} \Psi(E)(T), \ n \geq 0.$$

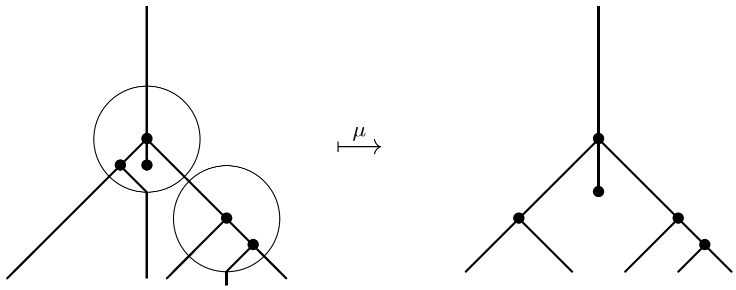
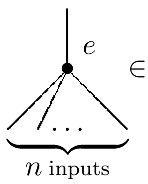


Figure 16: Bracketed trees. The left picture shows an element of $\Psi\Psi(E)(5)$ while the right picture shows the same element interpreted, after erasing the braces indicated by thin cycles, as an element of $\Psi(E)(5)$. For simplicity, we did not show the decoration of vertices by elements of E.

The elements in the right hand side are represented by rooted trees T with vertices decorated by elements of $\Psi(E)$, while elements of $\Psi(E)$ are represented by rooted trees with vertices decorated by E. We may therefore imagine elements of $\Psi\Psi(E)$ as 'bracketed' rooted trees, in the sense indicated in Figure 16. The triple multiplication $\mu_E : \Psi\Psi(E) \to \Psi(E)$ then simply erases the braces. The triple unit $\upsilon_E : E \to \Psi(E)$ identifies elements of E with decorated corollas:

$$E(n) \ni e \quad \longleftrightarrow$$



$$\in \Psi(E)(n),\ n \geq 0.$$

It is not difficult to verify that the above constructions indeed make Ψ a triple, compare [16]. Now we can define non-unital operads as algebras over the triple (Ψ, μ, υ). The advantage of this approach is that, by replacing $\mathtt{Tree}_n$ in (46) by another category of trees or graphs, one may obtain triples defining other types of operads and their generalizations.

We have already seen in (48) that enlarging $\mathtt{Tree}_n$ into $\mathtt{UTree}_n$ by adding the exceptional tree, one gets the triple Γ describing (unital) operads. It is not difficult to see that non-unital May's operads are related to the category $\mathtt{MTree}_n$ of *May's trees* which are, by definition, rooted trees whose vertices can be arranged into

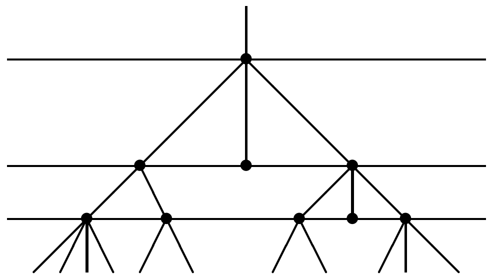


Figure 17: A May's tree.

levels as in Figure 17. Non-unital May's operads are then algebras over the triple $\mathrm{M} : \Sigma\text{-}\mathtt{mod} \to \Sigma\text{-}\mathtt{mod}$ defined by

$$\mathrm{M}(E)(n) := \operatorname*{colim}_{T \in \mathtt{MTree}_n} E(T), \ n \geq 0.$$

These observations are summarized in the first three lines of the table in Figure 24 on page 188.

8 Cyclic Operads and non-rooted Trees

In the following two Sections we use the approach developed in Section 7 to introduce cyclic and modular operads. We recalled, in Example 3.9, the operad $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n\geq 0}$ of Riemann spheres with parametrized labeled holes. Each $\widehat{\mathfrak{M}}_0(n)$ was a right Σ_n-space, with the operadic right Σ_n-action permuting the labels $1, \ldots, n$ of the holes $u_1, \ldots, u_n$. But each $\widehat{\mathfrak{M}}_0(n)$ obviously admits a higher type of symmetry which interchanges labels $0, \ldots, n$ of *all* holes, including the label of the 'output' hole u_0. Another example admitting a similar higher symmetry is the configuration (non-unital) operad $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geq 2}$.

These examples indicate that, for some operads, there is no clear distinction between 'inputs' and the 'output.' Cyclic operads, introduced by E. Getzler and M.M. Kapranov in [8], formalize this phenomenon. They are, roughly speaking, operads with an extra symmetry that interchanges the output with one of the inputs. Let us recall some notions necessary to give a precise definition.

We remind the reader that in this Section, as well as everywhere in Sections 3–11, main definitions are formulated over the underlying category of $\mathbf{k}$-modules, where $\mathbf{k}$ is a commutative associative unital ring. However as in Sections 1–2, for some constructions, we will require $\mathbf{k}$ to be a *field*; we will indicate this as usual by speaking about *vector spaces* instead of $\mathbf{k}$-modules.

Let Σ_n^+ be the permutation group of the set $\{0,\ldots,n\}$. The group Σ_n^+ is, of course, non-canonically isomorphic to the symmetric group Σ_{n+1}. We identify Σ_n with the subgroup of Σ_n^+ consisting of permutations $\sigma \in \Sigma_n^+$ such that $\sigma(0) = 0$. If $\tau_n \in \Sigma_n^+$ denotes the cycle $(0,\ldots,n)$, that is, the permutation with $\tau_n(0) = 1,\ \tau_n(1) = 2,\ \ldots, \tau_n(n) = 0$, then τ_n and Σ_n generate Σ_n^+.

Recall that a *cyclic* Σ*-module* or a Σ^+*-module* is a sequence $W = \{W(n)\}_{n\geq 0}$ such that each $W(n)$ is a (right) $\mathbf{k}[\Sigma_n^+]$-module. Let Σ^+-$\mathtt{mod}$ denote the category of cyclic Σ-modules. As (ordinary) operads were Σ-modules with an additional structure, cyclic operads are Σ^+-modules with an additional structure.

We will also need the following 'cyclic' analog of (44): if X is a set with $n+1$ elements and $W \in \Sigma^+$-$\mathtt{mod}$, then

$$W(\!(X)\!) := W(n) \times_{\Sigma_n^+} \mathit{Bij}([n]^+, X), \qquad (51)$$

where $[n]^+ := \{0,\ldots,n\}$, $n \geq 0$. Double brackets in $W(\!(X)\!)$ remind us that the nth piece of the cyclic Σ-module $W = \{W(n)\}_{n\geq 0}$ is applied on a set with $n+1$ elements, using the extended Σ_n^+-symmetry. Therefore

$$W(\!(\{0,\ldots,n\})\!) \cong W(n) \ \text{ while } \ W(\{0,\ldots,n\}) \cong W(n+1),\ n \geq 0.$$

Pasting schemes for cyclic operads are *cyclic (leg-) labeled n-trees*, by which we mean <u>non-rooted</u> trees, with legs labeled by the set $\{0,\ldots,n\}$. An example of such a tree is given in Figure 18. Since we do not assume a choice of the root, the edges of a cyclic tree C are not directed and it does not make sense to speak about inputs and the output of a vertex $v \in \mathit{vert}(C)$. Let $\mathtt{Tree}_n^+$ be the category of cyclic labeled n-trees and their bijections.

For a cyclic Σ-module W and a cyclic labeled tree T we have the following cyclic version of the product (45)

$$W(\!(T)\!) := \bigotimes_{v \in \mathit{vert}(T)} W(\!(\mathit{edge}(v))\!).$$

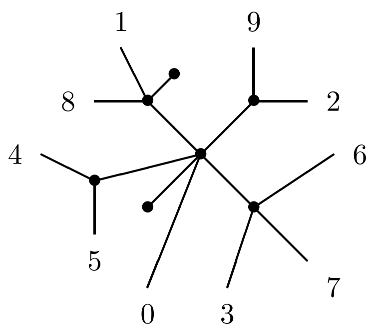


Figure 18: A cyclic labeled tree from $\mathtt{Tree}_9^+$.

The conceptual difference between (45) and the above formula is that instead of the set $in(v)$ of incoming edges of a vertex v of a rooted tree, here we use the set $edge(v)$ of *all* edges incident with v. Let, finally, $\Psi_+ : \Sigma^+\text{-}\mathtt{mod} \to \Sigma^+\text{-}\mathtt{mod}$ be the functor

$$\Psi_+(W)(n) := \operatorname*{colim}_{T \,\in\, \mathtt{Tree}_n^+} W(\!(T)\!), \ n \geq 0, \tag{52}$$

equipped with the triple structure of 'forgetting the braces' similar to that reviewed on page 161. We will use also the 'extended' triple $\Gamma_+ : \Sigma^+\text{-}\mathtt{mod} \to \Sigma^+\text{-}\mathtt{mod}$,

$$\Gamma_+(W)(n) := \operatorname*{colim}_{T \,\in\, \mathtt{UTree}_n^+} W(\!(T)\!), \ n \geq 0,$$

where $\mathtt{UTree}_n^+$ is the obvious extension of the category $\mathtt{Tree}_n^+$ by the exceptional tree $|$.

DEFINITION 8.1 *A* ***cyclic*** *(resp.* ***non − unital cyclic****)* ***operad*** *is an algebra over the triple* Γ_+ *(resp. the triple* Ψ_+*) introduced above.*

In the following proposition, which slightly improves [8], $\tau_n \in \Sigma_n^+$ denotes the cycle $(0, \ldots, n)$.

PROPOSITION 8.1 *A non-unital cyclic operad is the same as a non-unital operad* $\mathcal{C} = \{\mathcal{C}(n)\}_{n\geq 0}$ *(Definition 3.4) such that the right* Σ_n*-action on* $\mathcal{C}(n)$ *extends, for each* $n \geq 0$*, to an action of* Σ_n^+ *with the property that for* $p \in \mathcal{C}(m)$ *and* $q \in \mathcal{C}(n)$*,*

$1 \le i \le m$, $n \ge 0$, *the composition maps satisfy*

$$(p \circ_i q)\tau_{m+n-1} = \begin{cases} (q\tau_n) \circ_n (p\tau_m), \text{ if} i = 1, \text{and} \\ (p\tau_m) \circ_{i-1} q, \text{ for} 2 \le i \le m. \end{cases}$$

The above structure is a (unital) cyclic operad if moreover there exists a Σ_1^+-invariant operadic unit $e \in \mathcal{C}(1)$.

Proposition 8.1 gives a biased definition of cyclic operads whose obvious modification (see [16]) makes sense in an arbitrary symmetric monoidal category. We can therefore speak about topological cyclic operads, differential graded cyclic operads, simplicial cyclic operads etc. Observe that there are no *non-unital cyclic May's operads* because it does not make sense to speak about levels in trees without a choice of the root.

EXAMPLE 8.1 *Let V be a finite dimensional vector space and $B : V \otimes V \to \mathbf{k}$ a nondegenerate symmetric bilinear form. The form B induces the identification*

$$Lin(V^{\otimes n}, V) \ni f \longmapsto \widehat{B}(f) := B(-, f(-)) \in Lin(V^{\otimes(n+1)}, \mathbf{k})$$

of the spaces of linear maps. The standard right Σ_n^+-action

$$\widehat{B}(f)\sigma(v_0, \ldots, v_n) = \widehat{B}(f)(v_{\sigma^{-1}(0)}, \ldots, v_{\sigma^{-1}(n)}),\ \sigma \in \Sigma_n^+,\ v_0, \ldots, v_n \in V,$$

defines, via this identification, a right Σ_n^+-action on $Lin(V^{\otimes n}, V)$, that is, on the nth piece of the endomorphism operad $\mathcal{E}nd_V = \{\mathcal{E}nd_V(n)\}_{n\ge 0}$ recalled in Example 3.1. It is easy to show that, with the above action, $\mathcal{E}nd_V$ is a cyclic operad in the monoidal category of vector spaces, called the ***cyclic endomorphism operad*** *of the pair $V = (V, B)$. The biased definition of cyclic operads given in Proposition 8.1 can be read off from this example.*

EXAMPLE 8.2 *We saw in Example 3.5 that a unital operad $\mathcal{A} = \{\mathcal{A}(n)\}_{n\ge 0}$ such that $\mathcal{A}(n) = 0$ for $n \ne 1$ is the same as a unital associative algebra. Similarly, it can be easily shown that a cyclic operad $\mathcal{C} = \{\mathcal{C}(n)\}_{n\ge 0}$ satisfying $\mathcal{C}(n) = 0$ for $n \ne 1$ is the same as a unital associative algebra A with a linear involutive antiautomorphism, by which we mean a $\mathbf{k}$-linear map $^* : A \to A$ such that*

$$(ab)^* = b^* a^*,\ (a^*)^* = a \ \text{ and } \ 1^* = 1,$$

for arbitrary $a, b \in A$.

Let $\mathcal{P} = \Gamma(E)/(R)$ be a quadratic operad as in Definition 6.2. The action of Σ_2 on E extends to an action of Σ_2^+, via the sign representation $\mathrm{sgn} : \Sigma_2^+ \to \{\pm 1\} = \Sigma_2$. It can be easily verified that this action induces a cyclic operad structure on the free operad $\Gamma(E)$. In particular, $\Gamma(E)(3)$ is a right Σ_3^+-module.

DEFINITION 8.2 *We say that the operad $\mathcal{P}$ is a* ***cyclic quadratic operad*** *if, in the above presentation, R is a Σ_3^+-invariant subspace of $\Gamma(E)(3)$.*

If the condition of the above definition is satisfied, $\mathcal{P}$ has a natural induced cyclic operad structure.

EXAMPLE 8.3 *By [8], all quadratic operads generated by a one-dimensional space are cyclic quadratic, therefore the operads $\mathcal{L}ie$ and $\mathcal{C}om$ are cyclic quadratic. Also the operads $\mathcal{A}ss$ and the operad $\mathcal{P}oiss$ for Poisson algebras are cyclic quadratic [8]. A surprisingly simple operad which is cyclic and quadratic, but not cyclic quadratic, is constructed in [62].*

The operad $\widehat{\mathfrak{M}}_0$ of Riemann spheres with labeled punctures reviewed in Example 3.9 is a topological cyclic operad. The configuration operad $\overline{\mathcal{M}}_0$ recalled on page 141 is a non-unital topological cyclic operad. Important examples of non-cyclic operads are the operad pre-$\mathcal{L}ie$ for pre-Lie algebras [62, Section 3] and the operad $\mathcal{L}eib$ for Leibniz algebras [8].

Let $\mathcal{C}$ be an operad, $\alpha : \mathcal{C}(n) \otimes V^{\otimes n} \to V$, $n \geq 0$, a $\mathcal{C}$-algebra with the underlying vector space V as in Proposition 5.1 and $B : V \otimes V \to U$ a bilinear form on V with values in a vector space U. We can form a map

$$\widetilde{B}(\alpha) : \mathcal{C}(n) \otimes V^{\otimes(n+1)} \to U,\ n \geq 0, \tag{53}$$

by the formula

$$\widetilde{B}(\alpha)(c \otimes v_0 \otimes \cdots v_n) := B(v_0, \alpha(c \otimes v_1 \otimes \cdots v_n)),\ c \in \mathcal{C}(n),\ v_0, \ldots, v_n \in V.$$

Suppose now that the operad $\mathcal{C}$ is cyclic, in particular, that each $\mathcal{C}(n)$ is a right Σ_n^+-module. We say that the bilinear form $B : V \otimes V \to U$ is *invariant* [8], if the maps $\widetilde{B}(\alpha)$ in (53) are, for each $n \geq 0$, invariant under the diagonal action of Σ_n^+ on $\mathcal{C}(n) \otimes V^{\otimes(n+1)}$. We leave as an exercise to verify that the invariance of

$\widetilde{B}(\alpha)$ for $n = 1$ together with the existence of the operadic unit implies that B is symmetric,

$$B(v_0, v_1) = B(v_1, v_0), \ v_0, v_1 \in V.$$

DEFINITION 8.3 *A* ***cyclic algebra*** *over a cyclic operad $\mathcal{C}$ is a $\mathcal{C}$-algebra structure on a vector space V together with a nondegenerate invariant bilinear form $B : V \otimes V \to \mathbf{k}$.*

By [16], a cyclic algebra is the same as a cyclic operad homomorphism $\mathcal{C} \to \mathcal{E}nd_V$, where $\mathcal{E}nd_V$ is the cyclic endomorphism operad of the pair (V, B) recalled in Example 8.1.

EXAMPLE 8.4 *A cyclic algebra over the cyclic operad $\mathcal{C}om$ is a* ***Frobenius algebra****, that is, a structure consisting of a commutative associative multiplication $\cdot : V \otimes V \to V$ as in Example 5.1 together with a non-degenerate symmetric bilinear form $B : V \otimes V \to \mathbf{k}$, invariant in the sense that*

$$B(a \cdot b, c) = B(a, b \cdot c), \quad \textit{for all} \ \ a, b, c \in V.$$

Similarly, a cyclic Lie algebra is given by a Lie bracket $[-,-] : V \otimes V \to V$ and a non-degenerate symmetric bilinear form $B : V \otimes V \to \mathbf{k}$ satisfying

$$B([a, b], c) = B(a, [b, c]), \quad \textit{for} \ \ a, b, c \in V.$$

For algebras over cyclic operads, one may introduce cyclic cohomology that generalizes the classical cyclic cohomology of associative algebras [63–65] as the non-abelian derived functor of the universal bilinear form [8], [16]. Let us close this Section by mentioning two examples of operads with other types of higher symmetries. The symmetry required for *anticylic operads* differs from the symmetry of cyclic operads by the sign [16]. *Dihedral operads* exhibit a symmetry governed by the dihedral groups [62].

9 Modular Operads

Let us consider again the Σ^+-module $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n \geq 0}$ of Riemann spheres with punctures. We saw that the operation $M, N \mapsto M \circ_i N$ of sewing the 0th hole of the surface N to the ith hole of the surface M defined on $\widehat{\mathfrak{M}}_0$ a cyclic operad structure. One may generalize this operation by defining, for $M \in \widehat{\mathfrak{M}}_0(m)$, $N \in \widehat{\mathfrak{M}}_0(n)$, $0 \leq i \leq m$, $0 \leq j \leq n$, the element $M_i\circ_j N \in \widehat{\mathfrak{M}}_0(m+n-1)$ by sewing the jth hole of M to the ith hole of N. Under this notation, $\circ_i = {}_i\circ_0$. In the same manner, one may consider a single surface $M \in \widehat{\mathfrak{M}}_0(n)$, choose labels i, j, $0 \leq i \neq j \leq n$, and sew the ith hole of M along the jth hole of the *same* surface. The result is a new surface $\xi_{\{i,j\}}(M)$, with $n-2$ holes and genus 1.

This leads us to the system $\widehat{\mathfrak{M}} = \{\widehat{\mathfrak{M}}(g,n)\}_{g\geq 0, n\geq -1}$, where $\widehat{\mathfrak{M}}(g,n)$ denotes now the moduli space of genus g Riemann surfaces with $n+1$ holes. Observe that we include $\widehat{\mathfrak{M}}(g,n)$ also for $n=-1$; $\widehat{\mathfrak{M}}(g,-1)$ is the moduli space of Riemann surfaces of genus g. The operations ${}_i\circ_j$ and $\xi_{\{i,j\}}$ act on $\widehat{\mathfrak{M}}$. Clearly, for $M \in \widehat{\mathfrak{M}}(g,m)$ and $N \in \widehat{\mathfrak{M}}(h,n)$, $0 \leq i \leq m$, $0 \leq j \leq n$ and $g, h \geq 0$,

$$M_i\circ_j N \in \widehat{\mathfrak{M}}(g+h, m+n-1) \tag{54}$$

and, for $m \geq 1$ and $g \geq 0$,

$$\xi_{\{i,j\}}(M) \in \widehat{\mathfrak{M}}(g+1, m-2). \tag{55}$$

A particular case of (54) is the non-operadic composition

$${}_0\circ_0 : \widehat{\mathfrak{M}}(g,0) \times \widehat{\mathfrak{M}}(h,0) \to \widehat{\mathfrak{M}}(g+h,-1), \ \ g,h \geq 0. \tag{56}$$

Modular operads are abstractions of the above structure satisfying a certain additional stability condition. The following definitions, taken from [66], are made for the category of $\mathbf{k}$-modules, but they can be easily generalized to an arbitrary symmetric monoidal category with finite colimits, whose monoidal product $\odot$ is distributive over colimits. Let us introduce the underlying category for modular operads.

A *modular* Σ*-module* is a sequence $\mathcal{E} = \{\mathcal{E}(g,n)\}_{g\geq 0, n\geq -1}$ of $\mathbf{k}$-modules such that each $\mathcal{E}(g,n)$ has a right $\mathbf{k}[\Sigma_n^+]$-action. We say that $\mathcal{E}$ is *stable* if

$$\mathcal{E}(g,n) = 0 \ \text{ for } 2g+n-1 \leq 0 \tag{57}$$

and denote `MMod` the category of stable modular Σ-modules.

Stability (57) says that $\mathcal{E}(g,n)$ is trivial for $(g,n) = (0,-1),\ (1,-1),\ (0,0)$ and $(0,1)$. We will sometimes express the stability of $\mathcal{E}$ by writing $\mathcal{E} = \{\mathcal{E}(g,n)\}_{(g,n)\in\mathfrak{S}}$, where

$$\mathfrak{S} := \{(g,n) \mid g \geq 0,\ n \geq -1 \text{ and } 2g+n-1 > 0\}.$$

Recall that a genus g Riemann surface with k marked points is stable if it does not admit infinitesimal automorphisms. This happens if and only if $2(g-1)+k > 0$, that is, excluded is the torus with no marked points and the sphere with less than three marked points. Thus the stability property of modular Σ-modules is analogous to the stability of Riemann surfaces.

Now we introduce graphs that serve as pasting schemes for modular operads. The naive notion of a graph as we have used it up to this point is not subtle enough; we need to replace it by a more sophisticated:

DEFINITION 9.1 *A **graph** Γ is a finite set Flag(Γ) (whose elements are called **flags** or **half− edges**) together with an involution σ and a partition λ. The **vertices** vert(Γ) of a graph Γ are the blocks of the partition λ, we assume also that the number of these blocks is finite. The **edges** Edg(Γ) are pairs of flags forming a two-cycle of σ. The **legs** Leg(Γ) are the fixed points of σ.*

We also denote by *edge*(v) the flags belonging to the block v or, in common speech, half-edges adjacent to the vertex v. We say that graphs Γ_1 and Γ_2 are *isomorphic* if there exists a set isomorphism $\varphi : \mathit{Flag}(\Gamma_1) \to \mathit{Flag}(\Gamma_2)$ that preserves the partitions and commutes with the involutions. We may associate to a graph Γ a finite one-dimensional cell complex $|\Gamma|$, obtained by taking one copy of $[0,\frac{1}{2}]$ for each flag, a point for each block of the partition, and imposing the following equivalence relation: The points $0 \in [0,\frac{1}{2}]$ are identified for all flags in a block of the partition λ with the point corresponding to the block, and the points $\frac{1}{2} \in [0,\frac{1}{2}]$ are identified for pairs of flags exchanged by the involution σ.

We call $|\Gamma|$ the *geometric realization* of Γ. Observe that empty blocks of the partition generate isolated vertices in the geometric realization. We will usually make no distinction between the graph and its geometric realization. As an example (taken from [66]), consider the graph with $\{a,b,\ldots,i\}$ as the set of flags, the involution $\sigma = (df)(eg)$ and the partition $\{a,b,c,d,e\} \cup \{f,g,h,i\}$. The geometric realization of this graph is the 'sputnik' in Fig. 19. Let us introduce labeled

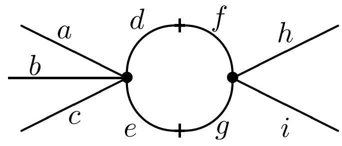


Figure 19: The sputnik.

versions of the above notions. A *(vertex-) labeled graph* is a connected graph Γ together with a map g (the *genus map*) from $vert(\Gamma)$ to the set $\{0, 1, 2, \ldots\}$. Labeled graphs Γ_1 and Γ_2 are isomorphic if there exists an isomorphism preserving the labels of the vertices. The *genus* $g(\Gamma)$ of a labeled graph Γ is defined by

$$g(\Gamma) := b_1(\Gamma) + \sum_{v \in vert(\Gamma)} g(v), \tag{58}$$

where $b_1(\Gamma) := \dim H_1(|\Gamma|)$ is the first Betti number of the graph $|\Gamma|$, i.e. the number of independent circuits of Γ. A graph Γ is *stable* if

$$2(g(v) - 1) + |edge(v)| > 0,$$

at each vertex $v \in vert(\Gamma)$.

For $g \geq 0$ and $n \geq -1$, let $\mathtt{MGr}(g, n)$ be the groupoid whose objects are pairs (Γ, ℓ) consisting of a stable (vertex-) labeled graph Γ of genus g and an isomorphism $\ell : Leg(\Gamma) \to \{0, \ldots, n\}$ labeling the legs of Γ by elements of $\{0, \ldots, n\}$. Morphisms of $\mathtt{MGr}(g, S)$ are isomorphisms of vertex-labeled graphs preserving the labeling of the legs. The stability implies, via an elementary combinatorial topology that, for each fixed $g \geq 0$ and $n \geq -1$, there is only a finite number of isomorphism classes of stable graphs $\Gamma \in \mathtt{MGr}(g, n)$, see [66].

We will also need the following obvious generalization of (51): if $\mathcal{E} = \{\mathcal{E}(g, n)\}_{g \geq 0, n \geq -1}$ is a modular Σ-module and X a set with $n + 1$ elements, then

$$\mathcal{E}(\!(g, X)\!) := \mathcal{E}(g, n) \times_{\Sigma_n^+} Bij([n]^+, X),\ g \geq 0,\ n \geq -1. \tag{59}$$

For a modular Σ-module $\mathcal{E} = \{\mathcal{E}(g, n)\}_{g \geq 0, n \geq -1}$ and a labeled graph Γ, let $\mathcal{E}(\!(\Gamma)\!)$ be the product

$$\mathcal{E}(\!(\Gamma)\!) := \bigotimes_{v \in vert(\Gamma)} \mathcal{E}(\!(g(v), edge(v))\!). \tag{60}$$

Evidently, the correspondence $\Gamma \mapsto \mathcal{E}(\!(\Gamma)\!)$ defines a functor from the category $\mathtt{MGr}(g,n)$ to the category of $\mathbf{k}$-modules and their isomorphisms. We may thus define an endofunctor $\mathbb{M}$ on the category $\mathtt{MMod}$ of stable modular Σ-modules by the formula

$$\mathbb{M}\mathcal{E}(g,n) := \operatorname*{colim}_{\Gamma \in \mathtt{MGr}(g,n)} \mathcal{E}(\!(\Gamma)\!),\ g \geq 0,\ n \geq -1.$$

Choosing a representative for each isomorphism class in $\mathtt{MGr}(g,n)$, one obtains the identification

$$\mathbb{M}\mathcal{E}(g,n) \cong \bigoplus_{[\Gamma]\in\{\mathtt{MGr}(g,n)\}} \mathcal{E}(\!(\Gamma)\!)_{\mathrm{Aut}(\Gamma)},\ g \geq 0,\ n \geq -1, \tag{61}$$

where $\{\mathtt{MGr}(g,n)\}$ is the set of isomorphism classes of objects of the groupoid $\mathtt{MGr}(g,n)$ and the subscript $\mathrm{Aut}(\Gamma)$ denotes the space of coinvariants. Stability (57) implies that the summation in the right-hand side of (61) is finite. Formula (61) generalizes (47) which does not contain coinvariants because there are no non-trivial automorphisms of leaf-labeled trees. On the other hand, stable labeled graphs with nontrivial automorphisms are abundant, an example can be easily constructed from the graph in Figure 19. The functor $\mathbb{M}$ carries a triple structure of 'erasing the braces' similar to the one used on pages 161 and 164.

Definition 9.2 *A* ***modular operad*** *is an algebra over the triple* $\mathbb{M} : \mathtt{MMod} \to \mathtt{MMod}$.

Therefore a modular operad is a stable modular Σ-module $\mathcal{A} = \{\mathcal{A}(g,n)\}_{(g,n)\in\mathfrak{S}}$ equipped with operations that determine coherent contractions along stable modular graphs. Observe that the stability condition is built firmly into the very definition. Very crucially, modular operads *do not have units*, because such a unit ought to be an element of the space $\mathcal{A}(0,1)$ which is empty, by (57).

One can easily introduce un-stable modular operads and their unital versions, but the main motivating example reviewed below is stable. We will consider an extension of the Grothendieck-Knudsen configuration operad $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geq 2}$ consisting of moduli spaces of stable curves of arbitrary genera in the sense of the following generalization of Definition 4.1:

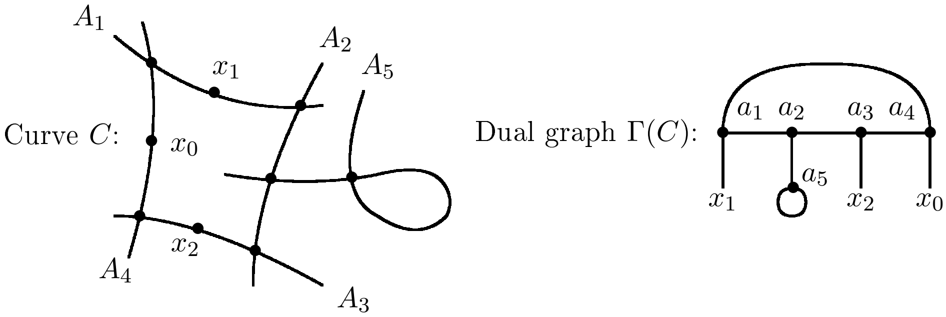


Figure 20: A stable curve and its dual graph. The curve C on the left has five components A_i, $1 \le i \le 5$, and three marked points x_0, x_1 and x_2. The dual graph $\Gamma(C)$ on the right has five vertices a_i, $1 \le i \le 5$, corresponding to the components of the curve and three legs labeled by the marked points.

DEFINITION 9.3 *A* ***stable*** $(n+1)$*-**pointed curve**, $n \geq 0$, is a connected complex projective curve C with at most nodal singularities, together with a 'marking' given by a choice $x_0, \dots, x_n \in C$ of smooth points. The stability means, as usual, that there are no infinitesimal automorphisms of C fixing the marked points and double points.*

The stability in Definition 9.3 is equivalent to saying that each smooth component of C isomorphic to the complex projective space $\mathbb{CP}^1$ has at least three special points and that each smooth component isomorphic to the torus has at least one special point, where by a special point we mean either a double point or a node.
The *dual graph* $\Gamma = \Gamma(C)$ of a stable $(n+1)$-pointed curve $C = (C, x_0, \dots, x_n)$ is a labeled graph whose vertices are the components of C, edges are the nodes and its legs are the points $\{x_i\}_{0 \leq i \leq n}$. An edge e_y corresponding to a nodal point y joins the vertices corresponding to the components intersecting at y. The vertex v_K corresponding to a branch K is labeled by the genus of the normalization of K. See [30] for the normalization and recall that a curve is normal if and only if it is nonsingular. The construction of $\Gamma(C)$ from a curve C is visualized in Figure 20.

Let us denote by $\overline{\mathcal{M}}_{g,n+1}$ the coarse moduli space [30] of stable $(n+1)$-pointed curves C such that the dual graph $\Gamma(C)$ has genus g, in the sense of (58). The genus of $\Gamma(C)$ in fact equals the arithmetic genus of the curve C, thus $\overline{\mathcal{M}}_{g,n+1}$ is the coarse moduli space of stable curves of arithmetic genus g with $n+1$ marked points. By a result of P. Deligne, F.F. Knudsen and D. Mumford [21, 31, 32], $\overline{\mathcal{M}}_{g,n+1}$ is a projective variety.

Observe that, for a curve $C \in \overline{\mathcal{M}}_{0,n+1}$, the graph $\Gamma(C)$ must necessarily be a tree and all components of C must be smooth of genus 0, therefore $\overline{\mathcal{M}}_{0,n+1}$ coincides with the moduli space $\overline{\mathcal{M}}_0(n)$ of genus 0 stable curves with $n+1$ marked points that we discussed in Section 4. Dual graphs of curves $C \in \overline{\mathcal{M}}_{g,n+1}$ are stable labeled graphs belonging to $\mathtt{MGr}(g, n+1)$.

The symmetric group Σ_n^+ acts on $\overline{\mathcal{M}}_{g,n+1}$ by renumbering the marked points, therefore

$$\overline{\mathcal{M}} := \{\overline{\mathcal{M}}(g,n)\}_{g\geq 0, n\geq -1},$$

with $\overline{\mathcal{M}}(g,n) := \overline{\mathcal{M}}_{g,n+1}$, is a modular Σ-module in the category of projective varieties. Since there are no stable curves of genus g with $n+1$ punctures if $2g+n-1 \leq 0$, $\overline{\mathcal{M}}$ is a *stable* modular Σ-module. Let us define the contraction along a stable graph $\Gamma \in \mathtt{MGr}(g,n)$

$$\alpha_\Gamma : \overline{\mathcal{M}}(\!(\Gamma)\!) = \prod_{v \in vert(\Gamma)} \overline{\mathcal{M}}(\!(g(v), edge(v))\!) \to \overline{\mathcal{M}}(g,n) \tag{62}$$

by gluing the marked points of curves from $\overline{\mathcal{M}}(\!(g(v), edge(v))\!)$, $v \in vert(\Gamma)$, according to the graph Γ. To be more precise, let

$$\prod_{v\in vert(\Gamma)} C_v, \quad \text{where} \quad C_v \in \overline{\mathcal{M}}(\!(g(v), edge(v))\!),$$

be an element of $\overline{\mathcal{M}}(\!(\Gamma)\!)$. Let e be an edge of the graph Γ connecting vertices v_1 and v_2, $e = \{y^e_{v_1}, y^e_{v_2}\}$, where $y^e_{v_i}$ is a marked point of the component C_{v_i}, $i = 1,2$, which is also the name of the corresponding flag of the graph Γ. The curve $\alpha_\Gamma(C)$ is then obtained by the identifications $y^e_{v_1} = y^e_{v_2}$, introducing a nodal singularity, for all $e \in Edg(\Gamma)$. The procedure is the same as that described for the tree level in Section 4. As proved in [66, § 6.2], the contraction maps (62) define on the stable modular Σ-module of coarse moduli spaces $\overline{\mathcal{M}} = \{\overline{\mathcal{M}}(g,n)\}_{(g,n)\in\mathfrak{S}}$ a modular operad structure in the category of complex projective varieties.

Let us look more closely at the structure of the modular triple $\mathbb{M}$. Given a (stable or unstable) modular Σ-module $\mathcal{E}$, there is, for each $g \geq 0$ and $n \geq -1$, a natural decomposition

$$\mathbb{M}(\mathcal{E})(g,n) = \mathbb{M}_0(\mathcal{E})(g,n) \oplus \mathbb{M}_1(\mathcal{E})(g,n) \oplus \mathbb{M}_2(\mathcal{E})(g,n) \oplus \cdots ,$$

with $\mathbb{M}_k(\mathcal{E})(g,n)$ the subspace obtained by summing over graphs Γ with $\dim H_1(|\Gamma|) = k$, $k \geq 0$. In particular, $\mathbb{M}_0(\mathcal{E})(g,n)$ is a summation over simply connected graphs. It is not difficult to see that $\mathbb{M}_0(\mathcal{E})$ is a *subtriple* of $\mathbb{M}(\mathcal{E})$. This shows that modular operads are $\mathbb{M}_0$-algebras with some additional operations (the 'contractions') that raise the genus and generate the higher components $\mathbb{M}_k$, $k \geq 1$, of the modular triple $\mathbb{M}$.

There seems to be a belief expressed in the proof of Theorem in [66] that, in the stable case, the triple $\mathbb{M}_0$ is equivalent to the non-unital cyclic operad triple Ψ_+, but it is not so. The triple $\mathbb{M}_0$ is *much bigger*, for example, if $a \in \mathcal{E}(1,0)$, then $\mathbb{M}_0(\mathcal{E})(2,-1)$ contains a non-operadic element

which can be also written, using (56), as $a_0 \circ_0 a$. The corresponding part $\Psi_+(\mathcal{E})(-1)$ of the cyclic triple is empty. In the Grothendieck-Knudsen modular operad $\overline{\mathcal{M}}$, an element of the above type is realized by two tori meeting at a nodal point.

On the other hand, the triple $\mathbb{M}_0$ restricted to the subcategory of stable modular Σ-modules $\mathcal{E}$ such that $\mathcal{E}(g,n) = 0$ for $g > 0$ indeed coincides with the non-unital cyclic operad triple Ψ_+, as was in fact proved in [66]. Therefore, given a modular operad $\mathcal{A} = \{\mathcal{A}(g,n)\}_{(g,n)\in\mathfrak{S}}$, there is an induced non-unital cyclic operad structure on the cyclic collection $\mathcal{A}^\flat := \{\mathcal{A}(0,n)\}_{n\geq 2}$. We will call $\mathcal{A}^\flat$ the *associated cyclic operad*. For example, the cyclic operad associated to the Grothendieck-Knudsen modular operad $\overline{\mathcal{M}}$ equals its genus zero part $\overline{\mathcal{M}}_0$.

A *biased* definition of modular operads can be found in [16]. It is formulated in terms of operations

$$\{{}_i\circ_j : \mathcal{A}(g,m) \otimes \mathcal{A}(h,n) \to \mathcal{A}(g+h, m+n);\ 0 \leq i \leq m,\ 0 \leq j \leq n,\ g,h \geq 0\}$$

together with contractions

$$\left\{\xi_{\{i,j\}} : \mathcal{A}(g,m) \to \mathcal{A}(g+1, m-2);\ m \geq 1,\ g \geq 0\right\}$$

that generalize (54) and (55).

Example 9.1 *Let $V = (V, B)$ be a vector space with a symmetric inner product $B : V \otimes V \to \mathbf{k}$. Denote, for each $g \geq 0$ and $n \geq -1$,*

$$\mathcal{E}nd_V(g,n) := V^{\otimes(n+1)}.$$

It is clear from definition (60) that, for any labeled graph $\Gamma \in \mathtt{MGr}(g,n)$, $\mathcal{E}nd_V(\!(\Gamma)\!) = V^{\otimes Flag(\Gamma)}$.

Let $B^{\otimes Edg(\Gamma)} : V^{\otimes Flag(\Gamma)} \to V^{\otimes Leg(\Gamma)}$ be the multilinear form which contracts the factors of $V^{\otimes Flag(\Gamma)}$ corresponding to the flags which are paired up as edges of Γ. Then we define $\alpha_\Gamma : \mathcal{E}nd_V(\!(\Gamma)\!) \to \mathcal{E}nd_V(g,n)$ to be the map

$$\alpha_\Gamma : \mathcal{E}nd_V(\!(\Gamma)\!) = V^{\otimes Flag(\Gamma)} \xrightarrow{B^{\otimes Edg(\Gamma)}} V^{\otimes Leg(\Gamma)} \xrightarrow{V^{\otimes \ell}} V^{\otimes(n+1)} = \mathcal{E}nd_V(g,n),$$

where $\ell : Leg(\Gamma) \to \{0, \ldots, n\}$ is the labeling of the legs of Γ. It is easy to show that the compositions $\{\alpha_\Gamma;\ \Gamma \in \mathtt{MGr}(g,n)\}$ define on $\mathcal{E}nd_V$ the structure of an un-stable unital modular operad, see [66].

An *algebra* over a modular operad $\mathcal{A}$ is a vector space V with an inner product B, together with a morphism $\rho : \mathcal{A} \to \mathcal{E}nd_V$ of modular operads. Several important structures are algebras over modular operads. For example, an algebra over the homology $H_*(\overline{\mathcal{M}})$ of the Grothendieck-Knudsen modular operad is the same as a cohomological field theory in the sense of [9]. Other physically relevant algebras over modular operads can be found in [16,33,66]. Relations between modular operads, chord diagrams and Vassiliev invariants are studied in [61].

10 PROPs

Operads are devices invented to describe structures consisting of operations with several inputs and *one* output. There are, however, important structures with operations having several inputs and *several* outputs. Let us recall the most prominent one:

Example 10.1 *A (associative)* ***bialgebra*** *is a* $\mathbf{k}$*-module V with a* ***multiplication*** *$\mu : V \otimes V \to V$ and a* ***comultiplication*** *(also called a* ***diagonal****) $\Delta : V \to V \otimes V$. The multiplication is associative:*

$$\mu(\mu \otimes id_V) = \mu(id_V \otimes \mu),$$

the comultiplication is coassociative:

$$(\Delta \otimes id_V)\Delta = (id_V \otimes \Delta)\Delta$$

and the usual compatibility between μ and Δ is assumed:

$$\Delta(u \cdot v) = \Delta(u) \cdot \Delta(v) \quad \textit{for} \quad u, v \in V, \tag{63}$$

where $u{\cdot}v := \mu(u, v)$ and the dot $\cdot$ in the right hand side denotes the multiplication induced on $V \otimes V$ by μ. Loosely speaking, bialgebras are Hopf algebras without unit, counit and antipode.

PROPs (an abbreviation of **pro**duct and **p**ermutation category) describe structures as in Example 10.1. Although PROPs are more general than operads, they appeared much sooner, in a 1965 Mac Lane's paper [34]. This might be explained by the fact that the definition of PROPs is more compact than that of operads – compare Definition 10.1 below with Definition 1.1 in Section 1 and Definition 3.1 in Section 3. PROPs then entered the 'renaissance of operads' in 1996 via [35].

Definition 10.1 uses the notion of a symmetric strict monoidal category, see [16, 34,60]. An example is the category $\mathtt{Mod}_{\mathbf{k}}$ of $\mathbf{k}$-modules, with the monoidal product $\odot$ given by the tensor product $\otimes = \otimes_{\mathbf{k}}$, the symmetry $S_{U,V} : U \otimes V \to V \otimes U$ defined as $S_{U,V}(u, v) := v \otimes u$ for $u \in U$ and $v \in V$, and the unit 1 the ground ring $\mathbf{k}$.

DEFINITION 10.1 *A ($\mathbf{k}$-linear)* PROP *(called a* theory *in [35]) is a symmetric strict monoidal category $\mathsf{P} = (\mathsf{P}, \odot, S, 1)$ enriched over $\mathtt{Mod}_{\mathbf{k}}$ such that*

(i) the objects are indexed by (or identified with) the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of natural numbers, and

(ii) the product satisfies $m \odot n = m + n$, for any $m, n \in \mathbb{N} = \mathrm{Ob}(\mathsf{P})$ (hence the unit 1 equals 0).

Recall that the $\mathtt{Mod}_{\mathbf{k}}$-enrichment in the above definition means that each hom-set $Mor_{\mathsf{P}}(m, n)$ is a $\mathbf{k}$-module and the operations of the monoidal category P (the composition $\circ$, the product $\odot$ and the symmetry S) are compatible with this $\mathbf{k}$-linear structure.

For a PROP P denote $\mathsf{P}(m,n) := Mor_{\mathsf{P}}(m,n)$. The symmetry S induces, via the canonical identifications $m \cong 1^{\odot m}$ and $n \cong 1^{\odot n}$, on each $\mathsf{P}(m,n)$ a structure of (Σ_m, Σ_n)-bimodule (left Σ_m- right Σ_n-module such that the left action commutes with the right one). Therefore a PROP is a collection $\mathsf{P} = \{\mathsf{P}(m,n)\}_{m,n\geq 0}$ of (Σ_m, Σ_n)-bimodules, together with two types of compositions, *horizontal*

$$\otimes : \mathsf{P}(m_1,n_1) \otimes \cdots \otimes \mathsf{P}(m_s,n_s) \to \mathsf{P}(m_1 + \cdots + m_s, n_1 + \cdots + n_s),$$

induced, for all $m_1, \ldots, m_s, n_1, \ldots, n_s \geq 0$, by the monoidal product $\odot$ of P, and *vertical*

$$\circ : \mathsf{P}(m,n) \otimes \mathsf{P}(n,k) \to \mathsf{P}(m,k),$$

given, for all $m,n,k \geq 0$, by the categorial composition. The monoidal unit is an element $e := 1 \in \mathsf{P}(1,1)$. In Definition 10.1, $\mathtt{Mod}_{\mathbf{k}}$ can be replaced by an arbitrary symmetric strict monoidal category.

Let $\mathsf{P} = \{\mathsf{P}(m,n)\}_{m,n\geq 0}$ and $\mathsf{Q} = \{\mathsf{Q}(m,n)\}_{m,n\geq 0}$ be two PROPs. A *homomorphism* $f : \mathsf{P} \to \mathsf{Q}$ is a sequence $f = \{f(m,n) : \mathsf{P}(m,n) \to \mathsf{Q}(m,n)\}_{m,n\geq 0}$ of bi-equivariant maps which commute with both the vertical and horizontal compositions. An *ideal* in a PROP P is a system $\mathsf{I} = \{\mathsf{I}(m,n)\}_{m,n\geq 0}$ of left Σ_m- right Σ_n-invariant subspaces $\mathsf{I}(m,n) \subset \mathsf{P}(m,n)$ which is closed, in the obvious sense, under both the vertical and horizontal compositions. Kernels, images, etc., of homomorphisms between PROPs, as well as quotients of PROPs by PROPic ideals, are defined componentwise, see [35–38] for details.

EXAMPLE 10.2 *The* ***endomorphism*** *PROP of a* $\mathbf{k}$*-module* V *is the system*

$$\mathcal{E}nd_V = \{\mathcal{E}nd_V(m,n)\}_{m,n\geq 0}$$

with $\mathcal{E}nd_V(m,n)$ *the space of linear maps* $Lin(V^{\otimes n}, V^{\otimes m})$ *with* n *'inputs' and* m *'outputs,'* $e \in \mathcal{E}nd_V(1,1)$ *the identity map, horizontal composition given by the tensor product of linear maps, and vertical composition by the ordinary composition of linear maps.*

Also algebras over PROPs can be introduced in a very concise way:

DEFINITION 10.2 *A* P-***algebra*** *is a strict symmetric monoidal functor* $\lambda : \mathsf{P} \to \mathtt{Mod}_{\mathbf{k}}$ *of enriched monoidal categories. The value* $\lambda(1)$ *is the* underlying space *of the algebra* ρ.

It is easy to see that a P-algebra is the same as a PROP homomorphism $\rho : \mathsf{P} \to \mathcal{E}nd_V$. As in Proposition 5.1, a P-algebra is determined by a system

$$\alpha : \mathsf{P}(m,n) \otimes V^{\otimes n} \to V^{\otimes m},\ m,n, \geq 0,$$

of linear maps satisfying appropriate axioms.

As before, the first step in formulating an unbiased definition of PROPs is to specify their underlying category. A Σ-*bimodule* is a system $E = \{E(m,n)\}_{m,n\geq 0}$ such that each $E(m,n)$ is a left $\mathbf{k}[\Sigma_m]$- right $\mathbf{k}[\Sigma_n]$-bimodule. Let Σ-$\mathtt{bimod}$ denote the category of Σ-bimodules. For $E \in \Sigma$-$\mathtt{bimod}$ and finite sets Y, X with m resp. n elements put

$$E(Y,X) := \mathit{Bij}(Y,[m]) \times_{\Sigma_m} E(m,n) \times_{\Sigma_n} \mathit{Bij}([n],X),\ m,n \geq 0,$$

where $\mathit{Bij}(-,-)$ is the same as in (43). Pasting schemes for PROPs are *directed* (m,n)*-graphs*, by which we mean finite, not necessary connected, graphs in the sense of Definition 9.1 such that

(i) each edge is equipped with a direction

(ii) there are no directed cycles and

(iii) the set of legs is divided into the set of inputs labeled by $\{1,\ldots,n\}$ and the set of outputs labeled by $\{1,\ldots,m\}$.

An example of a directed graph is given in Figure 21. We denote by $\mathtt{Gr}(m,n)$ the category of directed (m,n)-graphs and their isomorphisms. The direction of edges determines at each vertex $v \in \mathit{vert}(G)$ of a directed graph G a disjoint decomposition

$$\mathit{edge}(v) = \mathit{in}(v) \sqcup \mathit{out}(v)$$

of the set of edges adjacent to v into the set $\mathit{in}(v)$ of incoming edges and the set $\mathit{out}(v)$ of outgoing edges. The pair $(\#(\mathit{out}(v)), \#(\mathit{in}(v))) \in \mathbb{N} \times \mathbb{N}$ is called the *biarity* of v. To incorporate the unit, we need to extend the category $\mathtt{Gr}(m,n)$, for $m = n$, into the category $\mathtt{UGr}(m,n)$ by allowing the exceptional graph

$$\uparrow\uparrow\uparrow\ \cdots \uparrow \in \mathtt{UGr}(n,n),\ n \geq 1,$$

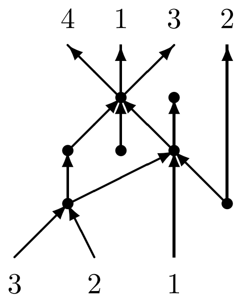


Figure 21: A directed graph from $\mathtt{Gr}(4,3)$.

with n inputs, n outputs and no vertices. For a graph $G \in \mathtt{UGr}(m,n)$ and a Σ-bimodule E, let

$$E(G) := \bigotimes_{v \in vert(G)} E(out(v), in(v)).$$

and

$$\Gamma_{\mathtt{P}}(E)(m,n) := \operatorname*{colim}_{G \,\in\, \mathtt{UGr}(m,n)} E(G),\ m,n \geq 0. \tag{64}$$

The Σ-bimodule $\Gamma_{\mathtt{P}}(E)$ is a PROP, with the vertical composition given by the disjoint union of graphs, the horizontal composition by grafting the legs, and the unit the exceptional graph $\uparrow \in \Gamma_{\mathtt{P}}(E)(1,1)$. The following proposition follows from [39] and [36–38]:

PROPOSITION 10.1 *The PROP $\Gamma_{\mathtt{P}}(E)$ is the* free PROP *generated by the Σ-bimodule E.*

As in the previous Sections, (64) defines a triple $\Gamma_{\mathtt{P}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$ with the triple multiplication of erasing the braces. According to general principles [29], Proposition 10.1 is almost equivalent to

PROPOSITION 10.2 *PROPs are algebras over the triple $\Gamma_{\mathtt{P}}$.*

One may obviously consider *non-unital PROPs* defined as algebras over the triple

$$\Psi_{\mathtt{P}}(E)(m,n) := \operatorname*{colim}_{G \,\in\, \mathtt{Gr}(m,n)} E(G),\ m,n \geq 0,$$

and develop a theory parallel to the theory of non-unital operads reviewed in Section 4.

EXAMPLE 10.3 *We will use the graphical language explained in Example 6.3. Let* $\Gamma(\curlywedge, \mathsf{Y})$ *be the free PROP generated by one operation* $\curlywedge$ *of biarity* $(1,2)$ *and one operation* Y *of biarity* $(2,1)$. *As we noticed already in [35, 40], the PROP* B *describing bialgebras equals*

$$\mathsf{B} = \Gamma(\curlywedge, \mathsf{Y})/\mathsf{I}_{\mathsf{B}},$$

where I_{B} *is the PROPic ideal generated by*

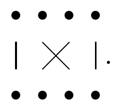
$$\curlywedge\!\!{}_{\curlywedge} - {}_{\curlywedge}\!\!\curlywedge,\ \ \mathsf{Y}\!\!{}^{\mathsf{Y}} - {}^{\mathsf{Y}}\!\!\mathsf{Y}\ \ \textit{and}\ \ \mathsf{X} - \bowtie. \tag{65}$$

In the above display we denoted

$$\curlywedge\!\!{}_{\curlywedge} := \curlywedge \circ (\curlywedge \otimes e),\ \ {}_{\curlywedge}\!\!\curlywedge := \curlywedge \circ (e \otimes \curlywedge),\ \ \mathsf{Y}\!\!{}^{\mathsf{Y}} := (\mathsf{Y} \otimes e) \circ \mathsf{Y},\ \ {}^{\mathsf{Y}}\!\!\mathsf{Y} := (e \otimes \mathsf{Y}) \circ \mathsf{Y},$$

$$\mathsf{X} := \mathsf{Y} \circ \curlywedge \ \ \textit{and}\ \ \bowtie := (\curlywedge \otimes \curlywedge) \circ \kappa \circ (\mathsf{Y} \otimes \mathsf{Y}),$$

where $\kappa \in \Sigma_4$ *is the permutation*

$$\kappa := \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} =$$  (66)

The above description of B *is 'tautological,' but B. Enriquez and P. Etingof found in [?] the following basis of the* $\mathbf{k}$*-linear space* $\mathsf{B}(m,n)$ *for arbitrary* $m, n \geq 1$. *Let* $\curlywedge \in \mathsf{B}(1,2)$ *be the equivalence class, in* $\mathsf{B} = \Gamma(\curlywedge, \mathsf{Y})/\mathsf{I}_{\mathsf{B}}$, *of the generator* $\curlywedge \in \Gamma(\mathsf{Y}, \curlywedge)(1,2)$ *(we use the same symbol both for a generator and its equivalence class). Define* $\curlywedge^{[1]} := e \in \mathsf{B}(1,1)$ *and, for* $a \geq 2$, *let*

$$\curlywedge^{[a]} := \curlywedge \circ (\curlywedge \otimes e) \circ (\curlywedge \otimes e^{\otimes 2}) \circ \cdots \circ (\curlywedge \otimes e^{\otimes(a-2)}) \in \mathsf{B}(1,a).$$

Let $\mathsf{Y}_{[b]} \in \mathsf{B}(b,1)$ *has the obvious similar meaning. The elements*

$$(\curlywedge^{[a_1]} \otimes \cdots \otimes \curlywedge^{[a_m]}) \circ \sigma \circ (\mathsf{Y}_{[b^1]} \otimes \cdots \otimes \mathsf{Y}_{[b^n]}), \tag{67}$$

where $\sigma \in \Sigma_N$ *for some* $N \geq 1$, *and* $a_1 + \cdots + a_m = b^1 + \cdots + b^m = N$, *form a* $\mathbf{k}$*-linear basis of* $\mathsf{B}(m,n)$. *This result can also be found in [41]. See also [42, 43] for the bialgebra PROP viewed from a different perspective.*

EXAMPLE 10.4 *Each operad $\mathcal{P}$ generates a unique PROP* P *such that* $\mathsf{P}(1,n) = \mathcal{P}(n)$ *for each* $n \geq 0$*. The components of such a PROP are given by*

$$\mathsf{P}(m,n) = \bigoplus_{r_1+\cdots+r_k=n} [\mathcal{P}(1,r_1) \otimes \cdots \otimes \mathcal{P}(1,r_k)] \times_{\Sigma_{r_1}\times\cdots\times\Sigma_{r_k}} \Sigma_n,$$

for each $m,n \geq 0$*. The (topological) PROPs considered in [23] are all of this type. On the other hand, Example 10.3 shows that not each PROP is of this form. A PROP* P *is generated by an operad if and only if it has a presentation* $\mathsf{P} = \Gamma_{\mathtt{P}}(E)/(R)$*, where* E *is a* Σ*-bimodule such that* $E(m,n) = 0$ *for* $m \neq 1$ *and* R *is generated by elements in* $\Gamma_{\mathtt{P}}(E)(1,n)$*,* $n \geq 0$*.*

11 Properads, Dioperads and $\frac{1}{2}$PROPs

As we saw in Proposition 6.1, under some mild assumptions, the components of free operads are finite-dimensional. In contrast, PROPs are huge objects. For example, the component $\Gamma_{\mathtt{P}}(\curlywedge, \curlyvee)(m,n)$ of the free PROP $\Gamma_{\mathtt{P}}(\curlywedge, \curlyvee)$ used in the definition of the bialgebra PROP B in Example 10.3 is infinite-dimensional for each $m,n \geq 1$, and also the components of the bialgebra PROP B itself are infinite-dimensional, as follows from the fact that the Enriquez-Etingof basis (67) of $\mathsf{B}(m,n)$ has, for $m,n \geq 1$, infinitely many elements.

To handle this combinatorial explosion of PROPs combined with lack of suitable filtrations, smaller versions of PROPs were invented. Let us begin with the simplest modification which we use as an example which explains the general scheme of modifying PROPs. Denote $\mathtt{UGr}_c(m,n)$ the full subcategory of $\mathtt{UGr}(m,n)$ consisting of *connected* graphs and consider the triple defined by

$$\Gamma_{\mathtt{c}}(E)(m,n) := \operatorname*{colim}_{G \in \mathtt{UGr}_c(m,n)} E(G), \ m,n \geq 0, \tag{68}$$

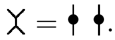
for $E \in \Sigma$-$\mathtt{bimod}$. The following notion was introduced by B. Vallette [36–38].

DEFINITION 11.1 ***Properads*** *are algebras over the triple* $\Gamma_{\mathtt{c}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$.

A properad is therefore a Σ-bimodule with operations that determine coherent contractions along connected graphs. A biased definition of properads is given in [36–38]. Since Γ_c is a subtriple of Γ_{P}, each PROP is automatically also a properad. Therefore one may speak about the *endomorphism properad* $\mathcal{E}nd_V$ and define *algebras* over a properad P as properad homomorphisms $\rho : P \to \mathcal{E}nd_V$. Algebras over other versions of PROPs recalled below can be defined in a similar way.

EXAMPLE 11.1 *Associative bialgebras reviewed in Example 10.3 are algebras over the properad B defined (tautologically) as the quotient of the free properad $\Gamma_{\mathsf{c}}(⅄, \mathsf{Y})$ by the properadic ideal generated by the elements listed in (65). We leave as an exercise to describe the sub-basis of (67) that span $B(m,n)$, $m,n \geq 1$.*

The following slightly artifical structure exists over PROPs but not over properads. It consists of a 'multiplication' $\mu = ⅄ : V \otimes V \to V$, a 'comultiplication' $\Delta = \mathsf{Y} : V \to V \otimes V$ and a linear map $f = ⍿ : V \to V$ satisfying $\Delta \circ \mu = f \otimes f$ or, diagrammatically



This structure cannot be a properad algebra because the graph on the right hand side of the above display is not connected.

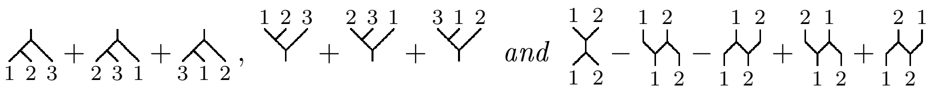
Properads are still huge objects. The first really small version of PROPs were dioperads introduced in 2003 by W.L. Gan [44]. As a motivation for his definition, consider the following:

EXAMPLE 11.2 *A* ***Lie bialgebra*** *is a vector space V with a Lie algebra structure $[-,-] = ⅄ : V \otimes V \to V$ and a Lie diagonal $\delta = \mathsf{Y} : V \to V \otimes V$. We assume that $[-,-]$ and δ are related by*

$$\delta[a,b] = \sum \left([a_{(1)}, b] \otimes a_{(2)} + [a, b_{(1)}] \otimes b_{(2)} + a_{(1)} \otimes [a_{(2)}, b] + b_{(1)} \otimes [a, b_{(2)}]\right) \tag{69}$$

for any $a, b \in V$, with the Sweedler notation $\delta a = \sum a_{(1)} \otimes a_{(2)}$ and $\delta b = \sum b_{(1)} \otimes b_{(2)}$.

Lie bialgebras are governed by the PROP $\mathsf{LieB} = \Gamma(⅄, \mathsf{Y})/\mathsf{I}_{\mathsf{LieB}}$, where ⅄ and Y are now antisymmetric *and $\mathsf{I}_{\mathsf{LieB}}$ denotes the ideal generated by*



(70)

with labels indicating the corresponding permutations of the inputs and outputs.

We observe that all graphs in (70) are not only connected as demanded for properads, but also simply-connected. This suggests considering the full subcategory $\mathtt{UGr_D}(m,n)$ of $\mathtt{UGr}(m,n)$ consisting of *connected simply-connected* graphs and the related triple

$$\Gamma_{\mathtt{D}}(E)(m,n) := \operatorname*{colim}_{G\,\in\,\mathtt{UGr_D}(m,n)} E(G),\ \ m,n \geq 0. \tag{71}$$

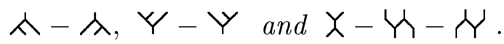
Definition 11.2 Dioperads *are algebras over the triple* $\Gamma_{\mathtt{D}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$.

A biased definition of dioperads can be found in [44]. As observed by T. Leinster, dioperads are more or less equivalent to polycategories, in the sense of [45], with one object. Lie bialgebras reviewed in Example 11.2 are algebras over a dioperad. Another important class of dioperad algebras is recalled in:

Example 11.3 *An* ***infinitesimal bialgebra*** *[46] (called in [47]* ***a mock bialgebra****) is a vector space V with an associative multiplication $\cdot : V \otimes V \to V$ and a coassociative comultiplication $\Delta : V \to V \otimes V$ such that*

$$\Delta(a \cdot b) = \sum \left(a_{(1)} \otimes a_{(2)} \cdot b + a \cdot b_{(1)} \otimes b_{(2)}\right)$$

for any $a, b \in V$. It is easy to see that the axioms of infinitesimal bialgebras are encoded by the following simply connected graphs:



Observe that associative bialgebras recalled in Example 10.1 cannot be defined over dioperads, because the rightmost graph in (65) is not simply connected. The following proposition, which should be compared to Proposition 6.1, shows that dioperads are of the same size as operads.

Proposition 11.1 *Let $E = \{E(m,n)\}_{m,n\geq 0}$ be a Σ-bimodule such that*

$$E(m,n) = 0 \text{ for } m+n \leq 2 \tag{72}$$

and that $E(m,n)$ is finite-dimensional for all remaining m,n. Then the components $\Gamma_{\mathtt{D}}(E)(m,n)$ of the free dioperad $\Gamma_{\mathtt{D}}(E)$ are finite-dimensional, for all $m,n \geq 0$.

The proof, similar to the proof of Proposition 6.1, is based on the observation that the assumption (72) reduces the colimit (71) to a summation over reduced trees (trees whose all vertices have at least three adjacent edges).

An important problem arising in connection with deformation quantization is to find a reasonably small, explicit cofibrant resolution of the bialgebra PROP B. Here by a resolution we mean a differential graded PROP R together with a homomorphism $\beta : \mathsf{R} \to \mathsf{B}$ inducing a homology isomorphism. Cofibrant in this context means that R is of the form $(\Gamma_{\mathtt{P}}(E), \partial)$, where the generating Σ-bimodule E decomposes as $E = \bigoplus_{n \geq 0} E_n$ and the differential decreases the filtration, that is

$$\partial(E_n) \subset \Gamma_{\mathtt{P}}(E)_{<n}, \text{ for each } n \geq 0,$$

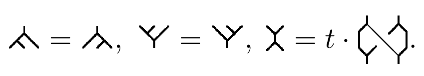
where $\Gamma_{\mathtt{P}}(E)_{<n}$ denotes the sub-PROP of $\Gamma(E)$ generated by $\bigoplus_{j<n} E_j$. This notion is an PROPic analog of the Koszul-Sullivan algebra in rational homotopy theory [48]. Several papers devoted to finding R appeared recently [41, 49–54]. The approach of [55] is based on the observation that B is a deformation, in the sense explained below, of the PROP describing structures recalled in the following:

Definition 11.3 *A* ***half − bialgebra*** *or simply a* $\frac{1}{2}$***bialgebra*** *is a vector space V with an associative multiplication $\mu : V \otimes V \to V$ and a coassociative comultiplication $\Delta : V \to V \otimes V$ that satisfy*

$$\Delta(u \cdot v) = 0, \quad \textit{for each } u, v \in V. \tag{73}$$

We chose this strange name because (73) is indeed one half of the compatibility relation (63) of associative bialgebras. $\frac{1}{2}$bialgebras are algebras over the PROP

Now define, for a formal variable t, B_t to be the quotient of the free PROP $\Gamma(\curlywedge, \curlyvee)$ by the ideal generated by



Thus B_t is a one-parametric family of PROPs with the property that $\mathsf{B}_0 = \frac{1}{2}\mathsf{B}$. At a generic t, B_t is isomorphic to the bialgebra PROP B. In other words, the PROP

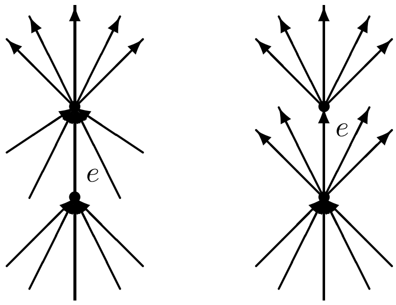


Figure 22: Edges allowed in a $\frac{1}{2}$graph.

for bialgebras is a deformation of the PROP for $\frac{1}{2}$bialgebras. According to general principles of homological perturbation theory [56], one may try to construct the resolution R as a perturbation of a cofibrant resolution $\frac{1}{2}\mathsf{R}$ of the PROP $\frac{1}{2}\mathsf{B}$. Since $\frac{1}{2}\mathsf{B}$ is simpler that B, one may expect that resolving $\frac{1}{2}\mathsf{B}$ would be a simpler task than resolving B.

For instance, one may realize that $\frac{1}{2}$bialgebras are algebras over a dioperad $\frac{1}{2}\mathtt{B}$, use [44] to construct a resolution $\frac{1}{2}\mathtt{R}$ of the dioperad $\frac{1}{2}\mathtt{B}$, and then take $\frac{1}{2}\mathsf{R}$ to be the PROP generated by $\frac{1}{2}\mathtt{R}$. More precisely, one denotes

$$F_1 : \mathtt{diOp} \to \mathtt{PROP} \tag{74}$$

the left adjoint to the forgetful functor $\mathtt{PROP} \xrightarrow{\square_1} \mathtt{diOp}$ and defines $\frac{1}{2}\mathsf{R} := F_1(\frac{1}{2}\mathtt{R})$.

The problem is that we do not know whether the functor F_1 is exact, so it is not clear if $\frac{1}{2}\mathsf{R}$ constructed in this way is really a resolution of $\frac{1}{2}\mathsf{B}$. To get around this subtlety, M. Kontsevich observed that $\frac{1}{2}$bialgebras live over a version of PROPs which is smaller than dioperads. It can be defined as follows.

Let an (m,n)-$\frac{1}{2}$*graph* be a connected simply-connected directed (m,n)-graph whose each edge e has the following property: either e is the unique outgoing edge of its initial vertex or e is the unique incoming edge of its terminal vertex, see Figure 22. An example of an (m,n)-$\frac{1}{2}$graph is given in Figure 23. Let $\mathtt{Gr}_{\frac{1}{2}}(m,n)$ be the category of (m,n)-$\frac{1}{2}$graphs and their isomorphisms. Define a triple $\Gamma_{\frac{1}{2}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$ by

$$\Gamma_{\frac{1}{2}}(E)(m,n) := \operatorname*{colim}_{G \in \mathtt{Gr}_{\frac{1}{2}}(m,n)} E(G), \ \ m,n \geq 0. \tag{75}$$

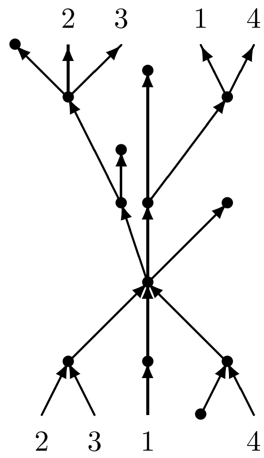


Figure 23: A graph from $\mathtt{Gr}_{\frac{1}{2}}(4,4)$.

DEFINITION 11.4 *A* $\frac{1}{2}$PROP *(called a* ***meager*** **PROP***) is an algebra over the triple* $\Gamma_{\frac{1}{2}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$.

A biased definition of $\frac{1}{2}$PROPs can be found in [39, 55]. We followed the convention that $\frac{1}{2}$PROPs do not have units; the unital version of $\frac{1}{2}$PROPs can be defined in an obvious way, compare also the remarks in [55].

EXAMPLE 11.4 $\frac{1}{2}$*bialgebras are algebras over a* $\frac{1}{2}$PROP *which we denote* $\frac{1}{2}\mathtt{b}$. *Another example of structures that can be defined over* $\frac{1}{2}$PROP*s are* Lie $\frac{1}{2}$bialgebras *consisting of a Lie algebra bracket* $[-,-] : V \otimes V \to V$ *and a Lie diagonal* $\delta : V \to V \otimes V$ *satisfying one-half of (69):*

$$\delta[a, b] = 0.$$

Let us denote by

$$F : \tfrac{1}{2}\mathtt{PROP} \to \mathtt{PROP}$$

the left adjoint to the forgetful functor $\mathtt{PROP} \xrightarrow{\Box} \frac{1}{2}\mathtt{PROP}$ from the category of PROPs to the category of $\frac{1}{2}$PROPs. M. Kontsevich observed that, in contrast to $F_1 : \mathtt{diOp} \to \mathtt{PROP}$ in (74), F is a *polynomial* functor, which immediately implies the following important theorem [39].

THEOREM 11.1 *The functor $F : \frac{1}{2}\mathtt{PROP} \to \mathtt{PROP}$ is exact.*

Now one may take a resolution $\frac{1}{2}\mathsf{r}$ of the $\frac{1}{2}$PROP $\frac{1}{2}\mathsf{b}$ and put $\frac{1}{2}\mathsf{R} := F(\frac{1}{2}\mathsf{r})$. Theorem 11.1 guarantees that $\frac{1}{2}\mathsf{R}$ defined in this way is indeed a resolution of the PROP $\frac{1}{2}\mathsf{B}$. Let us mention that there are also other structures invented to study resolutions of the PROP B, as $\frac{2}{3}$PROPs of Shoikhet [52], matrons of Saneblidze and Umble [49], or special PROPs considered in [55].

The constructions reviewed in this Section can be organized into the following chain of inclusions of full subcategories:

$$\mathtt{Oper} \subset \tfrac{1}{2}\mathtt{PROP} \subset \mathtt{diOp} \subset \mathtt{Proper} \subset \mathtt{PROP}.$$

The general scheme behind all these constructions is the following. We start by choosing a subgroupoid $\mathtt{SGr} = \coprod_{m,n\geq 0} \mathtt{SGr}(m,n)$ of $\mathtt{Gr} := \coprod_{m,n\geq 0} \mathtt{Gr}(m,n)$ (or a subgroupoid of $\mathtt{UGr} := \coprod_{\mathtt{m},\mathtt{n}\geq 0} \mathtt{UGr}(\mathtt{m},\mathtt{n})$ if we want units). Then we define a functor $\Gamma_{\mathtt{S}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$ by

$$\Gamma_{\mathtt{S}}(E)(m,n) := \operatorname*{colim}_{G\,\in\,\mathtt{SGr}(m,n)} E(G),\ \ m,n \geq 0.$$

It is easy to see that $\Gamma_{\mathtt{S}}$ is a subtriple of the PROP triple $\Gamma_{\mathtt{P}}$ if and only if the following two conditions are satisfied:

(i) the groupoid $\mathtt{SGr}$ is *hereditary* in the sense that, given a graph from $\mathtt{SGr}$ with vertices decorated by graphs from $\mathtt{SGr}$, then the graph obtained by 'forgetting the braces' again belongs to $\mathtt{SGr}$, and

(ii) $\mathtt{SGr}$ contains all directed corollas.

Hereditarity (i) is necessary for $\Gamma_{\mathtt{S}}$ to be closed under the triple multiplication of $\Gamma_{\mathtt{P}}$ while (ii) guarantees that $\Gamma_{\mathtt{S}}$ has an unit. Plainly, all the three choices used above – $\mathtt{UGr}_c$, $\mathtt{UGr}_{\mathtt{D}}$ and $\mathtt{Gr}_{\frac{1}{2}}$ – satisfy the above assumptions. Let us mention that one may modify the definition of PROPs also by *enlarging* the category $\mathtt{Gr}(m,n)$, as was done for *wheeled PROPs* in [57]. Pasting schemes and the corresponding structures reviewed in this article are listed in Figure 24.

Pasting schemes	**corresponding structures**
rooted trees	non-unital operads
May's trees	non-unital May's operads
extended rooted trees	operads
cyclic trees	non-unital cyclic operads
extended cyclic trees	cyclic operads
stable labeled graphs	modular operads
extended directed graphs	PROPs
extended connected directed graphs	properads
extended connected 1-connected dir. graphs	dioperads
$\frac{1}{2}$graphs	$\frac{1}{2}$PROPs

Figure 24: Pasting schemes and the structures they define.

12 Sums over Trees

In this Section we prove basic formal identities for certain infinite sums (partition functions) taken over graphs of various topological types. The simplest "Euler product" identity relates sums over not necessarily connected graphs to those over connected ones. Summation over trees is interpreted as a calculation of the critical value of a formal potential. Finally, summation over graphs of arbitrary topology is interpreted as the perturbation series for a formal Feynman integral.

12.1 Application to sums over graphs

Definition 12.1 *Let $(\mathcal{E}, \circ)$ be a symmetric monoidal category with the identity object* **1** *satisfying the following conditions.*

a) $\mathcal{E}$ has a countable set of isomorphism classes of objects. Every object has a finite automorphism group.

b) Every object of $\mathcal{E}$ is isomorphic to a product $\bigcirc_i \pi_i^{a_i}$ where π_i are indecomposable with respect to $\circ$ ("primes"), $\pi_i^{a_i}$ is the $\circ$-product of a_i copies of π_i, and

$\pi_i \neq \pi_j$ for $i \neq j$. This product is defined uniquely up to permutation of factors.

c) We have

$$|\mathrm{Aut}\ \bigcirc_i\, \pi_i^{a_i}| = \prod_i a_i!|\mathrm{Aut}\ \pi_i|^{a_i}\ , \tag{76}$$

in particular, $|\mathrm{Aut}\,(\mathbf{1})| = 1$.

In addition let R be a commutative topological ring and let $w : \mathrm{Ob}\,\mathcal{E} \to R$ be a weight function depending only on the isomorphism class of the object and multiplicative: $w(\sigma \circ \tau) = w(\sigma)w(\tau)$.

THEOREM 12.1 *If the sums and products involved absolutely converge, we have*

$$\prod_{\{\pi\}/(iso)} \exp\frac{w(\pi)}{|\mathrm{Aut}\,\pi|} = \sum_{\sigma\in\mathrm{Ob}\mathcal{E}/(iso)} \frac{w(\sigma)}{|\mathrm{Aut}\,\sigma|} = \exp\left(\sum_{\{\pi\}/(iso)} \frac{w(\pi)}{|\mathrm{Aut}\,\pi|}\right)\ . \tag{77}$$

Proof. We have

$$\prod_{\{\pi\}/(iso)} \exp\frac{w(\pi)}{|\mathrm{Aut}\,\pi|} = \prod_{\{\pi\}/(iso)} \sum_{a=0}^{\infty} \frac{w(\pi)^a}{a!|\mathrm{Aut}\,\pi|^a}\ ,$$

and it remains to apply (76). ■

Throughout this Section, we will take for $\mathcal{E}$ various categories of finite graphs, $\circ$ will denote the disjoint sum, and "primes" π will be connected graphs. Property (76) will be evident from the definition of isomorphisms. The second equality in (77) says that a weighted sum taken over all graphs can be obtained by exponentiation from the similar sum taken only over connected graphs.

We will now introduce a family of weights which will be called standard.

DEFINITION 12.2 *A* ***standard weight*** *on a category of finite graphs is defined by the following choices:*

a) A set of "colors" A, finite or countable.

b) A family of symmetric tensors $C_{a_1,\dots,a_k}$, $k = 1, 2, \dots$, whose subscripts belong to A and coordinates belong to a topological commutative ring R.

c) A symmetric tensor g^{ab} with the same properties. The matrix (g^{ab}) must be invertible, and we put $(g_{ab}) = (g^{ab})^{-1}$.

In other words, we have a free R-module H with metric and a sequence of symmetric polynomials of all degrees on H expressed in terms of a basis indexed by A. (We can generalize this setting considering supercommutative R and $\mathbf{Z}_2$-graded H.)

With these choices made, we put for a graph τ:

$$w(\tau) = \sum_{u:F_\tau \to A} \prod_{e\in E_\tau} g^{u(\partial e)} \prod_{v\in V_\tau} C_{u(F_\tau(v))}\ . \tag{78}$$

REMARK 12.1 *a) The expression ∂e in (78) means the set of two flags constituting the edge e. When a marking $u : F_\tau \to A$ is given, $u(\partial e) = \{a,b\}$ consists of two elements of A which produce g^{ab}. We can similarly define $C_{u(F_\tau(v))}$ thanks to the symmetry.*

b) If A is finite, the whole sum (78) is finite. Otherwise we have to postulate convergence already at this step. In our applications R will be a formal series ring. The multiplicativity of w with respect to disjoint union is evident.

c) Consider now the sum of type (77) with a standard weight:

$$Z_{\mathcal{E}}(w) = \sum_{\tau/(iso)} \frac{1}{|\mathrm{Aut}\,\tau|} \sum_{u:F_\tau \to A} \prod_{e\in E_\tau} g^{u(\partial e)} \prod_{v\in V_\tau} C_{u(F_\tau(v))}\ . \tag{79}$$

*Such sums occur in some models of statistical and quantum physics. Coloring of flags corresponds to the picture of A types of particles propagating along the edges with amplitudes g^{ab} and interacting at vertices with amplitudes $C_{a_1,\dots,a_k}$. In this context, graphs are Feynman diagrams, and (79) can be called the **partition function**. The same formalism emerges in the general operadic context and in the topology of moduli spaces.*

12.2 Summation over trees

In this Subsection, we will calculate the partition function (78) in which the summation is taken over the set T of isomorphism classes of all (connected) trees *without tails* and having at least one edge.

We will treat here $C_{a_1,\dots,a_k}$ independent formal variables over a subring $R_0 \subset R$ containing g_{ab}, g^{ab}, and $\mathbf{Q}$. Then all our sums make sense as formal series.

We will express Z via a simpler formal function of auxiliary variables $\phi = \{\phi^a | a \in A\}$ independent over R:

$$\Phi(\phi) = -\frac{1}{2}\sum_{a,b} g_{ab}\phi^a\phi^b + \sum_{k=1}^{\infty}\frac{1}{k!}\sum_{a_1,\dots,a_k\in A} C_{a_1,\dots,a_k}\phi^{a_1}\cdots\phi^{a_k}\ . \qquad (80)$$

Put $C^a = \sum_{b\in A} g^{ab}C_b$ and denote by $N \subset R$ the ideal generated by $C_{a_1,\dots,a_k}$ for all $k \geq 2$.

THEOREM 12.2 *a) The equations*

$$\frac{\partial\Phi(\phi)}{\partial\phi^a} = 0, \quad \forall a \in A, \qquad (81)$$

admit the unique solution $\phi_0 = \{\phi_0^a\} \in R^A$ *satisfying the condition*

$$\phi_0^a \equiv C^a \operatorname{mod} N\ . \qquad (82)$$

b) The partition function $Z = Z_T$ *satisfies the differential equations*

$$\frac{\partial Z}{\partial C_a} = \phi_0^a, \quad a \in A\ , \qquad (83)$$

and is the critical value of $\Phi(\phi)$*:*

$$Z = \Phi(\phi_0)\ . \qquad (84)$$

REMARK 12.2 *The assumption that* $C_{a_1,\dots,a_k}$ *are independent formal variables is used several times in the statements and proofs: to locate the critical point* ϕ_0*, to make sense of the left-hand side of (83), etc. However, when the identities (80) and (81) are proved in the formal context, they can be specialized to other topological rings* R*.*

Proof. a) Rewrite (81) as

$$\sum_{b\in A} g_{ab}\phi^b = C_a + \sum_{k\geq 2}\frac{1}{k!}\sum_{a_1,\dots,a_k\in A}\frac{\partial}{\partial\phi^a}(C_{a_1,\dots,a_k}\phi^{a_1}\cdots\phi^{a_k})\ , \quad \forall a \in A, \qquad (85)$$

that is,

$$\phi^a = C^a + \sum_{k\geq 2} \frac{1}{k!} \sum_{a_1,\ldots,a_k,b\in A} g^{ab} \frac{\partial}{\partial \phi^a}(C_{a_1,\ldots,a_k}\phi^{a_1}\cdots\phi^{a_k}) , \quad \forall a \in A . \tag{86}$$

Comparing (82) and (86) one sees that the critical point in question can be calculated by iterating (86). More precisely, consider the formal operator T mapping $\psi = (\psi^a | a \in A)$ to $(T^a(\psi)|a \in A)$ where

$$T^a(\psi) = \sum_{k\geq 2} \frac{1}{k!} \sum_{i=1}^{k} \sum_{a_1,\ldots,a_k,b\in A} g^{ab} C_{a_1,\ldots,a_k} \psi^{a_1} \cdots \widehat{\psi}^{a_i} \cdots \psi^{a_k} \delta_{a_i,b} . \tag{87}$$

The equation (86) can be rewritten as $\phi_0 = C + T(\phi_0)$ and solved by means of a version of the geometric progression formula

$$\phi_0 = C + T(C + T(C + T(C + \ldots))) . \tag{88}$$

The solution is clearly unique.

b) In order to make more transparent the formal structure of (88) as a sum over trees, we will consider the case when $A = \{*\}$ is a one-element set.

Put $g^{**} = g, g_{**} = g^{-1}, C_{*\ldots*}$ (k subscripts) $= C_k, \phi^* = \phi$, and $C^1 = gC_1$. Then (87) becomes

$$T(\psi) = \sum_{k=1}^{\infty} \frac{gC_{k+1}}{k!} \psi^k$$

and (88) takes the form

$$\phi_0 = \sum_{k=0}^{\infty} \frac{gC_{k+1}}{k!} \left(\sum_{l=0}^{\infty} \frac{gC_{l+1}}{l!} \left(\sum_{m=0}^{\infty} \frac{gC_{m+1}}{m!} (gC_1 + \ldots)^m \right)^l \right)^k . \tag{89}$$

Opening the brackets we will represent ϕ_0 sum of monomials in $\frac{gC_{i+1}}{i!}$. We will say that such a monomial has height $\leq N$ if it is a product of terms situated before the N-th opening bracket in (89) or directly after it (the terms of the latter type are gC_1).

E.g. the only monomial of height 0 is gC_1. Monomials of height 1 are

$$\frac{gC_{k=1}}{k!}(gC_1)^k, \quad k \geq 1 .$$

Monomials of height 2 are indexed by the families of integers $\{k_i; l_1, \dots, l_k\}$, $k \geq 1, l_i \geq 0$, each such family contributing

$$\frac{C_{k+1}}{k!}\frac{C_{l_1+1}}{l_1!}\cdots\frac{C_{l_k+!}}{l_k!}(gC_1)^{l_1+\cdots+l_k} \ . \tag{90}$$

To establish the general pattern, we need a definition. Consider a tree without tails τ. *The pinning* of τ is given by the choice of the following data:

a) The choice of a vertex $v_0 \in V_\tau$ with $|F_\tau(v_0)| = 1$ called *the root.*

Such choice determines a unique orientation of all flags (or edges) of τ such that the unique flag of v_0 is incoming and every vertex $v \neq v_0$ has exactly one outgoing flag.

Such choice determines a unique orientation of all flags (or edges) of τ such that the unique flag of v_0 is incoming and very vertex $v \neq v_0$ has exactly one outgoing flag.

b) A total ordering of all sets $V_\tau(k) \subset V_\tau$ where $V_\tau(k)$ denotes the set of all vertices of τ separated by k edges from v_0. This total ordering must satisfy the following condition. Let $f_k : V_\tau(k+1) \to V_\tau(k)$ be the map "going along the outgoing edge to the next vertex". Then f_k must be monotone with respect to the chosen orderings.

A pinned tree is a tree with pinning. An isomorphism of pinned trees is an isomorphism of trees compatible with orientation and pinning. *The height* of a pinned tree is max $\{k|V_\tau(k+1) \neq \emptyset$.

A contemplation shows that there is a natural bijection between the isomorphism classes of pinned trees (τ, p) with $|E_\tau| \geq 1$ of height $\leq N$ and monomials of height $\leq N$ which can be directly obtained from (89). Moreover, various pinnings of the same τ generate the differently ordered but equal monomials which can be written in the form dependent only on τ:

$$\frac{1}{C_1} g^{|E_\tau|} \prod_{v \in V_\tau} C_{|v|}/(|v|-1)! \ . \tag{91}$$

Now, the number of different pinnings of τ is $|T_\tau| \prod_{v \in V_\tau}(|v|-1)!$ where T_τ is the set of potential roots and factorials count orderings of incoming edges. The automorphism group of τ effectively acts on the set of pinnings. Hence (91) appears with the coefficient

$$|T_\tau| \prod_{v \in V_\tau} (|v|-1)!/|\mathrm{Aut}\,\tau| \ .$$

We now turn to the proof of (83) which for the one-element A becomes

$$\phi_0 = \frac{\partial}{\partial C_1}\left(\sum_\tau \frac{1}{|\mathrm{Aut}\,\tau|} g^{|E_\tau|} \prod_{v\in V_\tau} C_{|v|}\right) . \tag{92}$$

In fact, the discussion above shows that the tree τ with all its pinnings contributes to ϕ_0 the term

$$\frac{|T_\tau|}{C_1}\frac{\prod_{v\in V_\tau}(|v|-1)!}{|\mathrm{Aut}\,\tau|}\, g^{|E_\tau|} \prod_{v\in V_\tau}\frac{C_{|v|}}{(|v|-1)!} .$$

In view of (79), this is the same as the contribution of τ to $\frac{\partial Z}{\partial C_1}$. This gives (83).

We leave to the reader the discussion of the case $|A| \geq 1$.

To derive (84) from (83), consider both sides of (84) as formal series in $C_a, a \geq 1$. Their constants terms (value at $(C_a) = 0$) vanish. For Z, this follows from the fact that any tree in (79) has at least two verticles with $|v| = 1$. For ϕ_0, this follows from (92). Hence it sufficient to check that $\frac{\partial}{\partial C_a} Z = \frac{\partial}{\partial C_a}\Phi(\phi_0)$ for all $a \in A$. But we have

$$\frac{\partial}{\partial C_a}\Phi(\phi_0) = \sum_{b\in A}\frac{\partial \Phi}{\partial \phi^b}(\phi_0)\frac{\partial \phi_0^b}{\partial C_a} + \frac{\partial \Phi}{\partial C_a}(\phi_0) .$$

The first sum vanishes because $(dS)(\phi_0) = 0$, and the second term equals ϕ_0^a because of (80). It remains to apply (83). ■

12.3 Summation over graphs of arbitrary topology

We will now study the partition function (79) for more general graphs, keeping the same assumptions about the coefficient ring R and tensors C, g as in Subsections 12.1 and 12.2. In order to keep track of the Euler characteristic of the graphs, we extend R to the Laurent formal series ring $R_\lambda = R((\lambda^{-1}))$.

DEFINITION 12.3 *An R_λ-linear functional*

$$\langle\cdot\rangle : R_\lambda[[\phi]] \to R_\lambda$$

*is called $\lambda^{-1}g$-**Gaussian** (mean value) if it is (λ^{-1}, ϕ)-adically continuous, and*

$$\langle \exp(\lambda^{-1}\sum_a C_a\phi^a)\rangle = \exp((2\lambda^{-1})\sum_a C_a g^{ab} C_b) . \tag{93}$$

Lemma 12.1 **(Wick)** *If (C_a) are independent variables over R_λ, then we have:*

a) $\langle \phi^{a_1} \dots \phi^{a_n} \rangle = 0$ for $n \equiv 1$ mod 2.

b) $\langle \phi^a \phi^b \rangle = \lambda g^{ab}$.

c) $\langle \phi^{a_1} \dots \phi^{a_{2m}} \rangle = \lambda^m \sum g^{a_{i_1} a_{j_1}} \dots g^{a_{i_m} a_{j_m}}$ where the summation is taken over all unordered partitions of $\{1, \dots, 2m\}$ into m unordered pairs $\{i_1, j_1\}, \dots, \{i_m, j_m\}$ (pairings).

Conversely, if a (λ^{-1}, ϕ)-adically continuous functional $\langle \cdot \rangle$ satisfies a), b), c), then it is $\lambda^{-1} g$-Gaussian.

Proof. We have

$$\langle \exp(\lambda^{-1} \sum_a C_a \phi^a) \rangle = \sum_{n=0}^{\infty} \frac{1}{\lambda^n n!} \sum_{a_1, \dots, a_n \in A} C_{a_1} \dots C_{a_n} \langle \phi^{a_1} \cdots \phi^{a_n} \rangle \ ,$$

$$\exp((2\lambda)^{-1} \sum_a C_a g^{ab} C_b) = \sum_{m=0}^{\infty} \frac{1}{2^m \lambda^m m!} \sum_{a_i, b_i \in A} C_{a_1} C_{b_1} \dots C_{a_m} C_{b_m} g^{a_1 b_1} \cdots g^{a_m b_m} \ .$$

Comparing the coefficients, we get the lemma. ■

Now put

$$\Phi_0(\phi) = -\frac{1}{2} \sum_{a,b \in A} g_{ab} \phi^a \phi^b, \quad \Phi_1(\phi) = \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{a_1, \dots, a_k \in A} C_{a_1, \dots, a_k} \phi^{a_1} \dots \phi^{a_k} \ ,$$

and denote by $w(\tau)$ the weight function (78).

Let Γ be the set of (isomorphism classes of) all finite graphs without tails, not necessarily connected, including the empty graph, and Γ_0 the subset of connected non-empty graphs. Let $\{\cdot\}$ be the $\lambda^{-1} g$-Gaussian mean value. Denote by $\chi(\tau)$ the Euler characteristic of $||\tau||$.

Theorem 12.3 *We have*

$$\sum_{\tau \in \Gamma} \frac{\lambda^{\chi(\tau)}}{|\mathrm{Aut}\, \tau|} w(\tau) = \langle \exp(\lambda^{-1} \Psi_1(\phi)) \rangle \ , \tag{94}$$

$$\sum_{\tau \in \Gamma_0} \frac{\lambda^{\chi(\tau)}}{|\mathrm{Aut}\, \tau|} w(\tau) = \log \langle \exp(\lambda^{-1} \Psi_1(\phi)) \rangle \ . \tag{95}$$

Proof. The second equality follows from the first one in view of (77) and the additivity of the Euler characteristic with relation to the disjoint union.

Let us now calculate the right-hand side of (94). By definition, it is

$$\left\langle \sum_{n=0}^{\infty} \lambda^{-n} \frac{1}{n!} \left(\sum_{k=1}^{\infty} \frac{1}{k!} \sum_{a_1,\dots,a_k \in A} C_{a_1,\dots,a_k} \phi^{a_1} \dots \phi^{a_k} \right)^n \right\rangle . \tag{96}$$

Choose some $(n; k_1, \dots, k_n)$. A typical monomial in the decomposition of (96) will be

$$\lambda^{-1} \frac{1}{n!} \prod_{i=1}^{n} \frac{1}{k_i!} C_{a_1^{(i)},\dots,a_{k_i}^{(i)}} \langle \prod_{i=1}^{n} \phi^{a_1^{(i)}} \dots \phi^{a_{k_i}^{(i)}} \rangle . \tag{97}$$

It vanishes if $k_1 + \dots + k_n$ is odd. Otherwise, in view of Wick's Lemma (97) can be rewritten as

$$\lambda^{-n+\frac{1}{2}\sum k_i} \frac{1}{n!} \prod_{i=1}^{n} \frac{1}{k_i!} C_{a_1^{(i)},\dots,a_{k_i}^{(i)}} \left(\sum g^{a_{l_1}^{(i_1)} a_{m_1}^{(j_1)}} \cdots g^{a_{l_r}^{(i_r)} a_{m_r}^{(j_r)}} \right) , \tag{98}$$

where $r = \frac{1}{2} \sum k_i$ and the inner sum is taken over all pairings of the set of ordered pairs $F = \bigcup_{i=1}^{n} \{(i,1), \dots, (i, k_i)\}$.

Construct the family of graphs τ whose set of flags is $F_\tau := F, V_\tau = \{1, \dots, n\}$, $\partial_\tau(i, l) = i$, and involutions bijectively correspond to various pairings in (98). If we color the flags of one such graph by the map $F_\tau \to A : (i, l) \mapsto a_l^{(i)}$, then the sum over all pairings will produce the same monomials as in (79). It remains to do the accurate bookkeeping in order to identify the coefficients.

The graphs constructed above bijectively correspond to all elements of Γ. In fact, a choice of $(n; k_1, \dots, k_n)$ determines the number of vertices of any valence, and the choice of a pairing determines which pairs of flags become edges ($n = 0$ produces the empty graph). Moreover, a non-empty graph comes thus equipped with a total ordering of its vertices and all sets of flags belonging to one vertex. The sum over graphs does not take care of these orderings. The group $\mathrm{Aut}\,\tau$ effectively acts on the whole set of them consisting of $n! \prod_{i=1}^{n} k_i!$ elements. Summing over isomorphism classes, we may replace the numerical coefficient in (98) by $|\mathrm{Aut}\,\tau|^{-1}$.

Finally,

$$-n + \frac{1}{2} \sum_{i=1}^{n} k_i = -|V_\tau| + |E_\tau| = \chi(\tau) .$$

■

13 Generating Functions

In this Section we calculate several generating functions related to moduli spaces and quantum cohomology, first representing them as sums over trees of the type treated in Section 12.

13.1 Virtual Poincare polynomial

Let Y be an algebraic variety over $\mathbf{C}$, possibly non-smooth and non-compact. Following [58] we denote by $P_Y(q)$ the virtual Poincaré polynomial of Y which is uniquely defined by the following properties.

a) If Y is smooth and compact, then

$$P_Y(q) = \sum_j \dim H^j(Y) q^j \ . \tag{99}$$

In particular

$$\chi(Y) = P_Y(-1) \ . \tag{100}$$

b) If $Y = \coprod_i Y_i$ is a finite union of pairwise disjoint locally closed strata, then

$$P_Y(q) = \sum_i P_{Y_i}(q) \ . \tag{101}$$

c) $P_{Y\times Z}(q) = P_Y(q)P_Z(q)$. It follows that if Y is a fibration over base B with fiber F locally trivial in the Zariski topology, then $P_Y(q) = P_B(q)P_F(q)$.

A definition of $P_Y(q)$ can be given using the weight filtration on the cohomology with compact support:

$$P_Y(q) = \sum_{i,j} (-1)^{i+j} \dim\,(\mathrm{gr}_W^j H_c^i(Y, \mathbf{Q})) q^j \ . \tag{102}$$

13.2 Generating function for moduli spaces of genus zero

We put

$$\varphi(q,t) := t + \sum_{n=2}^{\infty} P_{\overline{M}_{0,n+1}}(q) \frac{t^n}{n!} \in \mathbf{Q}[q][[t]] \ , \tag{103}$$

$$\chi(t) := \varphi(-1,t) = t + \sum_{n=2}^{\infty} \chi(\overline{M}_{0,n+1}) \frac{t^n}{n!} \in \mathbf{Q}[[t]] \ . \tag{104}$$

THEOREM 13.1 *a) $\varphi(q,t)$ is the unique root in $t+t^2\mathbf{Q}[q][[t]]$ of any one of the following functional/differential equations in t with parameter q:*

$$(1+\varphi)^{q^2} = q^4\varphi - q^2(q^2-1)t + 1 \ , \tag{105}$$

$$(1+q^2t-q^2\varphi)\varphi_t = 1+\varphi \ . \tag{106}$$

b) χ is the unique root in $t+t^2\mathbf{Q}[[t]]$ of any one of the similar equations

$$(1+\chi)\log(1+\chi) = 2\chi - t \ , \tag{107}$$

$$(1+t-\chi)\chi_t = 1+\chi \ . \tag{108}$$

Equation (106) is equivalent to the following recursive formulas for the Poincaré polynomials. Put $p_n = p_n(q) = P_{\overline{M}_{0,n+1}}/n!$.

COROLLARY 13.1 *We have for $n \geq 1$:*

$$(n+1)p_{n+1} = p_n + q^2 \sum_{\substack{i+j=n+1 \\ i\geq 2}} jp_ip_j \ , \tag{109}$$

$$P_{\overline{M}_{0,n+2}}(q) = P_{\overline{M}_{0,n+1}}(q) + q^2 \sum_{\substack{i+j=n+1 \\ i\geq 2}} \binom{n}{i} P_{\overline{M}_{0,i+1}}(q)P_{\overline{M}_{0,j+1}}(q) \ . \tag{110}$$

From (108) one sees that the function inverse to χ has a critical point at $t=e-2$. From this one can derive the following asymptotical formula:

$$\chi(\overline{M}_{0,n+1}) \cong \frac{1}{\sqrt{n}}\left(\frac{n}{e^2-2e}\right)^{n-\frac{1}{2}} \ .$$

In order to prove Theorem 13.1, we will first apply the additivity formula (101) to the open boundary strata of $\overline{M}_{0,n}$ and then use Theorem 12.2. However, the classes of trees involved in the labeling of stable curves, on the one hand, and the summation formula (84), on the other, are slightly different: we need tails in the first problem and do not allow them in the second. In order to unify the combinatorial pictures, and *only in this Section*, we will eliminate tails by putting end-point vertices on them. This will lead to the following temporary modification of our conventions:

A tree without tails is called *stable* if $|v| \neq 2$ for all vertices v. If $|v| = 1$ we call v an end vertex. Let V_τ^1 be the set of end vertices. An n-marking of τ is a bijection $\mu : V_\tau^1 \to \{1, \ldots, n\}$. We also put $V_\tau^0 = V \backslash V_\tau^1$ and refer to it as the set of interior vertices.

Now let $(C; x_1, \ldots, x_n)$ be a stable compact connected curve of arithmetical genus zero with $n \geq 3$ labeled non-singular points. The combinatorial structure of this curve is described by the following stable tree with n-marking (τ, μ) : $V_\tau^0 = \{\text{irreducible components of } C\}$, $V_\tau^1 = \{x_1, \ldots, x_n\}$; $\mu : x_i \mapsto i$; an edge connects two interior vertices if the respective components of C have non-empty intersection; an edge connects an interior vertex to an end vertex if the respective point belongs to the respective component.

Denote now by $M(\tau, \mu) \subset \overline{M}_{0,n}$ the set of points parametrizing stable curves of the type (τ, μ). If τ has only one interior vertex, $M(\tau, \mu) := M_{0,n}$ is the big cell. The following statement summarizes the main properties of these sets; for a proof, see [59],

PROPOSITION 13.1 *a) $M(\tau, \mu)$ is a locally closed subset of $\overline{M}_{0,n}$ depending only on (the isomorphism class of) (τ, μ).*

b) $\overline{M}_{0,n}$ is the union of pairwise disjoint strata $M(\tau, \mu)$ for all marked stable n-trees (τ, μ).

c) For any (τ, μ),

$$M(\tau, \mu) \cong \prod_{v \in V_\tau^0} M_{0,|v|} \ . \tag{111}$$

Notice that there exists exactly one stable tree $\bullet\!\!-\!\!\!-\!\!\bullet$ which does not correspond to any stable curve.

We can now calculate Poincaré polynomials.

PROPOSITION 13.2 *We have*

$$P_{M(\tau,\mu)}(q) = \prod_{v \in V_\tau^0} P_{M_{0,|v|}}(q) \ , \tag{112}$$

$$P_{M_{0,k}}(q) = \begin{pmatrix} q^2 - 2 \\ k - 3 \end{pmatrix} (k-3)! \ . \tag{113}$$

Proof. (112) follows from (111) and the multiplicativity of Poincaré polynomials.

To prove (113), one can use the following geometric facts. First, the morphism $\pi : \overline{M}_{0,n+1} \to \overline{M}_{0,n}$ forgetting the last marked point is (canonically isomorphic to) the universal curve. Second, the boundary of the source consists of structure sections and fibers at infinity of the target. Therefore, over the big cell $M_{0,n}$ this morphism is a Zariski locally trivial fibration with fiber $\mathbf{P}^1$, and $\overline{M}_{0,n+1} = \pi^{-1}(M_{0,n})\backslash\{\textit{union of structure sections}\}$.

From the additivity of Poincaré polynomials it follows that

$$P_{M_{0,n+1}}(q) = P_{M_{0,n}}(q)P_{\mathbf{P}^1}(q) - nP_{M_{0,n}}(q) = (q^2+1-n)P_{M_{0,n}}(q) \; .$$

Since $P_{M_{0,3}}(q) = 1$. we get (113). ■

Summarizing, we have for $n \geq 3$:

$$P_{\overline{M}_{0,n}}(q)t^n = \sum_{\substack{(\tau,\mu)/(iso)\\ |V_\tau^1|=n}} \prod_{v\in V_\tau^0} \left(\begin{array}{c} q^2-2 \\ |v|-3 \end{array} \right) (|v|-3)! \prod_{v\in V_\tau^1} t \; , \tag{114}$$

where t is a new formal variable, and the sum is taken over n-marked stable trees.

We want to present (103) as a partition function. Comparing (114) to (78) and (80), we are more or less compelled to choose $A = \{*\}$ (one element set), $g^{**} = 1$, $C_* = t$, $C_{**} = 0$ (this gives weight zero to non-stable trees), and finally, denoting by C_k the component with $k \geq 3$ subscripts, we get

$$C_k = \left(\begin{array}{c} q^2-2 \\ k-3 \end{array} \right) (k-3)! \; . \tag{115}$$

In particular, we can forget about $u : F_\tau \to \{*\}$.

If $|V_\tau^1| = n$, the set of all n-markings of τ consists of $n!$ elements and is effectively acted upon by the group $\mathrm{Aut}\,\tau$. We see finally that $\psi(q,t) = Z$ where

$$\psi(q,t) := \frac{t^2}{2!} + \sum_{n\geq 3} \frac{t^n}{n!} P_{\overline{M}_{0,n}}(q) \; , \tag{116}$$

$$Z := \sum_{\tau/(iso)} \frac{1}{|\mathrm{Aut}\,\tau|} \prod_{v\in V_\tau} C_{|v|} \; . \tag{117}$$

The summation in (117) is now taken over all trees, and the term $t^2/2$ in (116) comes from the two-vertex tree.

We will now use (83) in order to calculate

$$\frac{\partial Z}{\partial t} = \frac{\partial \psi(q,t)}{\partial t} =: \varphi(q,t) \ .$$

From (80) and (115) one sees that

$$\Phi(\varphi) = -\frac{\varphi^2}{2} + t\varphi + \sum_{k\geq 3} C_k \frac{\varphi^k}{k!} = -\frac{\varphi^2}{2} + t\varphi + \sum_{k\geq 3} \left(\begin{array}{c} q^2-2 \\ k-3 \end{array} \right) \frac{\varphi^k}{k(k-1)(k-2)} \ .$$

This can easily be summed. We need only the derivative.

For generic q we have

$$\frac{\partial}{\partial \varphi}\Phi(\varphi) = \frac{(1+\varphi)^{q^2} - 1 - q^4\varphi}{q^2(q^2-1)} + t \ , \tag{118}$$

and for $q = -1$,

$$\frac{\partial}{\partial \varphi}\Phi(\varphi) = (1+\varphi)\log(1+\varphi) - 2\varphi + t \ . \tag{119}$$

(4.21)

We see now that (105), resp. (107), are equations for the critical point $d_\varphi \Phi = 0$. Differentiating them in t and eliminating $(1+\varphi)^{q^2}$, resp. $\log(1+\varphi)$, we get (106), resp. (108).

13.3 Generating function for configuration spaces

Let X be a smooth compact algebraic variety. The configuration space $X[n]$, $n \geq 2$, is defined in [58] as the closure of its big cell $X^n \backslash (\bigcup_{i<j} \Delta_{ij})$ (Δ_{ij} is the diagonal $x_i = x_j$) in $X^n \times \prod_S \widetilde{X}^S$, where S runs over subsets $S \subset \{1,\dots,n\}$, $|S| \geq 2$; X^S denotes the respective partial product of X's, and $\widetilde{X}^S$ is the blow up of the small diagonal Δ_S in X^S.

Every S determines a divisor at infinity $D(S) \subset X[n]$. Namely, let $\pi_S : X[n] \to X^S$ be the canonical projection. Then $\pi_S^{-1}(\Delta_S) = \bigcup_{T \supset S} D(T)$.

The natural stratification of $X[n]$ described in [58] consists of (open subsets of) intersections $\overline{X(\mathcal{S})} = \bigcap_{i=1}^r D(S_i)$ corresponding to sets $\mathcal{S} = \{S_1,\dots,S_r\}$ of subsets in $\{1,\dots,n\}$ called *nests*.

We put

$$\psi_X(q,t) = 1 + \sum_{n\geq 1} P_{X[n]}(q)\frac{t^n}{n!} \in \mathbf{Q}[q][[t]] \ ,$$

$$\chi_X(t) = \psi_X(-1,t) = 1 + \sum_{n\geq 1} \chi(X[n])\frac{t^n}{n!} \in \mathbf{Q}[[\mathbf{t}]] \ .$$

Put also

$$\kappa_m = \frac{q^{2m}-1}{q^2-1} = P_{\mathbf{P}^{m-1}}(q) \ .$$

THEOREM 13.2 *Denote by* $y^0 = y^0(g,t)$ *the unique root in* $t + t^2\mathbf{Q}[q^2][[t]]$ *of any one of the following equations:*

$$\kappa_m(1+y^0)^{q^{2m}} = q^{2m}(q^{2m}+\kappa_m-1)y^0 - q^{2m}(q^{2m}-1)t + \kappa_m \ , \tag{120}$$

$$[q^{2m}t + 1 - (q^{2m}-1+\kappa_m)y^0]y^0_t = 1 + y^0 \ . \tag{121}$$

Then we have in $\mathbf{Q}[q][[t]]$*:*

$$\psi_X(q,t) = (1+y^0)^{P_{X(q)}} \ . \tag{122}$$

THEOREM 13.3 *Denote by* $\eta = \eta(t)$ *the unique root in* $t + t^2\mathbf{Q}[[t]]$ *of any one of the following equations:*

$$m(1+\eta)\log(1+\eta) = (m+1)\eta - t \ , \tag{123}$$

$$(t+1-m\eta)\eta_t = 1 + \eta \ . \tag{124}$$

Then we have in $\mathbf{Q}[[t]]$*:*

$$\chi_X(t) = (1+\eta)^{\chi(X)} \ . \tag{125}$$

We start with combinatorics of the strata.

DEFINITION 13.1 *a)* $\mathcal{S} = \{S_1, \dots, S_r\}$ *is a nest (or n-nest) if* $|S_i| \geq 2$ *for all* i*, and either* $S_i \subset S_j$ *or* $S_j \subset S_i$ *for all* i,j *such that* $S_i \bigcap S_j \neq 0$*.*

In particular, $\mathcal{S} = \emptyset$ *is a nest, and* $\mathcal{S} = \{S\}$ *is a nest, if* $|S| \geq 2$*.*

b) A nest $\mathcal{S}$ *is called* ***whole*** *(resp.* ***broken****) if* $\{1,\dots,n\} \in \mathcal{S}$ *(resp.* $\{1,\dots,n\} \notin \mathcal{S}$*).*

Denote by $X(\mathcal{S}) \subset \overline{X(\mathcal{S})} = \bigcap_{S\in\mathcal{S}} D(S)$ the subset of points not belonging to smaller closed strata. The following facts are proved in [58].

PROPOSITION 13.3 *a) For any $n \geq 2$ and n-nest $\mathcal{S}$, $X(\mathcal{S})$ is a locally closed subset of $X[n]$.*

b) $X[n]$ is the union of pairwise disjoint strata $X(\mathcal{S})$ for all n-nests $\mathcal{S}$.

Now we will show how to pass from nests to marked trees. As above, we consider a bijection $\mu : V_\tau^1 \to [1,\ldots,n]$ as a part of the appropriate marking for our problem. The remaining data is supplied by choosing *orientations of all edges.*

DEFINITION 13.2 *A tree τ marked in this way is called* ***admissible*** *iff:*
a) Every vertex of τ except one has exactly one incoming edge.
b) The exceptional vertex has only outgoing edges, and their number is ≥ 2. This vertex is called the ***source****.*
c) All interior vertices with possible exception of the source have valency ≥ 3.

PROPOSITION 13.4 *The following maps are (1,1):*
$\{$broken n-nests$\}$ $\to$ $\{$whole n-nests$\}$ $\to$ $\{$admissible marked n-trees$\}/(iso)$,

$$\mathcal{S} \mapsto \mathcal{S} \cup \{\{1,\ldots,n\}\} \mapsto \tau(\mathcal{S}) = \tau(\mathcal{S} \cup \{\{1,\ldots,n\}\}) \ .$$

Here τ is defined by its sets of vertices and edges: if $\mathcal{S} = \{S_1,\ldots,S_r\}$, then

$$V_\tau = \{\widetilde{S}_1,\ldots,\widetilde{S}_{n+r}\} := \{S_1,\ldots,S_r,\{1\},\ldots,\{n\}\} \ ,$$

and an edge oriented from $\widetilde{S}_i$ to $\widetilde{S}_j$ connects these two vertices iff $\widetilde{S}_j \subset \widetilde{S}_i$ and no $\widetilde{S}_k$ lies strictly in between these two subsets.

This is proved by direct observation. The following facts are worth mentioning.
a) $\{1,\ldots,n\}$ is the source of $\tau(\mathcal{S})$ for any $\mathcal{S}$.
b) $\{1\},\ldots,\{n\}$ are all end vertices.
c) $i \in S_j$ iff one can pass from $S_j \in V_\tau$ to $\{i\} \in V_\tau$ in τ by always going in the positive direction.
The reader is advised to convince him- or herself that the source has valency ≥ 2 and all other interior vertices have valency ≥ 3.
Denote the source by s and the set of the remaining interior vertices by V_τ^0.

PROPOSITION 13.5 *([58]) The virtual Poincaré polynomials of the strata $X(\mathcal{S})$ are given by the following formulas (we add a formal variable t):*
If $\mathcal{S}$ is a broken n-nest, $s \in V_{\tau(\mathcal{S})}$:

$$t^n P_{X(\mathcal{S})}(q) = \left(\begin{array}{c} P_X(q) \\ |s| \end{array} \right) |s|! \times \prod_{v \in V^0_{\tau(\mathcal{S})}} \kappa_m \left(\begin{array}{c} q^{2m} - 2 \\ |v| - 3 \end{array} \right) (|v| - 3)! \times \prod_{v \in V^1_{\tau(\mathcal{S})}} t \,. \tag{126}$$

If $\mathcal{S}$ is a whole n-nest:

$$t^n P_{X(\mathcal{S})}(q) = P_X(q) \kappa_m \left(\begin{array}{c} q^{2m} - 2 \\ |s| - 2 \end{array} \right) (|s| - 2)! \tag{127}$$

$$\times \prod_{v \in V^0_{\tau(\mathcal{S})}} \kappa_m \left(\begin{array}{c} q^{2m} - 2 \\ |v| - 3 \end{array} \right) (|v| - 3)! \times \prod_{v \in V^1_{\tau(\mathcal{S})}} t \,.$$

Comparing (126) and (127) one sees that one can express the joint contribution of two nests corresponding to an admissible marked tree τ as a product of local weights corresponding to all vertices of τ. The local weight of the source will be

$$\left(\begin{array}{c} P_{X(q)} \\ |s| \end{array} \right) |s|! + P_X(q) \kappa_m \left(\begin{array}{c} q_{2m} - 2 \\ |s| - 2 \end{array} \right) (|s| - 2)!$$

and the remaining local weights coincide and depend only on the valency.

In order to find the appropriate standard weights of marked trees (summands in (78)), we make the following choices.

Put $A = \{+, -\}$. Interpret a mark $+$ (resp. $-$) on a flag as incoming (resp. outgoing) orientation of this flag. Thus, $f : F_\tau \to A$ is a choice of orientation of all flags.

Put $g^{+-} = g^{-+} = 1$, $g^{++} = g^{--} = 0$. This makes the standard weight of (τ, f) vanish unless all edges are unambiguously oriented by f.

Put $C_+ = t$ (see (120) and (121)) and $C_- = 0$. The last choice makes the standard weight vanish unless all end edges are oriented outwards.

Put $C_{+-} = C_{-+} = 0$. This excludes vertices of the type $\to \bullet \to$.

Put also $C_{a_1,\dots,a_k} = 0$ if $\{+,+\} \subset \{a_1,\dots,a_k\}$. This eliminates vertices with ≥ 2 incoming edges.

For tensors with $k \geq 2$ minuses among the indices we put

$$C_{-\dots-} = \begin{pmatrix} P_X(q) \\ k \end{pmatrix} k! + \kappa_m P_X(q) \begin{pmatrix} q^{2m}-2 \\ k-2 \end{pmatrix} (k-2)! \tag{128}$$

(because only the source has all outgoing edges), and

$$C_{+-\dots-} = \kappa_m \begin{pmatrix} q^{2m}-2 \\ k-2 \end{pmatrix} (k-2)! \tag{129}$$

(cf. (126) and (127)).

The standard weight of a marked tree defined by this data again is independent on the part $\mu : V_\tau^1 \to \{1,\dots,n\}$ of the initial marking, which accounts for the factor $\frac{n!}{|\mathrm{Aut}\,\tau|}$ below.

Summarizing, we put

$$\psi_X(q,t) := \sum_{n\geq 2} \frac{t^n}{n!} P_{X|n|}(q)\ , \tag{130}$$

$$Z := \sum_{\tau/(iso)} \frac{1}{|\mathrm{Aut}\,\tau|} \sum_{f:F_\tau\to\{+,-\}} \prod_{\alpha\in E_\tau} g^{f(\partial\alpha)} \prod_{v\in V_\tau} C_{f(\sigma v)}\ , \tag{131}$$

and get from the previous discussion

$$Z = \psi_X(q,t), \quad \frac{\partial}{\partial t} Z := \phi_x(q,t)\ . \tag{132}$$

The arguments in the potential will be denoted $\varphi_+ = x$, $\varphi_- = y$. We see that the potential is

$$\Phi(x,y) = -xy + tx + \kappa_m \sum_{k=2}^{\infty} \begin{pmatrix} q^{2m}-2 \\ k-2 \end{pmatrix} \frac{xy^k}{k(k-1)} +$$

$$+ \sum_{k=2}^{\infty} \begin{pmatrix} P_X(q) \\ k \end{pmatrix} y^k + \kappa_m P_x(q) \sum_{k=2}^{\infty} \begin{pmatrix} q^{2m}-2 \\ k-2 \end{pmatrix} \frac{y^k}{k(k-1)} \tag{133}$$

(we have two arguments x, y but only one $t = t_+$ because $C_- = 0$).

We must solve the system

$$\frac{\partial\Phi}{\partial x}|_{x^0,y^0} = \frac{\partial\Phi}{\partial y}|_{x^0,y^0} = 0 \ , \tag{134}$$

and (83) then tells us that

$$\frac{\partial}{\partial t} Z = \varphi_X(q,t) = x^0 \ . \tag{135}$$

Again, $\Phi(x, y)$ can be easily summed. To write down the functional equation, we need only the x-derivative which for general q is

$$\frac{\partial\Phi}{\partial x} = -y + t + \kappa_m \frac{(1+y)^{q^{2m}} - 1 - q^{2m} y}{q^{2m}(q^{2m}-1)} \ . \tag{136}$$

For $q = -1$:

$$\frac{\partial\Phi}{\partial x} = -y + t + m[(1+y)\log(1+y) - y] \ . \tag{137}$$

We now see that (120), resp. (123), are the equations defining y^0. Taking the derivative in t we get (121) and (124). And since $\Phi(x, y)$ is linear in x, the vanishing of the y-derivative provides an explicit expression of x^0 via y^0 :

$$\varphi_X(q,t) = P_X(q) \frac{(1+y^0)^{P_X(q)} + (q^{2m} + \kappa_m - 1)y^0 - q^{2m}t - 1}{1 + (1 - q^{2m} - \kappa_m)y^0 + q^{2m}t} \ .$$

To see that this is equivalent to (122) one can derivate (122) in t and use (121).

14 Method of VKS-Trees

Let us briefly review the method of VKS-trees [3,4]. Irreducible class 1 representations of the orthogonal group $O\,(n)$ are realized in the space of scalar functions on a unit sphere S_{n-1} in Euclidean space $\mathbb{R}R^n$, which are eigenfunctions of the Laplace operator on sphere [3]:

$$\Delta_\Omega Y_k(\Omega) + k(k+n-2)Y_k(\Omega) = 0. \tag{138}$$

The method of VKS-trees is based on consequent factorization of Laplacian Δ_Ω, for which it is required to introduce polyspherical coordinates.

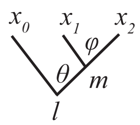
Definition 14.1 *Let $x_0, x_1, \ldots, x_{n-1}$ be the Cartesian coordinates of a point on a sphere S_n. We depict them as lines (see Figure 25a). We join the lines in such way that, at first stage, no more than two lines (edges) meet in a vertex and, at the second stage, no more that two previous joined lines meet in a vertex and so on. As a result we obtain a **VKS − tree** (see Figure 25b).*

As can be seen from the figure, there are two kinds of lines: lines with nodes at both ends and lines with one node. The former lines are called internal and the latter free (or dangling) ends. The number of dangling ends is equal to dimension of space. The number of vertices is equal to number of parameters, determining the location of a point on sphere. To each vertex we assign an angle θ_k. The lines outgoing from the vertex θ_k to the left correspond to $\cos\theta_k$ and those outgoing to the right correspond to $\sin\theta_k$. The vertex θ is a vertex of a graph. Then the path, say, from vertex θ to vertex φ_2 (see Figure 25b) can be represented as a product of internal lines, i.e. a product of cosines and sines, which occur along this path

$$h_{\varphi_2} = \prod_{\theta_k=\theta}^{\varphi_2} = \overset{\theta_k}{\bullet}\!\!-\!\!-\!\!-\!\!\overset{\theta_{k+1}}{\bullet} = \cos\theta\cos\beta\sin\alpha\ . \tag{139}$$

Cartesian coordinates can be determined in terms of angles by multiplying the coefficients h_{φ_k} by the dangling end associated with the angle φ_k.

In three-dimensional space, the following VKS-tree corresponds to the spherical coordinate system



$$x_0 = \cos\theta,\quad x_1 = \sin\theta\cos\varphi,\quad x_2 = \sin\theta\sin\varphi. \tag{140}$$

A relation of equivalence can be introduced on the set of VKS-trees: two VKS-trees belong to the same class if one of them can be transformed into the other by rotation around vertical (perpendicular to the plane of the diagram) axis (or axes), passing through vertex (or vertices). This relation of equivalence enables to split the set of VKS-trees into classes.

VKS-trees belonging to different classes, are truly different structures, i.e. truly topologically inequivalent constructions. There is one class of VKS-trees (systems of polyspherical coordinates) in three-dimensional space, two classes in four-dimensional space, and three classes in five-dimensional space.

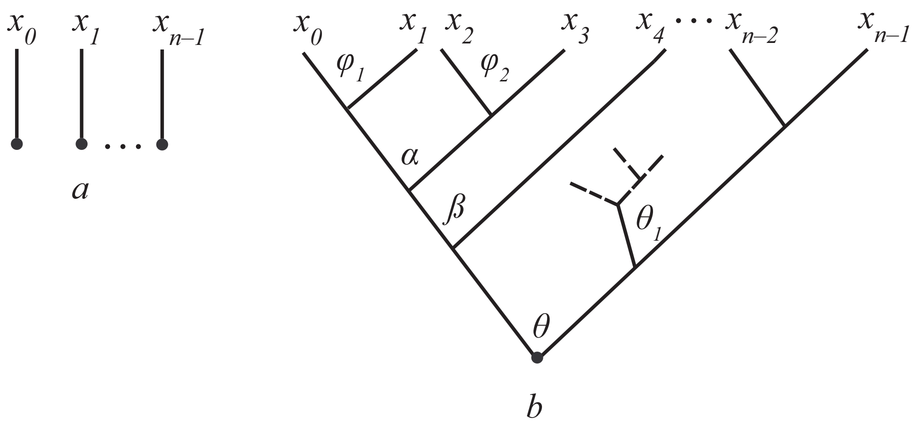


Figure 25: VKS-tree.

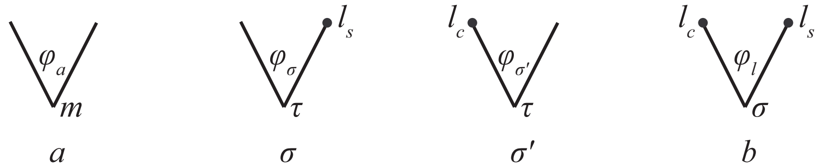


Figure 26: Elementary cells.

In his studies of topologically different VKS-trees, G.I. Kuznetsov [67] has given the following inductive algorithm for writing down Laplace operator Δ_Ω: Equation (138) for Figure 25b in polyspherical coordinates can be written as follows

$$\begin{aligned}\Big\{\frac{1}{\cos^c\theta\sin^s\theta}\frac{\partial}{\partial\theta}\cos^c\theta\sin^s\theta\frac{\partial}{\partial\theta}+\frac{\Delta_{\Omega_c}(\beta)}{\cos^2\theta}+ \\ +\frac{\Delta_{\Omega_s}(\theta_1)}{\sin^2\theta}+k(k+n-1)\Big\}Y_k(\Omega)=0\,,\end{aligned} \tag{141}$$

where, c is number of subsequent vertices to the left of vertex θ; s is number of subsequent vertices to the right of vertex θ, and $c+s=n-2$.

Laplacians $\Delta_{\Omega_c}(\beta)$ and $\Delta_{\Omega_s}(\theta_1)$ are given, respectively, on c- and s-dimensional spheres. For $\Delta_{\Omega_c}(\beta)$ and $\Delta_{\Omega_s}(\theta_1)$, the VKS-trees have left and right parts, counting from the vertex θ of initial VKS-tree, and the roots are β and θ_1. The algorithm for writing these Laplacians in polyspherical coordinates remains the same as for Δ_Ω. Consructing the Laplacian Δ_Ω in this way, gives its form in polyspherical coordinates.

The equation (141) can be solved by the method of variable separation. This yields a constant at each vertex. The constant the associated angle endow the vertex with additional characteristics. To solve equation (141), first solve the partial differential equation for each vertex. The complete solution given by the product of the partial solutions. The three kinds of vertices (cells) which occur with this approach are shown in Figure 26. The requirement of one-fold covering of the sphere imposes the following restrictions on the range of angles [3]:

$$0\le\varphi_a\le 2\pi,\quad 0\le\varphi_\sigma\le\pi,\quad -\frac{\pi}{2}\le\varphi_{\sigma'}\le\frac{\pi}{2},\quad 0\le\varphi_l\le\frac{\pi}{2}. \tag{142}$$

Let us consider an elementary cell of the graph shown at Figure 26*b*. Here m, τ, l_c, l_s, σ are constants of separation of variables. The equation for the variable $\theta=\varphi_l$ is as follows:

$$\begin{aligned}\Big[\frac{1}{\cos^c\theta\sin^s\theta}\frac{\partial}{\partial\theta}\cos^c\theta\sin^s\theta\frac{\partial}{\partial\theta}-\frac{l_c(l_c+c-1)}{\cos^2\theta}- \\ -\frac{l_s(l_s+s-1)}{\sin^2\theta}+\sigma(\sigma+c+s)\Big]\Psi(\theta)=0,\end{aligned} \tag{143}$$

and its solution can be written as

$$\begin{aligned}\Psi^{\alpha,\beta}_{n,c,s}(\theta) = N\cos^{l_c}\theta\sin^{l_s}\theta\;\; \mathcal{P}^{\alpha,\beta}_n(\cos 2\theta),\\ 2n=\sigma-l_c-l_s,\quad \alpha=l_s+\frac{s-1}{2}\equiv l_s+\frac{1}{2}S_{l_s},\\ \beta=l_c+\frac{c-1}{2}\equiv l_c+\frac{1}{2}S_{l_c}, n=0,1,2,\dots,\end{aligned} \tag{144}$$

if and only if $n \geq 0$ i.e. $\sigma \geq l_c + l_s$. Here, $c\,(s)$ is number of left (right) vertices corresponding to the vertex σ; S_{l_s}, S_{l_c} are the numbers of vertices above corresponding to vertices l_s and l_c, $\mathcal{P}^{\alpha,\beta}_n$ are Jakobi polynomials or hyperspherical functions, N is normalizing factor.

Let us consider the cell shown in Figure 26σ. It is worth of noticing that the cell in Figure 26σ' is the same as that of Figure 26σ except for one detail: $\cos\varphi_\sigma$ is substituted for $\sin\varphi_\sigma$.

The equation for the variable $\theta=\varphi_\sigma$ is

$$\left[\frac{1}{\sin^s\theta}\frac{\partial}{\partial\theta}\sin^2\theta\frac{\partial}{\partial\theta}-\frac{l_s(l_s+s-1)}{\sin^2\theta}+\tau(\tau+s)\right]\Psi(\theta)=0, \tag{145}$$

and its solution can be written as follows:

$$\begin{aligned}&\Psi^{\alpha,\alpha}_{n,0,s}(\theta)=N_1\sin^{l_s}\theta\mathcal{P}^{\alpha,\alpha}_n(\cos\theta), &(146)\\ &0\leq\theta\leq\pi,\quad n=\tau-l_s,\quad \alpha=l_s+\frac{s-1}{2}, n\geq 0,1,2,\dots,\end{aligned}$$

where $\mathcal{P}^{\alpha,\alpha}_n$ are Gegenbauer polynomials, N_1 is normalizing coefficient.

For the cell shown in Figure 26a, there is the corresponding the obvious solution

$$\Psi_m(\varphi_a)=\frac{1}{\sqrt{2\pi_a}}e^{im\varphi_a},\quad 0\leq\varphi_a\leq 2\pi,\quad m\in\mathbb{Z}. \tag{147}$$

As an example, let us consider a class 1 representation of the group $O\,(3)$. In spherical coordinates, the factorized solution of equation (138) can be written as follows (see Figure 140):

$$Y_{lm}(\theta,\varphi)=\overline{P}_{lm}(\theta)\frac{e^{im\varphi}}{\sqrt{2\pi}},\quad 0\leq\theta\leq\pi,\quad 0\leq\varphi\leq 2\pi, \tag{148}$$

where $Y_{lm}\,(\theta,\varphi)$ is the spherical function which is an eigenfunction of Laplace operator on sphere S_2.

15 Polyspherical Coordinates in Cayley-Klein Spaces

Cayley-Klein geometries of dimension n are realized on spheres

$$S^n(\boldsymbol{j}) = \left\{ \boldsymbol{x} \in \mathbb{R}R^{n+1}(\boldsymbol{j}) \mid x_0^2 + \sum_{k=1}^{n} x_k^2 \prod_{m=1}^{k} j_m^2 = 1 \right\} \tag{149}$$

in spaces $\mathbb{R}R^{n+1}(\boldsymbol{j})$ resulting from the Euclidean space $\mathbb{R}R^{n+1}$ under the mapping [5]

$$\begin{aligned} &\psi : \mathbb{R}R^{n+1} \to \mathbb{R}R^{n+1}(\boldsymbol{j}) \equiv \mathbb{R}V_{n+1}(\boldsymbol{j}), \quad \psi x_0^* = x_0, \quad \psi x_k^* = x_k \prod_{m=1}^{k} j_m, \\ &\boldsymbol{j} = (j_1, \ldots, j_n); \quad j_k = 1, \iota_k, i, \quad k = 1, 2, \ldots, n. \end{aligned} \tag{150}$$

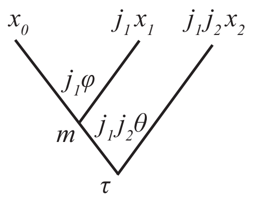
The combination of all possible values of the parameters $\boldsymbol{j}$ produces 3^n different real Cayley-Klein spaces $\mathbb{R}R^{n+1}(\boldsymbol{j}) \equiv \mathbb{R}V_{n+1}(\boldsymbol{j})$ [5]. A unified description of all 3^n Cayley-Klein geometries (geometries of constant curvature space) can be given as a domain of an n-dimensional spherical space parametrized by "concrete" coordinates[21]. Here the total number of nonisomorphic geometries is $N + n = [(3+\sqrt{5})^{n+1} - (3-\sqrt{5})^{n+1}]/2^{n+1}\sqrt{5}$ [5]. The operation of some transition between their groups is based on introducing a set of the parameters ($\boldsymbol{j} := j_1 \ldots, j_n$). Each of the parameters can take on three values: real, purely imaginary and dual units $\boldsymbol{\iota}$ [5].

Under the mapping (150) transforming Euclidean space $\mathbb{R}R^n$ into Cayley-Klein space $\mathbb{R}R^n(\boldsymbol{j})$, Cartesian coordinates in $\mathbb{R}R^n$ are multiplied by products of parameters $\boldsymbol{j}$. In [5, § 5.2–5.5] it is shown that angles are multiplied by some products of parameters $\boldsymbol{j}$ as well. Under the mapping ψ, the symmetry ("equality") of Cartesian coordinates disappears. In method of VKS-trees, this reveals is revealed by the fact that the operation of rotation around vertical axis passing through vertex does not transform a VKS-tree into an equivalent VKS-tree; the partition of the set of VKS-trees into classes is not possible. The other peculiarity is that, for imaginary values of the parameters $\boldsymbol{j}$, the sphere $S_{n-1}(\boldsymbol{j})$ can not

[21] By "concrete" coordinates, we mean real, purely imaginary or dual numbers. The latter were introduced by Clifford [68] in such a way that a dual number itself differs from zero but vanishes when squared.

be covered by one (polyspherical) map. For example, the Minkowski plane can be covered by four system of polar coordinates [69]. To simplify the exposition, we shall consider in this paragraph only contractions of groups, i.e. parameters $j_k = 1, \iota_k$, $k = 1, 2, \ldots, n-1$.

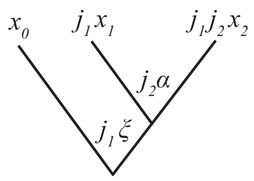
In the space $\mathbb{R}R_3\,(\boldsymbol{j})$, the sphere $S_2\,(\boldsymbol{j}) = \{\boldsymbol{x} \mid x_0^2 + j_1^2 x_1^2 + j_1^2 j_2^2 x_2^2 = 1\}$ admits two systems of spherical coordinates:



$$x_0 = \cos j_1 j_2 \theta \cos j_1 \varphi,\quad x_1 = \frac{1}{j_1} \cos j_1 j_2 \theta \sin j_1 \varphi,\quad x_2 = \frac{1}{j_1 j_2} \sin j_1 j_2 \theta, \tag{151}$$

$$\varphi \in \Phi(j_1) = \begin{cases} [0, 2\pi], & j_1 = 1, \\ \mathbb{R}, & j_1 = \iota_1, \end{cases} \tag{152}$$

$$\theta \in \Theta(j_1 j_2) = \begin{cases} \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], & j = 1, \\ \mathbb{R}, & j \neq 1; \end{cases} \tag{153}$$



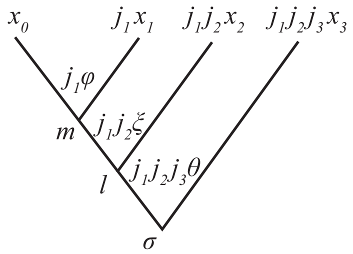
$$x_0 = \cos j_1 \xi,\quad x_1 = \frac{1}{j_1} \sin j_1 \xi \cos j_2 \alpha,\quad x_2 = \frac{1}{j_1 j_2} \sin j_1 \xi \sin j_2 \alpha, \tag{154}$$

$$\alpha \in \Phi(j_2), \quad \xi \in \mathbb{Z}(j_1, j_2) = \begin{cases} \mathbb{Z}_0(j_1), & j_2 = 1, \\ \Phi(j_1), & j_2 = \iota_2, \end{cases} \tag{155}$$

$$\mathbb{Z}_0(j_1) = \begin{cases} [0, \pi], & j_1 = 1, \\ \mathbb{R}^+, & j_1 = \iota_1. \end{cases} \tag{156}$$

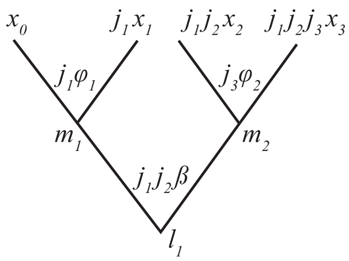
For $j_1 = \iota_1$ both systems of coordinates describe the connected component of sphere.

The sphere $S_3(\boldsymbol{j}) = \{\boldsymbol{x} \,|\, x_0^2 + j_1^2 x_1^2 + j_1^2 j_2^2 x_2^2 + j_1^2 j_2^2 j_3^2 x_3^2 = 1\}$ in the space $\mathbb{R}R^4(\boldsymbol{j})$ admits three systems of polyspherical coordinates:



$$
\begin{aligned}
x_0 &= \cos j_1\varphi \cos j_1 j_2 \xi \cos j_1 j_2 j_3 \theta, \\
x_1 &= \frac{1}{j_1} \sin j_1\varphi \cos j_1 j_2 \xi \cos j_1 j_2 j_3 \theta, \\
x_2 &= \frac{1}{j_1 j_2} \sin j_1 j_2 \xi \cos j_1 j_2 j_3 \theta, \\
x_3 &= \frac{1}{j_1 j_2 j_3} \sin j_1 j_2 j_3 \theta,
\end{aligned}
\tag{157}
$$

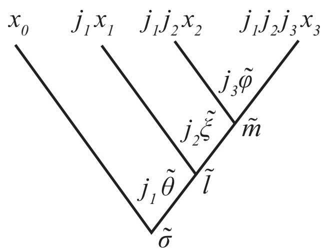
$$
\varphi \in \Phi(j_1), \quad \xi \in \Theta(j_1 j_2), \quad \theta \in \Theta(j_1 j_2 j_3) = \begin{cases} \left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right], & \boldsymbol{j} = \boldsymbol{1}, \\ \mathbb{R}, & \boldsymbol{j} \neq \boldsymbol{1}; \end{cases}
$$



$$
\begin{aligned}
x_0 &= \cos j_1\varphi_1 \cos j_1 j_2 \beta, \\
x_1 &= \frac{1}{j_1} \sin j_1\varphi_1 \cos j_1 j_2 \beta, \\
x_2 &= \frac{1}{j_1 j_2} \cos j_3\varphi_2 \sin j_1 j_2 \beta, \\
x_3 &= \frac{1}{j_1 j_2 j_3} \sin j_3\varphi_2 \sin j_1 j_2 \beta,
\end{aligned}
\tag{158}
$$

$$
\begin{aligned}
\varphi_1 &\in \Phi(j_1), \\
\varphi_2 &\in \Phi(j_3),
\end{aligned}
\quad \beta \in \mathcal{B}(j_1, j_2, j_3) = \begin{cases} Z^+(j_1 j_2), & j_3 = 1, \\ \Theta(j_1 j_2), & j_3 = \iota_3, \end{cases}
$$

$$
Z^+(j_1 j_2) = \begin{cases} [0, \pi], & \boldsymbol{j} = \boldsymbol{1}, \\ \mathbb{R}^+, & \boldsymbol{j} \neq \boldsymbol{1}; \end{cases}
$$



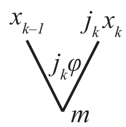
$$
\begin{aligned}
x_0 &= \cos j_1 \widetilde{\theta}, \\
x_1 &= \frac{1}{j_1} \sin j_1 \widetilde{\theta} \cos j_2 \widetilde{\xi}, \\
x_2 &= \frac{1}{j_1 j_2} \sin j_1 \widetilde{\theta} \sin j_2 \widetilde{\xi} \cos j_3 \widetilde{\varphi}, \\
x_3 &= \frac{1}{j_1 j_2 j_3} \sin j_1 \widetilde{\theta} \sin j_2 \widetilde{\xi} \sin j_3 \widetilde{\varphi},
\end{aligned}
\tag{159}
$$

$$\widetilde{\varphi} \in \Phi(j_3), \quad \widetilde{\xi} \in Z(j_2, j_3), \quad \widetilde{\theta} \in Z(j_1, j_2).$$

For $j_1 = \iota_1$, three systems of polyspherical coordinates describe the connected component $x_0 = 1$ of the sphere $S_3(\iota_1, j_2, j_3)$.

16 Equations for Elementary Cells

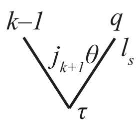
Let us begin with the cell shown in Figure 26a. We obtain



$$\Psi_m(\varphi, j_k) = N e^{im\varphi}. \tag{160}$$

Dependence on the parameter j_k is included in the range of values $\Phi(j_k)$ of the variable φ. The constant of separation is $m \in \mathbb{Z}$ for $j_k = 1$ and $m \in \mathbb{R}$ for $j_k = \iota_k$. The normalizing factor is $N = 1/\sqrt{2\pi}$ for $j_k = 1$ and $N = 1$ for $j_k = \iota_k$. In the latter case, the solution $\Psi_m(\varphi, \iota_k)$ is normalized to a delta-function.

The cell shown in Figure 26σ is transformed into the cell



$$\theta \in Z(j_k, j_{k+1}), \tag{161}$$

where $k-1$ is the order number of coordinate x_{k-1} connected with the left branch outgoing from vertex τ, q is the order number of the of the last coordinate, connected with node l_s, and s is the number of nodes to the right from vertex τ.

To the cell (161), there corresponds the sphere

$$S_{q-k+1}(j_k, \dots, j_q) = \left\{ \boldsymbol{x} \,|\, x_{k-1}^2 + j_k^2 x_k^2 + \cdots + \left(\prod_{r=k}^{q} j_r^2 \right) x_q^2 = 1 \right\}$$

and to the vertex l_s – sphere $S_{q-k}(j_{k+1}, \dots, j_q) = \left\{ \boldsymbol{x} \,|\, x_k^2 + j_{k+1}^2 x_{k+1}^2 + \cdots + \left(\prod_{r=k+1}^{q} j_r^2 \right) x_q^2 = 1 \right\}$. Under mapping ψ, the Laplacian (or, otherwise, Casimir operator of second order) is transformed according to the rule $\Delta_\theta(\boldsymbol{j}) =$

$\left(\prod_{r=k}^{q} j_r^2\right)\Delta^*_{\theta^*}$ $(\rightarrow)$, where the asterisk indicates the quantities entering equations (145) which are transformed as follows: $\psi\theta^* = j_{k+1}\theta$, $\tau = \tau^* j_k A$, $l_s = l_s^* A$, and $A = \prod_{r=k+1}^{q} j_r$. Transforming (145), we obtain the equation

$$\left\{\frac{A^2}{\sin^2 j_k\theta}\frac{\partial}{\partial\theta}\sin^s j_k\theta\frac{\partial}{\partial\theta} - j_k^2\frac{l_s[l_s+(s-1)A]}{\sin^2 j_k\theta} + \tau(\tau+sA)\right\}\Psi(\theta) = 0 \qquad (162)$$

which corresponds to cell (161). Its formal solution is the function

$$\Psi^{\alpha,\alpha}_{n,0,s}(\theta) = N(\sin j_k\theta)^{l_s}\mathcal{P}^{\alpha,\alpha}_{\tau-j_k l_s}(\cos j_k\theta), \qquad (163)$$
$$\alpha = A\alpha^*(\rightarrow) = l_s + \frac{s-1}{2}A, \quad n = j_k A n^*(\rightarrow) = \tau - j_k l_s,$$

where N is a normalizing factor.

Let $j_k = \iota_k$ and $A \neq \boldsymbol{\iota}$, i.e. $A = 1$. Then $\sin\iota_k\theta = \iota_k\theta$, and equation (162) turns into

$$\Psi'' + \frac{A_s}{\theta}\Psi' + \left[\tau^2 - \frac{l_s(l_s+(s-1)A)}{\theta^2}\right]\Psi = 0. \qquad (164)$$

This is the Lommel equation [70], which can be expressed in terms of Bessel functions:

$$\Psi_{\tau,l_s,s}(\theta) = \theta^{\frac{1-s}{2}} J_{l_s+\frac{s-1}{2}}(\tau\theta). \qquad (165)$$

Let $A = \boldsymbol{\iota}$, i.e. some of the parameters j_r, $k+1 \le r \le q$, take on dual values. Then (162) can be rewritten as the algebraic equation

$$\left(\tau^2 - j_k^2\frac{l_s^2}{\sin^2 j_k\theta}\right)\Psi(\theta) = 0, \qquad (166)$$

which connects the constants of separation in neighbouring vertices

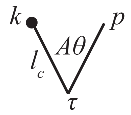
$$l_s = \tau\frac{1}{j_k}\sin j_k\theta. \qquad (167)$$

The case $j_k = \iota_k$, $A = \boldsymbol{\iota}$ is obtained from (167) with $j_k = \iota_k$; the result is $l_s = \tau\theta$. Comparing the formal solution (163) with the solution (165), we obtain the final relation for Gegenbauer polynomials

$$(\iota_k\theta)^{l_s}\mathcal{P}^{\alpha,\alpha}_{\tau-\iota_k l_s}(\cos\iota_k\theta) = \theta^{\frac{1-s}{2}} J_\alpha(\tau\theta) \qquad (168)$$

written up to normalizing factors.

The cell shown in Figure $26\sigma'$ is transformed into



$$\theta \in \Theta(A) = \begin{cases} \left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right], & A = 1, \\ \mathbb{R}, & A \neq 1, \end{cases} \qquad A = \prod_{\iota=k+1}^{p} j_r, \tag{169}$$

where k is the order number of the first coordinate x_k connected with vertex l_c, p is the order number of coordinate, connected with the right branch outgoing from vertex τ and c is number of nodes to the left of vertex τ.

To cell (169) there corresponds the sphere

$$S_{p-k}\left(j_{k+1}, \ldots, j_p\right) = \left\{ \boldsymbol{x} \,|x_k^2 + j_{k+1}^2 x_{k+1}^2 + \cdots + \left(\prod_{r=k+1}^{p} j_r^2 \right) x_p^2 = 1 \right\}$$

and to the vertex l_c, the sphere $S_{p-k-1}(j_{k+1}, \ldots, j_{p-1})$. Under mapping ψ, the Laplacian is transformed according to the rule $\Delta_\theta\left(\boldsymbol{j}\right) = A^2 \Delta^*_{\theta^*}\left(\rightarrow\right)$, where the asterisk indicates the quantities in the equation of the cell shown in Figure 26σ. These quantities can be written as follows: $\psi\theta^* = A\theta$, $\tau = \tau^* A$, $l_c = l_c^* B$, where $B = \prod\limits_{r=k+1}^{p-1} j_r$, i.e. $A = j_p B$. To cell (169), there corresponds the equation

$$\left\{ \frac{1}{\cos^c A\theta} \frac{\partial}{\partial\theta} \cos^c A\theta \frac{\partial}{\partial\theta} - j_p^2 \frac{l_c[l_c + (c-1)B]}{\cos^2 A\theta} + \tau(\tau + cA) \right\} \Psi(\theta) = 0. \tag{170}$$

Its formal solution can be written as follows

$$\begin{aligned} &\Psi_{n,0}^{\alpha,\alpha}(\theta) = N(\cos A\theta)^{l_c} \mathcal{P}_{\tau - j_p l_s}^{\alpha,\alpha}(\sin A\theta), \\ &\alpha = B\alpha^*(\rightarrow) = l_c + \frac{c-1}{2} B, \quad n = An^*(\rightarrow) = \tau - j_p l_c. \end{aligned} \tag{171}$$

Let $B = \boldsymbol{\iota}$, some of the parameters j_r, $k+1 \le r \le p-1$, are equal to dual units, and $j_p \neq \iota_p$. Then the equation (170) takes following form:

$$\Psi''(\theta) + (\tau^2 - j_p^2 l_c^2)\Psi(\theta) = 0, \tag{172}$$

and its solution can be written as

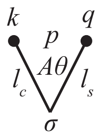
$$\Psi_{\tau,l_c}(\theta) = e^{i\theta\sqrt{\tau^2 - j_p^2 l_c^2}}. \tag{173}$$

The case $j_p = \iota_p$ is obtained from (172) and (173). Comparing (171) and (173), we obtain limit relations for Gegenbauer polynomials

$$\begin{aligned} &\mathcal{P}^{\alpha,\alpha}_{\tau - j_p l_c}(\sin \iota j_p \theta) = e^{i\theta\sqrt{\tau^2 - j_p^2 l_c^2}}, \quad \alpha = l_c + \iota\frac{c-1}{2}, \\ &\mathcal{P}^{\alpha,\alpha}_{\tau - \iota_p l_c}(\sin \iota_p \theta) = e^{i\theta\tau}, \quad \alpha = l_c + \frac{c-1}{2} \end{aligned} \tag{174}$$

written up to normalizing factors.

The cell, shown at Figure 26b, is transformed into the cell



$$\theta \in \begin{cases} Z^+(A), & j_{p+2} \neq l_{p+2}, \\ \Theta(A), & j_{p+2} = l_{p+2}, \end{cases} \quad A = \prod_{r=k+1}^{p+1} j_r,$$

$$Z^+(A) = \begin{cases} \left[0, \dfrac{\pi}{2}\right], & A = 1, \\ \mathbb{R}^+, & A \neq 1, \end{cases} \quad \Theta(A) = \begin{cases} \left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right], & A = 1, \\ \mathbb{R}, & A \neq 1, \end{cases} \tag{175}$$

where k is order number of the first coordinate x_k, connected with the left vertex l_c; p is order number of the last coordinate connected with vertex l_c; q is order number of the last coordinate connected with the right vertex l_s.

To cell (175) there corresponds sphere

$$S_{q-k}\left(j_{k+1}, \ldots, j_q\right) = \Bigg\{\boldsymbol{x} \,|x_k^2 + j_{k+1}^2 x_{k+1}^2 + \cdots + \left(\prod_{r=k+1}^{p} j_r^2\right) x_p^2 +$$
$$+A\left(x_{p+1}^2 + j_{p+2}^2 x_{p+2}^2 + \cdots + \left(\prod_{r=p+2}^{q} j_r^2\right) x_q^2\right) = 1\Bigg\}.$$

Equating the expression in round brackets to unit, we can obtain sphere $S_{q-p-1}\left(j_{p+2}, \ldots, j_q\right)$, which corresponds to vertex l_s, and equating to unit the expression in front of round brackets, we are able to write the sphere $S_{p-k}\left(j_{k+1}, \ldots, j_p\right)$, corresponding to vertex l_c. Under mapping ψ Laplacian (143) is transformed according to the rule $\Delta_\theta\left(\boldsymbol{j}\right) = \left(\prod\limits_{r=k+1}^{q} j_r^2\right) \Delta^*_{\theta^*}(\rightarrow)$, and quantities,

entering (143), are transformed as follows:
$\psi\theta^* = A\theta$, $\sigma = \sigma^* \prod\limits_{r=k+1}^{q} j_r$, $l_c = l_c^* A/j_{p+1}$, $l_s = l_s^* B$, where $B = \prod\limits_{r=p+2}^{q} j_r$.

Transforming (143), we obtain equation, corresponding to cell (175):

$$\left\{ \frac{B^2}{\cos^c A\theta \sin^s A\theta} \frac{\partial}{\partial\theta} \cos^c A\theta \sin^s A\theta \frac{\partial}{\partial\theta} - j_{p+1}^2 B^2 \frac{l_c[l_c + (c-1)A/j_{p+1}]}{\cos^2 A\theta} - \right.$$
$$\left. - A^2 \frac{l_s[l_s + (s-1)B]}{\sin^2 A\theta} + \sigma[\sigma + (c+s)AB] \right\} \Psi(\theta) = 0. \qquad (176)$$

Its formal solution can be written as follows:

$$\Psi_{n,l_c,l_s}^{\alpha,\beta}(\theta) = N(\sin A\theta)^{l_s}(\cos A\theta)^{l_c} \mathcal{P}_n^{\alpha,\beta}(\cos 2A\theta), \qquad (177)$$
$$2n = \sigma - l_c B j_{p+1} - l_s A, \quad \alpha = l_s + \frac{s-1}{2}B, \quad \beta = l_c + \frac{c-1}{2j_{p+1}}A.$$

For $A = 1$, $B = \boldsymbol{\iota}$, i.e. when one or more parameters j_r, $p+2 \le r \le q$, take dual values, equation (176) is transformed into algebraic equation, connecting constants of separation of variables in neighbouring vertices by relation

$$l_s^2 = \sigma^2 \frac{1}{A^2} \sin^2 A\theta. \qquad (178)$$

For $B = 1$, $A = \boldsymbol{\iota}$, $j_{p+1} \ne \iota_{p+1}$, equation (176) takes form

$$\Psi'' + \frac{s}{\theta}\Psi' + \left[\sigma^2 - j_{p+1}^2 l_c^2 - \frac{l_s(l_s + s - 1)}{\theta^2} \right] \Psi = 0. \qquad (179)$$

This is Lommel equation [70], which solution can be expressed in terms of Bessel function

$$\Psi_{\sigma,l_s,l_c,s}(\theta) = N\theta^{\frac{l-s}{2}} J_{l_s + \frac{s-1}{2}} \left(\theta \sqrt{\sigma^2 - j_{p+1}^2 l_c^2} \right). \qquad (180)$$

For $j_{p+1} = \iota_{p+1}$ equation and its solution come out of (179) and (180).

17 Representations of Groups $SO\,(3;\boldsymbol{j})$, $SO\,(4;\boldsymbol{j})$

Let us consider representations of class 1 of groups $SO\,(3;\boldsymbol{j})$, $j_1 = 1, \iota_1$; $j_2 = 1, \iota_2$ in spherical system of coordinates, corresponding to the VKS-tree (151). Metrics on sphere $S_2(\boldsymbol{j})$ in these coordinates is as follows

$$dl^2(\boldsymbol{j}) = d\varphi^2 \cos^2 j_1 j_2 \theta + j_2^2 d\theta^2. \tag{181}$$

Laplacian on sphere, corresponding to this metrics, is

$$\Delta_3(\boldsymbol{j}) = \frac{1}{\cos j_1 j_2 \theta} \frac{\partial}{\partial \theta} \cos j_1 j_2 \theta \frac{\partial}{\partial \theta} + \frac{j_2^2}{\cos^2 j_1 j_2 \theta} \frac{\partial^2}{\partial \varphi^2}. \tag{182}$$

Factorized eigenfunctions of Laplacian (182), which are solutions of equation $\Delta_3\,(\boldsymbol{j})\,\Psi\,(\theta,\varphi) = -\tau\,(\tau+1)\,\Psi\,(\theta,\varphi)$, can be found, multiplying solution (161) by solution (171) for $c = 1$, $l_c \equiv m$:

$$\Psi_{\tau,m}(\theta,\varphi) = N e^{im\varphi} (\cos j_1 j_2 \theta)^m \mathcal{P}^{m,m}_{\tau - j_2 m}(\sin j_1 j_2 \theta). \tag{183}$$

In coordinate system, corresponding to the VKS-tree (154), metrics on sphere $S_2\,(\boldsymbol{j})$ can be written

$$dl^2(\boldsymbol{j}) = d\zeta^2 + j_2^2 \frac{1}{j_1^2} \sin^2 j_1 \zeta d\alpha^2, \tag{184}$$

and Laplacian takes the following form

$$\Delta_3(\boldsymbol{j}) = \frac{j_2^2}{\sin j_1 \zeta} \frac{\partial}{\partial \zeta} \sin j_1 \zeta \frac{\partial}{\partial \zeta} + \frac{j_1^2}{\sin^2 j_1 \zeta} \frac{\partial^2}{\partial \alpha^2}. \tag{185}$$

Its eigenfunctions can be found, multiplying solution (160) and (163) for $s = 1$, $l_s \equiv l$, $k = 1$, $q = 2$, which gives

$$\Psi_{\tau,l}(\zeta,\alpha) = N e^{im\alpha} (\sin j_1 \zeta)^l \mathcal{P}^{l,l}_{\tau - j_1 l}(\cos j_1 \zeta). \tag{186}$$

On sphere $S_3\,(\boldsymbol{j})$ metrics in polyspherical coordinates, corresponding to the VKS-tree (157), is given by expression

$$dl^2(\boldsymbol{j}) = \cos^2 j_1 j_2 j_3 \theta (\cos^2 j_1 j_2 \xi \; d\varphi^2 + j_2^2 \; d\xi^2) + j_2^2 j_3^2 \; d\theta^2. \tag{187}$$

It corresponds to Laplacian

$$\Delta_4(\boldsymbol{j}) = \frac{1}{\cos^2 j_1 j_2 j_3 \theta}\left(\frac{\partial}{\partial\theta}\cos^2 j_1 j_2 j_3 \theta \frac{\partial}{\partial\theta} + \right.$$
$$\left. + \frac{j_3^2}{\cos j_1 j_2 \xi}\frac{\partial}{\partial\xi}\cos j_1 j_2 \xi \frac{\partial}{\partial\xi} + \frac{j_2^2 j_3^2}{\cos j_1 j_2 \xi}\frac{\partial^2}{\partial\varphi^2}\right), \tag{188}$$

which eigenfunctions, realizing the representation of class 1 of group $SO(4;\boldsymbol{j})$, can be found by multiplying functions (171) and have the form

$$\Psi_{\sigma,l,m}(\theta,\xi,\varphi) = Ne^{im\varphi}(\cos j_1 j_2 \xi)^m (\cos j_1 j_2 j_3 \theta)^l \times$$
$$\times \mathcal{P}^{m,m}_{l-j_2 m}(\sin j_1 j_2 \xi)\mathcal{P}^{l+\frac{1}{2}j_1 j_2, l+\frac{1}{2}j_1 j_2}_{\sigma - j_3 l}(\sin j_1 j_2 j_3 \theta). \tag{189}$$

In polyspherical coordinates (158) metrics on sphere $S_3(\boldsymbol{j})$ is as follows:

$$dl^2(\boldsymbol{j}) = j_2^2 d\beta^2 + \cos^2 j_1 j_2 \beta d\varphi_1^2 + j_3^2 \frac{1}{j_1^2}\sin^2 j_1 j_2 \beta d\varphi_2^2. \tag{190}$$

To this metrics there corresponds Laplacian

$$\begin{aligned}\Delta_4(\boldsymbol{j}) &= \frac{j_3^2}{\sin j_1 j_2 \beta \cos j_1 j_2 \beta}\frac{\partial}{\partial\beta}\sin j_1 j_2 \beta \cos j_1 j_2 \beta \frac{\partial}{\partial\beta} + \\ &+ \frac{j_2^2 j_3^2}{\cos^2 j_1 j_2 \beta}\frac{\partial^2}{\partial\varphi^2} + \frac{j_1^2 j_2^2}{\sin^2 j_1 j_2 \beta}\frac{\partial^2}{\partial\varphi_2^2}\ ,\end{aligned} \tag{191}$$

and its eigenfunctions, realizing representation of class 1 of group $SO(4;\boldsymbol{j})$ in coordinate system (158), can be found by using (160) and (177).

These functions are

$$\Psi_{l_1,m_1,m_2}(\beta,\varphi_1,\varphi_2) = \tag{192}$$
$$= Ne^{im_1\varphi_1 + im_1\varphi_2}(\cos j_1 j_2 \beta)^{m_1}(\sin j_1 j_2 \beta)^{m_2}\mathcal{P}^{m_2,m_1}_{l_1 - j_2 j_3 m_1 - j_1 j_2 m_2}(\cos 2 j_1 j_2 \beta).$$

Finally, in coordinates (159) metrics on sphere $S_3(\boldsymbol{j})$ can be written as

$$dl^2(\boldsymbol{j}) = d\widetilde{\theta}^2 + \frac{1}{j_1^2}\sin^2 j_1 \widetilde{\theta}(j_2^2\, d\widetilde{\xi}^2 + j_3^2 \sin^2 j_2 \widetilde{\xi}\, d\widetilde{\varphi}^2). \tag{193}$$

To this metrics there corresponds Laplace operator

$$\Delta_4(\boldsymbol{j}) = \frac{1}{\sin^2 j_1\widetilde{\theta}}\biggl(j_2^2 j_3^2\,\frac{\partial}{\partial\widetilde{\theta}}\sin^2 j_1\widetilde{\theta}\,\frac{\partial}{\partial\widetilde{\theta}} + \\ + \frac{j_1^2 j_3^2}{\sin j_2\widetilde{\xi}}\,\frac{\partial}{\partial\widetilde{\xi}}\sin j_2\widetilde{\xi}\,\frac{\partial}{\partial\widetilde{\xi}} + \frac{j_1^2 j_2^2}{\sin^2 j_1\widetilde{\xi}}\,\frac{\partial^2}{\partial\widetilde{\varphi}^2}\biggr), \tag{194}$$

its eigenfunctions

$$\Psi_{\widetilde{\sigma},\widetilde{l},\widetilde{m}}(\widetilde{\theta},\widetilde{\xi},\widetilde{\varphi}) = Ne^{i\widetilde{m}\widetilde{\varphi}}(\sin j_2\widetilde{\xi})^{\widetilde{m}}(\sin j_1\widetilde{\theta})^{\widetilde{l}}\times \\ \times\mathcal{P}^{\widetilde{m},\widetilde{m}}_{\widetilde{l}-j_2\widetilde{m}}(\cos j_2\widetilde{\xi})\mathcal{P}^{\widetilde{l}+\frac{1}{2}j_2j_3,\widetilde{l}+\frac{1}{2}j_2j_3}_{\widetilde{\sigma}-j_1\widetilde{l}}(\cos j_1\widetilde{\theta}) \tag{195}$$

can be found, using the described algorithm. They realize representation of class 1 of group $SO\,(4;\boldsymbol{j})$ in polyspherical coordinate system (159). Using relations from § 16, it is easy to find eigenfunctions of Laplace operators of contracted groups [71–75].

18 Functor category of VKS-trees

Let us briefly review a category **Euclid** and classical Cayley-Klein categories [5, § 7.1.5].

A Euclidean vector space $\mathbb{R}R^{n+1}$ is a vector space over the field $\mathbb{R}$ of real numbers-equipped with an inner product function $(,) : \mathbb{R}R^{n+1}\times\mathbb{R}R^{n+1}\to\mathbb{R}$ which is bilinear, symmetric, and positive definite. These spaces are the objects of a category **Euclid**, with morphisms those linear maps which preserve inner product. There are two functors

$$U : \mathbf{Euclid}\to\mathbf{Vct}_{\mathbb{R}}, \qquad * : (\mathbf{Euclid})^{op}\to\mathbf{Vct}_{\mathbb{R}}$$

to the category of real vector spaces: The (covariant) forgetful functor U "forget the inner product" and the contravariant functor "take the dual space".

On the other hand, in the previous Section we have found representations of orthogonal groups in Cayley-Klein spaces and we have shown that their Laplace (Casimir) operators on sphere and other algebraic constructions can be obtained by transfering the corresponding constructions for classical Lie groups in accord with (150). Such an approach is natural and justified by the fact that classical

Lie groups and their characteristic algebraic constructions are well studied. But is such an approach the only one? Is it possible to take as the initial space not only the Euclidean space but also a Cayley-Klein space? The positive answer to this question was given by Theorem on the structure of transitions between Cayley-Klein spaces [76–79].
The transitions from the $(n+1)$-dimensional real Cayley-Klein space $\mathbb{R}V^{n+1}(\boldsymbol{j})$ to the real Cayley-Klein space $\mathbb{R}V^{n+1}(\boldsymbol{j}')$, and from the groups $SO(n+1,\mathbb{R};\boldsymbol{j})$, $Sp\,(n,\mathbb{R};\boldsymbol{j})$ to the groups $SO(n{+}1,\mathbb{R};\boldsymbol{j}')$, $Sp\,(n,\mathbb{R};\boldsymbol{j}')$ as well, can be, respectively, obtained from (150) and the transitions

$$\psi : \mathbb{R}V^{n+1}(\boldsymbol{j}) \to \mathbb{R}V^{n+1}(\boldsymbol{j}'), \quad \psi' x_0 = x_0', \quad \psi' x_k = x_k' \prod_{m=1}^{k} j_m' j_m^{-1}, \tag{196}$$

by the same substitution of parameters j_k for $j_k' j_k^{-1}$, where $\boldsymbol{j} = (j_1, \ldots, j_n)$ and each of parameters j_k assuming three values: $j_k = 1, \iota_k, i$.

Similarly the permissibility of these transitions can be justified for complex Cayley-Klein space $\mathbb{C}V^{n+1}(\boldsymbol{j})$ which emerge from the $(n{+}1)$-dimensional complex Euclidean space $\mathbb{C}R^{n+1}$ by the mapping

$$\begin{gathered} \psi : \mathbb{C}R^{n+1} \to \mathbb{C}V^{n+1}(\boldsymbol{j}), \quad \psi z_0^* = z_0, \\ \psi z_k^* = z_k \prod_{m=1}^{k} j_m, \quad k = 1, 2, \ldots, n, \end{gathered} \tag{197}$$

where $z_0^*, z_k^* \in \mathbb{C}R^{n+1}$, $z_0, z_k \in \mathbb{C}V^{n+1}(\boldsymbol{j})$ are complex Cartesian coordinates. The totality of all possible values of the parameter $\boldsymbol{j}$ gives 3^n different real $\mathbb{R}V^{n+1}(\boldsymbol{j})$ and, correspondingly, complex $\mathbb{C}V^{n+1}(\boldsymbol{j})$ Cayley-Klein spaces.

The quadratic form $(\boldsymbol{z}^*, \boldsymbol{z}^*) = \sum\limits_{m=0}^{n} |z_m^*|^2$ of the space $\mathbb{C}R^{n+1}$ turns into the quadratic form

$$(\boldsymbol{z}, \boldsymbol{z}) = |z_0|^2 = \sum_{k=1}^{n} |z_k|^2 \prod_{m=1}^{k} j_m^2. \tag{198}$$

of the space $\mathbb{C}R^{n+1}(\mathbf{j})$ under the mapping (197). Here $|z_k| = (x_k^2 + y_k^2)^{1/2}$ is the absolute value (modulus) of the complex number $z_k = x_k + iy_k$, and $\boldsymbol{z}$ is the complex vector: $\boldsymbol{z} = (z_0, z_1, \ldots, z_n)$.

Let us define (formally) the transition from the space $\mathbb{C}V^{n+1}(\mathbf{j})$ and generators to the space $\mathbb{C}V^{n+1}(\boldsymbol{j}')$ by transformations, which can be obtained from the transformations (197), substituting in the latter the parameters j_k for $j'_k j_k^{-1}$, i.e.

$$\psi' : \mathbb{C}V^{n+1}(\boldsymbol{j}) \to \mathbb{C}V^{n+1}(\mathbf{j}'), \quad \psi' z_0 = z'_0,$$
$$\psi' z_k = z'_k \sum_{m=1}^{k} j'_m j_m^{-1}, \quad k = 1, 2, \ldots, n. \tag{199}$$

A Cayley-Klein space $\mathbb{K}(\boldsymbol{j}) := \mathbb{R}V^{n+1}(\boldsymbol{j})$, $\mathbb{C}V^{n+1}(\boldsymbol{j})$, $\mathbb{R}V^{n}(\boldsymbol{j}) \otimes \mathbb{R}V^{n}(\boldsymbol{j})$ is called non-fiber space, if none of the parameters $j_1, \ldots, j_n$ assumes a dual value. A space $\mathbb{K}(\boldsymbol{j})$: is called $(k_1, k_2 \ldots, k_p)$–fiber space, if $1 \le k_1 < k_2 < \cdots < k_p \le n$ and $j_{k_1} = \iota_{k_1}, \ \ldots, \ j_{k_p} = \iota_{k_p}$, and the other $j_k = 1, i$.

However, transitions (196) and (199) do not make sense for all Cayley-Klein groups and spaces, because for the dual values of parameters $\boldsymbol{j}$ the expressions ι_k^{-1}, $\iota_m \iota_k^{-1}$ for $k \ne m$ are not defined. We have defined (see [5]) only expressions $\iota_k \iota_k^{-1}$, $k = 1, 2, \ldots, n$. So if for some k we put $j_k = \iota_k$, then the transformations (196) and (199) will be defined only in the case when the dashed parameter with the same number is equal to the same purely dual number, i.e. $j'_k = \iota_k$.

These Cayley-Klein spaces are the objects of a Cayley-Klein category **CK**, with morphisms (196) or (199) which preserve quadratic form.

Let us introduce the notations: $TV^{n+1}(\boldsymbol{j})$ for VKS-trees in Cayley-Klein spaces $\mathbb{K}(\boldsymbol{j})$.

Given Cayley-Klein categories $\mathbf{CK}C$ and $\mathbf{CK}B$, we consider all VKS-tree functors $RV^{n+1}(\boldsymbol{j})$, $SV^{n+1}(\boldsymbol{j}), TV^{n+1}(\boldsymbol{j}), \ldots : \mathbf{CK}C \to \mathbf{CK}B$. If $\sigma : R \rightarrowtail S$ and $\tau : S \rightarrowtail T$, are two natural transformations, their components for each $c \in C$ define composite morphisms $(\tau \cdot \sigma)c \to \tau c \circ \sigma c$ which are the components of a transformation $\tau \cdot \sigma \rightarrowtail T$. To show $\tau \cdot \sigma$ natural, take any $f : c \to c'$ in C and consider the diagram

$$\begin{tikzcd}
Rc \arrow[r, "Rf"] \arrow[d, "\sigma c"'] \arrow[dd, bend right=60, "(\tau\cdot\sigma)c"'] & Rc' \arrow[d, "\sigma c'"] \arrow[dd, bend left=60, "(\tau\cdot\sigma)c'"] \\
Sc \arrow[r, "Sf"] \arrow[d, "\tau c"'] & Sc' \arrow[d, "\tau c'"] \\
Tc \arrow[r, "Tf"] & Tc'
\end{tikzcd}$$

Since σ and τ are natural, both small squares are commutative. Hence the rectangle commutes, so the composite $\tau \cdot \sigma$ is natural.

This composition of transformations is associative; moreover it has for each functor T an identity, the natural transformation $1_T : T \to T$ with components $1_T c = 1_{T_c}$. Hence, given the Cayley-Klein categories $\mathbf{CK}B$ and $\mathbf{CK}C$, we may construct formally a *functor category of VKS-trees* $\mathbf{VKS}|\mathtt{Tree}$ with objects the functors $R, S, T : \mathbf{CK}C \to \mathbf{CK}B$ and morphisms the natural transformations between two such functors.

References

1. Y. I. Manin. *Frobenius Manifolds, Quantum Cohomology, and Moduli Spaces*, volume 47 of *American Mathematical Society Colloquium Publications.* American Mathematical Society, Providence, RI, 1999.

2. Aguiar M., Sottile F. Structure of the Loday-Ronco Hopf Algebra of Trees. Preprint `math.CO/0409022`, 2004.

3. Kuznetsov G.I., Moskaliuk S.S., Smirnov Yu.F., Shelest V.P. *Graphical Theory of Representations of Orthogonal and Unitary Groups and its Physical Applications.* MMM-1. – Kiev: Naukova Dumka Publishers, 1992 (in Russian).

4. Gromov N.A., Moskaliuk S.S., Smirnov Yu.F. *The Tree, Contractions and Analytical Continuations Methods in Representations Theory of Classical Groups and Quantum Algebras.* MMM-2. – Kiev: Naukova Dumka Publishers, 1993 (in Russian).

5. Moskaliuk S.S. *From Cayley-Klein Groups to Categories.* MMM-11. – Kyiv: TIMPANI Publishers, 2006.

6. Günter M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics, 152, Springer-Verlag, 1995.

7. Richard P. Stanley, *Enumerative combinatorics. Vol. 2*, Cambridge University Press, Cambridge, 1999, with a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

8. Getzler E. and Kapranov M.M. Cyclic operads and cyclic homology. In S.-T. Yau, editor, *Geometry, Topology and Physics for Raoul Bott*, volume 4 of *Conf. Proc. Lect. Notes. Geom. Topol.*, pages 167–201. International Press, 1995.

9. Kontsevich M. and Manin Yu. Gromov-Witten classes, quantum cohomology, and enumerative geometry. *Comm. Math. Phys.*, 164:525–562, 1994.

10. Stasheff J.D. Homotopy associativity of H-spaces I,II. *Trans. Amer. Math. Soc.*, 108:275–312, 1963.

11. Graeme Segal, Classifying spaces and spectral sequences, *Institut des Hautes Études Scientifiques Publications Mathématiques* 34 (1968), 105–112.

12. Gabriel P., Zisman M. , *Calculus of Fractions and Homotopy Theory*, Springer, 1967.

13. T. Leinster. *Higher operads, higher categories*, volume 298 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 2004.

14. Gordon R., Power A.J., Street R., *Coherence for Tricategories*, Memoirs of the AMS 117, no. 558, 1995.

15. Křž I., May J. P. *Operads, Algebras, Modules and Motives*, volume 233 of *Astérisque.* Société Mathématique de France, 1995.

16. Markl M., Shnider S., and Stasheff J. D. *Operads in Algebra, Topology and Physics*, volume 96 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, Rhode Island, 2002.

17. May J.P. Definitions: Operads, algebras and modules. In J.L. Loday, J.D Stasheff, and A.A. Voronov, editors, *Operads: Proceedings of Renaissance Conferences*, volume 202 of *Contemporary Math.*, pages 1–7. Am. Math. Soc., 1997.

18. Huang Y.-Z. *Two-dimensional conformal geometry and vertex operator algebras.* Birkhäuser Boston Inc., Boston, MA, 1997.

19. Wells R.O. *Differential Analysis on Complex Manifolds*, volume 65 of *Graduate Texts in Mathematics*. Springer-Verlag, 1980.

20. Deligne P. Resumé des premiérs exposés de A. Grothendieck. In *Groupes de Monodromie en Géometrie Algébrique*, volume 288, pages 1–24. Springer-Verlag, Berlin, 1972.

21. Knudsen F.F. The projectivity of the moduli space of stable curves, II: The stacks $M_{g,n}$. *Math. Scand.*, 52:161–199, 1983.

22. Markl M. Models for operads. *Comm. Algebra*, 24(4):1471–1500, 1996.

23. Boardman J.M. and Vogt R.M. *Homotopy Invariant Algebraic Structures on Topological Spaces*. Springer-Verlag, 1973.

24. May J.P. *The Geometry of Iterated Loop Spaces*, volume 271 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, 1972.

25. Devadoss S. L. Tessellations of moduli spaces and the mosaic operad. In J.P. Meyer, J. Morava, and W.S. Wilson, editors, *Homotopy Invariant Algebraic Structures*, volume 239 of *Contemporary Math.*, pages 91–114, 1999.

26. Etingof P., Henriques A., Kamnitzer J., and Rains E. The cohomology ring of the real locus of the moduli space of stable curves of genus 0 with marked points. Preprint `math.AT/0507514`, August 2005.

27. Hilton P.J. and Stammbach U. *A Course in Homological Algebra*, volume 4 of *Graduate texts in Mathematics*. Springer-Verlag, 1971.

28. Ginzburg V. and Kapranov M.M. Koszul duality for operads. *Duke Math. J.*, 76(1):203–272, 1994.

29. Eilenberg S. and Moore J.C. Adjoint functors and triples. *Illinois J. Math.*, 9:381–389, 1965.

30. Hartshorne R. *Algebraic Geometry*, volume 52 of *Graduate Texts in Mathematics*. Springer-Verlag, 1977.

31. Deligne P. and Mumford D. *The Irreducibility of the Space of Curves of Given Genus*, volume 36 of *Inst. Hautes Études Sci. Publ. Math.* I.H.E.S., 1969.

32. Knudsen F.F. and Mumford D. The projectivity of the moduli space of stable curves I: Preliminaries on “det” and “Div”. *Math. Scand.*, 39:19–55, 1976.

33. Markl M. Loop homotopy algebras in closed string field theory. *Comm. Math. Phys.*, 221:367–384, 2001.

34. Mac Lane S. Categorical algebra. *Bull. Amer. Math. Soc.*, 71:40–106, 1965.

35. Markl M. Cotangent cohomology of a category and deformations. *J. Pure Appl. Algebra*, 113:195–218, 1996.

36. Vallette B. *Dualité de Koszul des PROPs.* PhD thesis, Université Louis Pasteur, 2003.

37. Vallette B. A Koszul duality for props. Preprint `math.AT/0411542`, November 2004.

38. Vallette B. Koszul duality for PROPs. *Comptes Rendus Math.*, 338(12):909–914, June 2004.

39. Markl M. and Voronov A.A. PROPped up graph cohomology. Preprint `math.QA/0307081`, July 2003.

40. Markl M. Deformations and the coherence. *Proc. of the Winter School ‘Geometry and Physics,’ Zdíkov, Bohemia, January 1993, Supplemento ai Rend. Circ. Matem. Palermo, Serie II*, 37:121–151, 1994.

41. Kontsevich M. and Soibelman Y. Deformation theory of bialgebras, Hopf algebras and tensor categories. Preprint, March 2002.

42. Lack S. Composing PROPS. *Theory Appl. Categ.*, 13:No. 9, 147–163 (electronic), 2004.

43. Pirashvili T. On the PROP corresponding to bialgebras. *Cah. Topol. Géom. Différ. Catég.*, 43(3):221–239, 2002.

44. Gan W.L. Koszul duality for dioperads. *Math. Res. Lett.*, 10(1):109–124, 2003.

45. Szabo M.E. Polycategories. *Comm. Algebra*, 3(8):663–689, 1975.

46. Joni S. A. and Rota G.-C.. Coalgebras and bialgebras in combinatorics. *Stud. Appl. Math.*, 61(2):93–139, 1979.

47. Fox T.F. and Markl M. Distributive laws, bialgebras, and cohomology. In J.-L. Loday, J.D. Stasheff, and A.A. Voronov, editors, *Operads: Proceedings of Renaissance Conferences*, volume 202 of *Contemporary Math.*, pages 167–205. Amer. Math. Soc., 1997.

48. Halperin S. *Lectures on Minimal Models*, volume 261 of *Memoirs Soc. Math. France, Nouv. Ser. 9-10.* Soc. Mat. France, 1983.

49. Saneblidze S. and Umble R. Matrons, A_∞-bialgebras and the polytopes KK. Preprint `math.AT/0508017`, August 2005.

50. Saneblidze S. and Umble R. The biderivative and A_∞-bialgebras. Preprint `math.AT/0406270`, April 2004.

51. Saneblidze S. and Umble R. The biderivative, matrons and A_∞-bialgebras. Preprint, July 2004.

52. Shoikhet B. A concept of $\frac{2}{3}$PROP and deformation theory of (co)associative coalgebras. Preprint `math.QA/0311337`, November 2003.

53. Shoikhet B. The CROCs, non-commutative deformations, and (co)associative bialgebras. Preprint `math.QA/0306143`, June 2003.

54. Shoikhet B. An explicit deformation theory of (co)associative bialgebras. Preprint `math.QA/0310320`, October 2003.

55. Markl M. A resolution (minimal model) of the PROP for bialgebras. Preprint `math.AT/0209007`, September 2002.

56. Gugenheim V.K.A.M. and Stasheff J.D. On perturbations and A_∞-structures. *Bull. Soc. Math. Belgique*, 38:237–246, 1986.

57. Merkulov S.A. PROP profile of deformation quantization. Preprint `math.QA/0412257`, December 2004.

58. Fulton R., MacPherson R. A compactification of configuration spaces. *Ann. of Math.*, 139: 183–225, 1994.

59. Keel S. Intersection theory of moduli spaces of stable n-pointed curves of genus zero. *Trans. AMS*, 330: 545–574, 1992.

60. Mac Lane S. Natural associativity and commutativity. *Rice Univ. Stud.*, 49(1):28–46, 1963.

61. Hinich V. and Vaintrob A. Cyclic operads and algebra of chord diagrams. *Selecta Math. (N.S.)*, 8(2):237–282, 2002.

62. Markl M. and Remm E. Algebras with one operation including Poisson and other Lie-admissible algebras. Preprint `math.AT/0412206`.

63. Cartier P. Homologie cycliques: rapport sur des travaux récents de Connes, Karoubi, Loday, Quillen... Séminaire Bourbaki 621, 1983–84.

64. Loday J.-L. and Quillen D. Cyclic homology and the Lie algebra homology of matrices. *Comment. Math. Helv.*, 59:565–591, 1984.

65. Tamarkin D. and Tsygan B. Cyclic formality and index theorems. *Lett. Math. Phys.*, 56:85–97, 2001.

66. Getzler E. and Kapranov M.M. Modular operads. *Compositio Math.*, 110(1):65–126, 1998.

67. Kildyushov M.S., Kuznetsov G.I. Graphic realization of representation of class 1 of orthogonal groups. Moscow, 1973.- 20 p. (Prepr./Int. Atom. Energy; no. 2257) (in Russian).

68. Clifford W.K. *Prelimianry Sketch of Biquaternions*, Proc. London Math. Soc. – 1973, **5**. – P. 235–310.

69. Vilenkin N.Ya. *Special Functions and Theory of Group Representations.* – Amer. Math. Soc. Providence, RI, 1968.

70. Nikiforov A. F., Uvarov V. B. *Special functions of mathematical physics*, Basel: Birkhäuser Verlag, 1988.

71. Celeghini E., Tarlini M. Contractions of group representations. I. *Nuovo Cimento B*, **61**, no. 2: 265–277, 1981.

72. Celeghini E., Tarlini M. Contractions of group representations. II. *Nuovo Cimento B*, **65**, no. 1: 172–180, 1981.

73. Celeghini E., Tarlini M. Contractions of group representations. III. *Nuovo Cimento B*, **68**, no. 1: 133–141, 1982.

74. Celeghini E., Giachetti R., Sorace E., Tarlini M. Contractions of quantum groups. *Lecture Notes in Math.*, **1510**, Berlin: Springer, 1992. P. 221–244.

75. Celeghini E., Giachetti R., Kulish P. P., Sorace E., Tarlini M. Hopf superalgebra contractions and R-matrix for fermions. *J. Phys. A*, **24**, no. 24: 5675–5682, 1991.

76. Gromov N.A., Moskaliuk S.S. Classification of Transition between Groups in Cayley-Klein Spaces and Kinematic Groups, (Preprint/ The Erwin Schrödinger International Institute for Mathematical Physics, Vienna: ESI, 1995, **223**.- 29 p.); *Hadronic Journal*, **19(4)**: 407–435, 1996.

77. G.I. Kuznetsov and S.S. Moskaliuk, "Unitary transformation and matrices of finite rotations in space of trees", *Hadronic Journal* **24** (2001) 17-51.

78. S.S. Moskaliuk "Method of Categorical Extension of Cayley-Klein Groups".// In: Abstracts of XXIVth International Colloquium on Group Theoretical Methods in Physics, July 15–20, 2002 Paris. – P. 50.– *Czechoslovak Journal of Physics*, **55**, no. 11, 2005.- P. 1495-1501.

79. McRae A. S. Clifford algebras and possible kinematics. *SIGMA Symmetry Integrability Geom. Methods Appl.*, **3**, 079: 29 p., 2007 (`math-ph/0707.2869`).

Anatoliy M. Pavlyuk

Generalized Alexander Polynomial Invariants[22]

Bogolyubov Institute for Theoretical Physics
of the National Academy of Sciences of Ukraine,
14b, Metrologichna str. Kyiv 03143, Ukraine
e-mail: pavlyuk@bitp.kiev.ua

We propose an algorithm which allows to derive the generalized Alexander polynomial invariants of knots and links with the help of the $q,p-$numbers, appearing in bosonic two-parameter quantum algebra. These polynomials turn into HOMFLY ones by applying special parametrization. The Jones polynomials can be also obtained by using this algorithm.

1 Introduction

The aim of this paper is to generalize one-parameter Alexander polynomial invariants, one of the main characteristics of knots and links, to two-parameter generalized Alexander polynomial invariants.

[22]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

First, we recall some basic notions of the knot theory. Applying to an initial link (knot) L_+ so called "surgery" operation - elimination of a crossing - we obtain a simpler link/knot L_O. Applying to the same initial link (knot) L_+ another "surgery" operation - switching of a crossing - we obtain another simpler link/knot L_-.

It is postulated:
1) every knot and link is described by the definite polynomial;
2) three concrete polynomials, namely $P_{L_+}(t)$, $P_{L_O}(t)$, $P_{L_-}(t)$ are connected with the help of the following (geometro-algebraic) recurrence relation, which is called the skein relationship:

$$P_{L_+}(t) = l_1 P_{L_O}(t) + l_2 P_{L_-}(t) \tag{1}$$

where l_1, l_2 are coefficients;
3) the normalization condition for the unknot:

$$P_{unknot} = 1\,. \tag{2}$$

Applying the operation of elimination for torus knots and links $L_{n,2}$ turns it into $L_{n-1,2}$, and the switching operation turns it into $L_{n-2,2}$, where n is a positive integer number. From these considerations and from (1) it follows the following recurrence relation :

$$P_{L_{n+1,2}}(t) = l_1 P_{L_{n,2}}(t) + l_2 P_{L_{n-1,2}}(t)\,,$$

or in the simpler notations:

$$P_{n+1,2}(t) = l_1 P_{n,2}(t) + l_2 P_{n-1,2}(t)\,. \tag{3}$$

Thus, the form of the recurrence relation (3) for torus knots and links $L_{n,2}$ coincides with the form of the skein relationship (1).

Recurrence relation only for torus knots $T(2m+1, 2)$ (or only for torus links $L(2m, 2)$), where $m = 0, 1, 2, \dots$. looks as follows:

$$P_{n+2,2}(t) = k_1 P_{n,2}(t) + k_2 P_{n-2,2}(t)\,, \tag{4}$$

where

$$k_1 = l_1^2 + 2l_2, \quad k_2 = -l_2^2. \tag{5}$$

We also have

$$P_{1,2} = 1, \quad P_{3,2} = k_1 + k_2. \tag{6}$$

2 Alexander polynomials

The Alexander polynomials $\Delta(t)$ of knots and links [1] can be defined by the skein relationship

$$\Delta_+(t) - \Delta_-(t) = (t^{\frac{1}{2}} - t^{-\frac{1}{2}})\Delta_O(t), \quad \Delta_{unknot} = 1. \tag{7}$$

From (7) (in analogy to (3)) it follows the following recurrence relation for torus knots and links $L_{n,2}(t)$:

$$\Delta_{n+1,2}(t) = (t^{\frac{1}{2}} - t^{-\frac{1}{2}})\Delta_{n,2}(t) + \Delta_{n-1,2}(t)\,. \tag{8}$$

From (8) (in analogy to (4)) one obtains the recurrence relation only for torus knots $T(2m+1,2)$ (or for torus links $L(2m,2)$)

$$\Delta_{n+2,2}(t) = (t + t^{-1})\Delta_{n,2}(t) - \Delta_{n-2,2}(t)\,. \tag{9}$$

The Alexander polynomials of torus knots $T(n,2)$ can be expressed through q-numbers characteristic to Biedenharn-Macfarlane quantum bosonic oscillator. The bosonic q-number corresponding to an integer n is defined as [2, 3]

$$[n]_q = \frac{q^n - q^{-n}}{q - q^{-1}}\,, \tag{10}$$

where q is a parameter. Some of the q-numbers are:

$$[1]_q = 1\,, \quad [2]_q = q + q^{-1}\,, \quad [3]_q = q^2 + 1 + q^{-2}\,, \quad [4]_q = q^3 + q + q^{-1} + q^{-3}, \ldots\,.$$

The recurrence relation for (10) looks as

$$[n+1]_q = (q + q^{-1})[n]_q - [n-1]_q\,. \tag{11}$$

It was found that [4, 5]:

$$\Delta_{2m+1,2}(t) = [m+1]_t - [m]_t\,, \qquad t \equiv q\,, \tag{12}$$

or, since $n = 2m+1$, as

$$\Delta_{n,2}(t) = \left[\frac{n+1}{2}\right]_t - \left[\frac{n-1}{2}\right]_t. \tag{13}$$

In the following section we generalize these results with the help of $q-$numbers.

3 Algorithm of obtaining of Alexander polynomials from bosonic $q-$numbers

Analyzing the results of previous sections we can formulate an algorithm of obtaining of the Alexander skein relationship (7). Afterwards this procedure will be used for obtaining another skein relations.

First step: we introduce polynomials $A_{n,2}(q)$, which refer to torus knots $T(2m+1,2)$, satisfying following recurrence relation (repeating (11)):

$$A_{n+2,2}(q) = (q+q^{-1})A_{n,2}(q) - A_{n-2,2}(q)\,. \tag{14}$$

According to (6):

$$A_{1,2}(q) = 1, \quad A_{3,2}(q) = q - 1 + q^{-1}. \tag{15}$$

Second step: we formulate full recurrence relation for all polynomials $A_{n,2}(q)$ and, thus, find corresponding skein relationship. From (14) we have $k_1 = q + q^{-1}$, $k_2 = -1$. Because of (5), we find

$$l_1 = q^{\frac{1}{2}} + q^{-\frac{1}{2}}, \quad l_2 = 1\,. \tag{16}$$

Therefore

$$A_{n+1,2}(q) = (q^{\frac{1}{2}} - q^{-\frac{1}{2}})A_{n,2}(q) + A_{n-1,2}(q)\,. \tag{17}$$

From (17) (in anology with (1) and (3)) we obtain the following skein relationship:

$$A_+(q) - A_-(q) = (q^{\frac{1}{2}} - q^{-\frac{1}{2}})A_O(q). \tag{18}$$

Third step: we find an expression for torus knots $A_{2m+1,2}(q)$. In analogy with (19), we put

$$A_{2m+1,2}(q) = a_1(q)[m+1]_q - a_2(q)[m]_t\,, \qquad t \equiv q\ , \tag{19}$$

Using (10), (15) and (19), we find $a_1(q) = 1,\ a_2(q) = 1$. Therefore,

$$A_{2m+1,2}(q) = [m+1]_q - [m]_q \,. \tag{20}$$

In general, we described three-step procedure of obtaining of: 1) skein relationship of knots and links, and 2) expression for polynomial invariants of torus knots $T(2m+1,2)$, from structural functions of bosonic deformed oscillators. In particular, we obtained the formulas (18), (28), which coincides with those for the Alexander polynomial invariants (7), (12) (if $q \equiv t$).

4 Generalized Alexander polynomials $A(q,p)$ from q,p-numbers

In this section we use the proposed three-step algorithm to obtain the generalized Alexander polynomials $A(q,p)$ from q,p-numbers, which reduce to the Alexander polynomials if $p = q^{-1}$.

The q,p-number corresponding to integer number n is introduced as [6]

$$[n]_{q,p} = \frac{q^n - p^n}{q-p} \,, \tag{21}$$

where $q\,,p$ are some complex parameters. If $p = q^{-1}$, then $[n]_{q,p} = [n]_q$. Here are some of the q,p-numbers:

$$[1]_{q,p}{=}1\,,\ [2]_{q,p}{=}q+p\,,\ [3]_{q,p}{=}q^2+qp+p^2\,,\ [4]_{q,p}{=}q^3+q^2p+qp^2+p^3, \dots \ .$$

The recurrence relation for q,p-numbers is

$$[n+1]_{q,p} = (q+p)[n]_{q,p} - qp[n-1]_{q,p} \,. \tag{22}$$

First, in anology with previous section, on the base of (22) we introduce polynomials $A_{n,2}(q,p)$, which generalize the Alexander polynomials:

$$A_{n+2,2}(q,p) = (q+p)A_{n,2}(q,p) - qpA_{n-2,2}(q,p) \,. \tag{23}$$

Thus from normalization condition and (6)

$$A_{1,2}(q,p) = 1, \quad A_{3,2}(q,p) = q - qp + p. \tag{24}$$

Second, from (23) it also follows

$$k_1 = l_1^2 + 2l_2 = q + p, \quad k_2 = -l_2^2 = -qp\,.$$

From here one finds

$$l_2 = q^{\frac{1}{2}} p^{\frac{1}{2}}, \quad l_1 = q^{\frac{1}{2}} - p^{\frac{1}{2}}\,,$$

which leads to the generalized Alexander skein relationship [7]:

$$A_+(q,p) = (q^{\frac{1}{2}} - p^{\frac{1}{2}}) A_O(q,p) + q^{\frac{1}{2}} p^{\frac{1}{2}} A_-(q,p). \tag{25}$$

Formula (25) can be written in the form

$$q^{-\frac{1}{4}} p^{-\frac{1}{4}} A_+(q,p) - q^{\frac{1}{4}} p^{\frac{1}{4}} A_-(q,p) = (q^{\frac{1}{4}} p^{-\frac{1}{4}} - q^{-\frac{1}{4}} p^{\frac{1}{4}}) A_O(q,p) \tag{26}$$

By putting $p = q^{-1}$, the generalized Alexander skein relationship turns into the Alexander skein relationship (7).

Third, we take

$$A_{2m+1,2}(q,p) = a_1(q,p)[m+1]_{q,p} - a_2(q,p)[m]_{q,p}\,. \tag{27}$$

From (24) we have $a_1(q,p) = 1,\ a_2(q,p) = qp$. Therefore,

$$A_{2m+1,2}(q,p) = [m+1]_{q,p} - qp[m]_{q,p}\,. \tag{28}$$

5 Generalized Alexander polynomials and HOMFLY polynomials

The HOMFLY polynomial invariants [8] are described by the skein relationship:

$$a^{-1} H_+(a,z) - a H_-(a,z) = z H_O(a,z). \tag{29}$$

Comparing (26) with the HOMFLY skein relationship (29) we obtain

$$a = q^{\frac{1}{4}} p^{\frac{1}{4}}, \quad z = q^{\frac{1}{4}} p^{-\frac{1}{4}} - q^{-\frac{1}{4}} p^{\frac{1}{4}}\,. \tag{30}$$

Substituting this result into (29), one obtains the generalized Alexander skein relationship (26).

6 Generalized Alexander polynomials and Jones polynomials

The Jones polynomial invariants [9] can be defined as

$$t^{-1}V_+(t) - tV_-(t) = (t^{\frac{1}{2}} - t^{-\frac{1}{2}})V_O(t). \tag{31}$$

Comparing (26) with the Jones skein relationship (31), we find that substitution

$$q = t^3, \quad p = t \tag{32}$$

reduces the generalized Alexander polynomials to Jones ones.

According to results of Section 3, the Jones skein relationship (31) can be obtained with the help of the proposed three-step algorithm from $q-$numbers defined as

$$[n]_{q^3,q} = \frac{q^{3n} - q^n}{q^3 - q}\,. \tag{33}$$

Acknowledgement

This research was partially supported by the Special Programme of Division of Physics and Astronomy of NAS of Ukraine.

References

1. J.W. Alexander, Topolgical invariants of knots and links, Trans. Am. Math. Soc. **30** (1928), 275-306.
2. L.C. Biedenharn, The quantum group $SU_q(2)$ and a $q-$analogue of the boson operators, J. Phys. A: Math. Gen. **22** (1989), L873-L878.
3. A.J. Macfarlane, On $q-$analogues of the quantum harmonic oscillator and the quantum group $SU_q(2)$, J. Phys. A: Math. Gen., **22** (1989), 4581-4585.
4. A.M. Gavrilik, $q-$Serre relations in $U_q(u_n)$, $q-$deformed meson mass sum rules, and Alexander polynomials, J. Phys. A: Math. Gen. **27** (1994), L91-L94.

5. A.M. Gavrilik, A.M. Pavlyuk, Alexander polynomial invariants of torus knots $T(n, 3)$ and Chebyshev polynomials, Ukr. J. Phys. **56** (2011), 680-687; arXiv:1107.5516v1 [math-ph].
6. A. Chakrabarti, R. Jagannathan, A (p, q)oscillator realization of two-parameter quantum algebras, J. Phys. A: Math. Gen. **24** (1991), L711-L718.
7. A.M. Gavrilik, A.M. Pavlyuk, On Chebyshev polynomials and torus knots, Ukr. J. Phys. **55** (2010), 129-134; arXiv:0912.4674v2 [math-ph].
8. P. Freyd, D. Yetter, J. Hoste, W.B.R. Lickorish, K. Millet, A. Ocneanu, A new polynomial invariant of knots and links, Bull. AMS **12** (1985), 239-246.
9. V.F.R. Jones, A polynomial invariant of knots and links via von Neumann algebras, Bull. AMS **12** (1985), 103-111.

Igor Salom[23]

Representations of Parabose Supersymmetry

Institute of Physics, University of Belgrade,
118, Pregrevica, Belgrade, Serbia
e-mail: isalom@ipb.ac.rs

Parabose symmetry (alternative names: Generalized conformal supersymmetry with tensorial central charges, conformal M-algebra, $osp(1|2n)$ supersymmetry) has been considered as an alternative to d-dimensional conformal superalgebra. Potential relevance of the corresponding superalgebra spreads to various subfields of High Energy Physics and Astrophysics (e.g. particle classification, gauging gravity, dark matter/energy candidates, etc.). Yet, due to mathematical difficulties, even classification and analysis of its unitary irreducible representations (UIR's) have not been entirely accomplished. We complete this classification for $n = 4$ case (corresponding to four dimensional space-time) and then show how the discrete subset of these UIR's can be constructed in a less abstract manner, that allows natural physical interpretation as spaces of particular composite particle states.

[23]This work was based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity*, Kyiv (Ukraine), April 29–June 15, 2012, and supported in parts by the Project-ON171031 of Ministry of Education, Science and Technological Development, Serbia and the Project-1202.094-12 of the Central European Initiative Cooperation Fund.

We also conjecture generalization of the obtained results to the cases relevant in the string/brane context ($n > 4$).

1 Introduction

In the standard Poincaré supersymmetry, anticommutator of two lefthanded (righthanded) supersymmetry generators either vanishes or, in the extended supersymmetry case, equals to a central charge. If this requirement is relaxed, in four space-time dimensions the following relations are obtained (in four component spinor notation):

$$\{Q_\alpha, Q_\beta\} = (C\gamma^\mu)_{\alpha\beta}P_\mu + (C\gamma^{\mu\nu})_{\alpha\beta}Z_{\mu\nu}, \{QQcovariantly\} \tag{1}$$

with C being the charge conjugation matrix, $\gamma_{\mu\nu} = [\gamma_\mu, \gamma_\nu]$, space-time indices take values $\mu, \nu = 0, 1, 2, 3$ and spinorial indices $\alpha, \beta = 1, 2, 3, 4$. The nonstandard second term on the righthand side contains six entities $Z_{\mu\nu}$ known as "tensorial central charges".
This sort of supersymmetry generalization conveys also to the superconformal case, introducing, as we will see, a number of additional bosonic generators into the algebra. The superconformal generalization turns out to form $osp(1|8)$ superalgebra, whose enveloping algebra coincides with the, so called, $n = 4$ parabose algebra [1, 2].
Historically, first to notice interesting properties of such a construct seems to have been C. Fronsdal [3], as early as in 1985, while investigating Penrose twistors and conformal field theory. He noticed that reduction from $osp(1|8)$ symmetry to conformal symmetry of Minkowski space ($osp(1|8) \supset su(2, 2)$) can be seen as a specific type of Kaluza-Klein reduction from 10 to 4 dimensions that leads to model with infinite tower of massless fields with increasing spins. Since then the construct of generalized supersymmetry reappeared, sometimes independently, in many physical contexts. In particular, it gained lot of interest when it was realized that tensorial central charges in higher dimensions appear naturally in relation to extended objects, such as branes and that it seems to be the underlying symmetry of M-theory [4–9]. Besides, exotic BPS particles were found and studied [10–13] in this framework, and field equations corresponding to higher spin fields were obtained [14–19]. Independently, generalized conformal supersymmetry showed

up as the result of a search for mathematically simple structures that could contain Poincaré symmetry and thus could be interesting as candidates for a larger space-time symmetry. The approach was based on Heisenberg [20, 21], bose and parabose [22–24] algebras.

In the first place we will be interested in the orthosymplectic generalization of supersymmetry as a candidate for a realistic symmetry of the space-time. This means that we will consider case $osp(1|8)$ that is related to four space-time dimension, but we also conjecture generalization of the results to higher dimensional cases (where $osp(1|2n)$ algebra appears in the context of branes and M-theory).

When considering a (super)group in the context of a space-time symmetry, one of the first and most natural steps to undertake is to find unitary irreducible representations (UIR's) of the group, as these give us basic information on the particle content of the free theory. In principle, only then one can know what types of fields can exist in the model, and is entitled to consider field theory, write action for the fields, attempt quantisation and/or introduce interactions. Yet, in spite of substantial interest in this type of generalized supersymmetry, no complete analysis of unitary irreducible representations, especially in this physical context, has been carried out. The probable reason is that this task is related with substantial mathematical difficulties.

The problems have been solved for low n cases: apart from the well understood case $n = 1$, even UIR's of $n = 2$ were successfully classified [25] and some families explicitly constructed [26]. We are familiar with only a few partial results pertaining to the representations of the $osp(1|2n)$ for $n > 2$ (for a brief review of the progress in the representation theory of the orthosymplectic superalgebras $osp(m|2n)$ in general, see [27]). Günayadin applied his oscillator construction to obtain some positive energy UIR's of $osp(1|2n)$ from discrete spectrum [28]. However, his approach was constructional and thus lacking in a few ways: no classification of UIR's was given, the question if there are more discrete UIR's was left open and there was no insight where is the limit of the continuous spectre. Taking parabosonic approach Lievens, Stoilova, and Van der Jeugt [29] obtained a narrow subclass of positive energy UIR's, called representations with unique vacuum (parastatistics terminology). To the best of our knowledge, the only systematic and general approach to the classification of (positive energy) $osp(1|2n)$ UIR's was attempted by Dobrev and Zhang [30], who analyzed reducibility of lowest weight Verma modules. Yet, it turned out that a complete classification of positive energy UIR's of $osp(1|8)$, at the present level of our mathematical understanding of Verma module structure, required some extremely lengthy and involved calculations that could be only performed by using computers. We thus followed the approach of Dobrev and Zhang, but developed computer algorithms to analyze Verma module structure: to search for singular and subsingular vectors and check their dependencies in each particular case. In this way we managed to make a complete list of positive energy $osp(1|8)$ UIR's, together with explicit forms of the corresponding Verma module singular and subsingular vectors. We demonstrate that there is a concrete number of discrete UIR families (precisely nine, or ten if the trivial representation is counted as a separate class), that physically should be related to elementary particles of $osp(1|8)$ models.

In addition, we also propose a method to explicitly construct discrete representations, allowing one to easily perform concrete calculations in these spaces and, in that way, give physical interpretation to the states within. The method is based on a specific generalization of the, so called, Green's ansatz (used in the context of parastatistics), but in such a way that no anticommuting operators appear when representing superalgebra elements. Curiously, it turns out that to realise all discrete families of UIR's, elementary Green's ansatz representations have to be grouped in pairs, and it takes exactly up to three such pairs to construct arbitrary discrete UIR. It is quite probably that our method for construction of representations can be connected with the one in [28], but, to our opinion, is advantageous due to lack of anticommuting operators that drastically simplifies calculations (and allows us to directly use mathematical machinery developed for non relativistic quantum mechanics).

2 Parabose algebra $n = 4$ as generalized superconformal symmetry

Parabose algebra is a generalization of the algebra of standard bose creation and annihilation operators, first suggested by H.S.Green [1]. In literature [1, 31], it is usually defined as algebra of n pairs of mutually hermitian conjugate operators $a_\alpha, a_\alpha^\dagger$, satisfying trilinear relations:

$$[\{a_\alpha, a_\beta^\dagger\}, a_\gamma] = -2\delta_{\beta\gamma}a_\alpha, \tag{2}$$

$$[\{a_\alpha, a_\beta\}, a_\gamma] = 0, \tag{3}$$

together with relations (additional four) that follow from these by hermitian conjugation and by use of Jacobi identities.[24]

Parabose operators, defined as above, together with all possible anticommutators $\{a_\alpha, a_\beta\}$, $\{a_\alpha, a^\dagger_\beta\}$ and $\{a^\dagger_\alpha, a^\dagger_\beta\}$ of the parabose operators, form a realization of orthosymplectic superalgebra $osp(1|2n)$. With the usual assumptions of positivity of Hilbert space metrics in the space where parabose operators act, list of unitary irreducible representation of parabose algebra reduces to, so called, "positive energy" class of $osp(1|2n)$ UIR's.

As announced in the introduction, we are primarily interested in the case of 4 physical dimensions, corresponding to $n = 4$.

Conformal ($c(1,3) \sim so(2,4)$) algebra is contained in the algebra closed by all anticommutators of parabose operators. We will demonstrate the connection by making a two-step change of basis. We first switch from operators a_α and $a^\dagger_\alpha$ to their hermitian combinations $S^\alpha \equiv (a_\alpha + a^\dagger_\alpha)$ and $Q_\alpha \equiv -i(a_\alpha - a^\dagger_\alpha)$. In the space of all anticommutators of S^α and Q_α we then introduce the following basis:

$$\begin{aligned}
&J_i \equiv \tfrac{1}{8}(\sigma_i)^\alpha{}_\beta \{Q_\alpha, S^\beta\}, && Y_{\underline{i}} \equiv \tfrac{1}{8}\left(\tau_{\underline{i}}\right)^\alpha{}_\beta \{Q_\alpha, S^\beta\}, && N_{\underline{i}j} \equiv \tfrac{1}{8}\left(\alpha_{\underline{i}j}\right)^\alpha{}_\beta \{Q_\alpha, S^\beta\},\\
&K_{\underline{i}j} \equiv -\tfrac{1}{8}\left(\alpha_{\underline{i}j}\right)_{\alpha\beta} \{S^\alpha, S^\beta\}, && K_0 \equiv \tfrac{1}{8}(\alpha_0)_{\alpha\beta} \{S^\alpha, S^\beta\}. && \\
&D \equiv \tfrac{1}{8}(\alpha_0)^\alpha{}_\beta \{Q_\alpha, S^\beta\}, && P_{\underline{i}j} \equiv \tfrac{1}{8}\left(\alpha_{\underline{i}j}\right)^{\alpha\beta} \{Q_\alpha, Q_\beta\}, && P_0 \equiv \tfrac{1}{8}(\alpha_0)^{\alpha\beta} \{Q_\alpha, Q_\beta\},
\end{aligned} \tag{4}$$

Matrices $\sigma_i, \tau_{\underline{i}}, \alpha_{\underline{i}j}$ and α_0, appearing here, represent a basis of four by four real matrices, defined as follows. Basis for antisymmetric matrices is given by six matrices σ_i and $\tau_{\underline{i}}$, $i, \underline{i} = 1,2,3$ that satisfy:

$$[\sigma_i, \sigma_j] = 2\varepsilon_{ijk}\sigma_k, \quad [\tau_{\underline{i}}, \tau_{\underline{j}}] = 2\varepsilon_{\underline{i}\underline{j}\underline{k}}\tau_{\underline{k}}, \quad [\sigma_i, \tau_{\underline{j}}] = 0. \tag{5}$$

Matrices $\alpha_{\underline{i}j} \equiv \tau_{\underline{i}}\sigma_j$, together with the unit matrix denoted as α_0, form a basis of symmetric matrices.

Algebra closed by parabose anticommutators, whose one particular basis is given by (4), has 36 generators and is isomorphic to $sp(8)$. Centralizer of element $Y_{\underline{i}}$ ($\underline{i}$ arbitrary) is a subalgebra isomorphic to Conformal algebra of Minkowski spacetime ($so(2,4) \subset sp(8)$) plus the element $Y_{\underline{i}}$ alone. Without loss of generality, we

[24] We note that, in a Hilbert space equipped with positive definite metrics (with respect to which one defines the adjoint $a^\dagger_\alpha$), all algebra relations actually follow from a single relation (2).

will consider centralizer of $Y_{\underline{3}}$, spanned by the operator $Y_{\underline{3}}$ itself and the operators:

$$J_k, N_i \equiv N_{\underline{3}i}, D, P_i \equiv P_{\underline{3}i}, P_0, K_i \equiv K_{\underline{3}i}, K_0, \tag{6}$$

that generate $so(2,4)$ algebra. Operators (6) play the roles of rotation generators, boost generators, dilatation generator, momenta and pure conformal generators, respectively.

We have thus demonstrated that the group generated by anticommutators of $n=4$ parabose algebra can be seen as a particular generalization, that is, extension of the conformal symmetry group in four dimensions. If we additionally include the parabose operators Q and S themselves in the even algebra, the overall structure becomes an extension of conformal superalgebra (hence the name generalized conformal supersymmetry). Mathematically, algebra extends from $sp(8)$ to $osp(1|8)$. Operators Q and S play roles of space-time supersymmetry generators. To see this we can "invert" relations (4):

$$\begin{aligned} &\{Q_\alpha, Q_\beta\} = (\alpha_0)_{\alpha\beta}\, P_0 + (\alpha_{\underline{i}j})_{\alpha\beta}\, P_{\underline{i}j}, \{S^\alpha, S^\beta\} = (\alpha_0)^{\alpha\beta}\, K_0 - (\alpha_{\underline{i}j})^{\alpha\beta}\, K_{\underline{i}j}, \\ &\{S^\alpha, Q_\beta\} = (\alpha_0)^\alpha{}_\beta\, D + (\alpha_{\underline{i}j})^\alpha{}_\beta\, N_{\underline{i}j} + (\sigma_i)^\alpha{}_\beta\, J_i + (\tau_{\underline{i}})^\alpha{}_\beta\, Y_{\underline{i}}. \end{aligned} \tag{7}$$

Comparison of these relations with the standard conformal superalgebra relations shows appearance of extra terms on righthand sides of (7) – these are exactly the tensorial central charges from relation (1), written in a different, Lorentz non-covariant notation. In the first of the relations, apart from the expected operators $P_{\underline{3}i}$ and P_0 that we have identified with spatial momentum and energy (6), there are additional operators $P_{\underline{1}i}$ and $P_{\underline{2}i}$. These operators transform as components of a second rank antisymmetric Lorentz tensor and are linear combinations of anticommutators $\{Q_\eta, Q_\xi\}$ and $\{\overline{Q}_{\dot\eta}, \overline{Q}_{\dot\xi}\}$ (that vanish by definition in the standard supersymmetry case).

3 Unitary irreducible representations

In this section we classify unitary irreducible representations of $n=4$ parabose algebra. We will begin with some basic observations.

As the metrics is positive definitive, an operator defined as $E \equiv \frac{1}{2}\sum_\alpha \{a_\alpha, a_\alpha^\dagger\}$ must be positive. Annihilation operators a_α reduce the eigenvalue of E, thus the Hilbert space must contain a subspace that these operators annihilate. This subspace is called vacuum subspace: $V_0 = \{|v\rangle, a_\alpha|v\rangle = 0\}$. From the parabose algebra relations follows:

$$|v\rangle \in V_0 \Rightarrow \{a_\alpha, a_\beta^\dagger\}|v\rangle \in V_0, \tag{8}$$

with α, β arbitrary. Therefore vacuum subspace carries a representation of an $U(1)\times SU(N)$ group generated by operators $\{a_\alpha, a_\beta^\dagger\}$ (with $U(1)$ part generated by E). Let $V_0^{(\mu)}$ be a subspace of V_0 carrying irreducible representation μ of $SU(N)$. For the reasons of unitarity we are interested in cases when this subspace is finite dimensional. Since generators $\{a_\alpha, a_\beta^\dagger\}$ commute with E, E acts as a multiple of unity in this subspace and its eigenvalue will be denoted as e_0. Therefore, we can uniquely label $V_0^{(\mu)}$ as $V_0^{(\mu,e_0)}$, and the parameters μ and e_0 in this way also label UIR's of parabose algebra. In the context of $osp(1|2n)$ algebra such representations are called positive energy UIR's. In analysis of this type of $osp(1|2n)$, or more concretely, of $osp(1|8)$ unitary irreducible representations we closely followed the approach from [30]: not only in method (analysis of reducibility and unitarity conditions for lowest weight Verma modules), but also in conventions, choice of root system, UIR labels, et cetera (only different letters will be sometimes used to denote quantities, in order to ensure compatibility with the rest of this paper). Thus we will run through preliminaries very briefly, referring to [30] for details.

We consider lowest weight Verma modules $V^\Lambda \cong U(\mathcal{G}^+) \otimes |v_0\rangle$. Here, $\mathcal{G}^+$ denotes subalgebra of positive roots in standard algebra decomposition $\mathcal{G}^{\mathbb{C}} = \mathcal{G}^+ \oplus \mathcal{H} \oplus \mathcal{G}^-$ ($\mathcal{G}$ denotes superalgebra $osp(1|8)$ and $\mathcal{G}^{\mathbb{C}}$ its complexification; $\mathcal{H}$ is Cartan subalgebra) and $|v_0\rangle$ is a lowest weight vector of weight Λ:

$$X \in \mathcal{G}^- \Rightarrow X|v_0\rangle = 0, \quad H \in \mathcal{H} \Rightarrow H|v_0\rangle = \Lambda(H)|v_0\rangle. \tag{9}$$

Roots, expressed using elementary functionals, are:

$$\begin{aligned} \Delta \;=\; & \{\pm\delta_\alpha, 1 \leq \alpha \leq 4; \pm\delta_\alpha \pm \delta_\beta, 1 \leq \alpha < \beta \leq 4; \\ & \pm 2\delta_\alpha, 1 \leq \alpha \leq 4\} \end{aligned} \tag{10}$$

(the two signs in $\pm\delta_\alpha \pm \delta_\beta$ not being correlated) and the corresponding root vectors we will denote as (in the same order):

$$\begin{aligned} \mathcal{G}^+ \oplus \mathcal{G}^- \;=\; & \{a_{\pm\alpha}^\dagger, 1 \leq \alpha \leq 4; a_{\pm\alpha,\pm\beta}^\dagger, 1 \leq \alpha < \beta \leq 4; \\ & a_{\pm\alpha,\pm\alpha}^\dagger, 1 \leq \alpha \leq 4\}. \end{aligned} \tag{11}$$

Here we introduced a compact notation for superalgebra elements, that emphasises the parabose connection:

$$a^\dagger_{-\alpha} \equiv a_\alpha, \quad a^\dagger_{\alpha,\beta} \equiv \{a^\dagger_\alpha, a^\dagger_\beta\}. \tag{12}$$

Simple root vectors are:

$$\{a^\dagger_{-2,1}, a^\dagger_{-3,2}, a^\dagger_{-4,3}, a^\dagger_4\} \tag{13}$$

and the corresponding positive root vectors are:

$$\begin{aligned}\Delta^+ = \ & \Big\{a^\dagger_4, a^\dagger_{1,4}, a^\dagger_{2,4}, a^\dagger_{3,4}, a^\dagger_3, a^\dagger_{1,3}, a^\dagger_{2,3}, a^\dagger_2, a^\dagger_{1,2}, a^\dagger_1, \\ & a^\dagger_{-4,3}, a^\dagger_{-4,2}, a^\dagger_{-3,2}, a^\dagger_{-4,1}, a^\dagger_{-3,1}, a^\dagger_{-2,1}\Big\},\end{aligned} \tag{14}$$

written in, so called, normal ordering [30] that we will use for ordering of the Poincaré-Birkhoff-Witt (PBW) basis of $U(\mathcal{G}^+)$.

We will label representations by the signature

$$\chi = \{s_1, s_2, s_3, d\}, \tag{15}$$

where parameters s_1, s_2, s_3 actually label the $su(4)$ representation μ and parameter d is related to e_0 by $e_0 = 4d + s_1 - s_3$. The connection between the signature and the lowest weight Λ is given by:

$$\begin{aligned}\Lambda = \ & (d - \frac{s_1}{2} - \frac{s_2}{2} - \frac{s_3}{2})\delta_1 + (d + \frac{s_1}{2} - \frac{s_2}{2} - \frac{s_3}{2})\delta_2 \\ & + (d + \frac{s_1}{2} + \frac{s_2}{2} - \frac{s_3}{2})\delta_3 + (d + \frac{s_1}{2} + \frac{s_2}{2} + \frac{s_3}{2})\delta_4.\end{aligned} \tag{16}$$

A corresponding shortened notation will be also used for weights: $\Lambda = (\frac{2d-s_1-s_2-s_3}{2}, \frac{2d+s_1-s_2-s_3}{2}, \frac{2d+s_1+s_2-s_3}{2}, \frac{2d+s_1+s_2+s_3}{2})$.

We introduce a (Shapovalov) norm on the Verma module via natural involutive antiautomorphism: $\omega : \omega(a_\alpha) = a^\dagger_\alpha$ (compatible with the assumed Hilbert space metric). Right away we note that simple unitarity considerations – calculating norms of vectors $a^\dagger_{-(\alpha+1),\alpha}|v_0\rangle$ and $a^\dagger_1|v_0\rangle$ – result in constraints: $s_1 \geq 0, s_2 \geq 0, s_3 \geq 0$, $d \geq (s_1 + s_2 + s_3)/2$. Parameters s_1, s_2, s_3 must be integer, labelling an $SU(4)$ Young tableau with $s_1 + s_2 + s_3$ boxes in the first row, $s_1 + s_2$ boxes in the second and s_1 boxes in the third row.

For certain values of Λ submodules appear in the structure of the Verma module V^Λ, and the module becomes reducible. Basic case is when this happens due to existence of a singular vector $|v_s\rangle \in V^\Lambda$:

$$X|v_s\rangle = 0, \qquad \forall X \in \mathcal{G}^-. \tag{17}$$

This singular vector, in turn, generates a submodule $V^{\Lambda'} \cong U(\mathcal{G}^+)|v_s\rangle$ within V^Λ.

To ensure irreducibility, all submodules corresponding to singular vectors must be factored out. However, after factoring out these submodules, new singular vectors may appear in the remaining space – called subsingular vectors. Namely, if the union of all submodules of singular vectors is denoted by $\tilde{I}^\Lambda$ then a vector $|v_{ss}\rangle \in V^\Lambda$ is called a subsingular vector [32] if $|v_{ss}\rangle \notin \tilde{I}^\Lambda$ and:

$$X|v_{ss}\rangle \in \tilde{I}^\Lambda, \qquad \forall X \in \mathcal{G}^-. \tag{18}$$

Just as singular vectors, subsingular vectors also generate submodules that have to be factored out when looking for irreducible representations.

In the particular case of $osp(1|2n)$ there are always, irrespectively of d value, singular vectors of the form:

$$|v_s^\alpha\rangle \equiv (a^\dagger_{-(\alpha+1),\alpha})^{s_\alpha+1}|v_0\rangle, \quad \alpha = 1, 2, \ldots n-1, \tag{19}$$

(when considering cases of unitary and therefore finite dimensional $SU(n)$ representations μ, related to integer values of s_α). The union of the submodules corresponding to these singular vectors we will denote as I^Λ_{SU}. We will always consider factor modules V^Λ/I^Λ_{SU}, and due to this fact subsingular vectors will play a significant role in the the analysis.

Our analysis of the Verma module structure heavily relied on the computer analysis and was carried out in the following general manner (that we just briefly describe). First, Kac determinant of a sufficiently high level was considered as a function of parameter d (for each given class of $SU(4)$ representation μ). In this way it was possible to locate the highest value of d for which the determinant vanishes and the Verma module becomes reducible. The singular or subsingular vector responsible for the singularity of the Kac matrix was then calculated, effectively by solving an (optimized) system of linear equations. Next we would find the norm of this vector and look for possible additional discrete reduction points at (lower) values of d for which the norm also vanishes. If new reduction points with new (sub)singular vectors were found it was also necessary to check that, upon removal of the corresponding submodules, no vectors with zero or negative norm remained. For this, it was enough to check that previously found (sub)singular vectors (i.e. those occurring for higher d values) belonged to the factored-out submodules. Optimized Wolfram Mathematica code was written to perform all these calculations.

We will illustrate the procedure on a few cases, and then give the final classification. More detailed account of the (sub)singular vectors and their interrelations will be given elsewhere.

First we consider unitary irreducible representations that appear when μ is the trivial representation ($s_1 = s_2 = s_3 = 0$), i.e. cases when the lowest weight vector of Verma module is invariant w.r.t. $SU(4)$ subgroup action (space V_0 is one dimensional). The structure of the Verma module in this case is as follows.

For values $d > \frac{3}{2}$ the Verma module is irreducible, all norms are positive and the corresponding representations are unitary and irreducible.

At value $d = \frac{3}{2}$ a subsingular vector appears. In PBW basis this vector has form:

$$\begin{aligned}|v_{ss}^{(1,1,1,1)}\rangle = \; & (-2a^\dagger_{3,4}a^\dagger_2a^\dagger_1 + 2a^\dagger_{2,4}a^\dagger_3a^\dagger_1 - 2a^\dagger_4a^\dagger_{2,3}a^\dagger_1 - 2a^\dagger_{1,4}a^\dagger_3a^\dagger_2 + 2a^\dagger_4a^\dagger_{1,3}a^\dagger_2 \\ & - 2a^\dagger_4a^\dagger_3a^\dagger_{1,2} + a^\dagger_{3,4}a^\dagger_{1,2} - a^\dagger_{2,4}a^\dagger_{1,3} + a^\dagger_{1,4}a^\dagger_{2,3} + 4a^\dagger_4a^\dagger_3a^\dagger_2a^\dagger_1)|v_0\rangle.\end{aligned}$$

The notation for labeling these (sub)singular vectors is the following: ss in the lower index stands for "subsingular" whereas s means "singular" vector; in the upper index we give "relative weight" of the vector – if the (sub)singular vector generates Verma submodule of weight Λ' the the relative weight is $\Lambda' - \Lambda$ (the relative weight alone will turn out to uniquely label these vectors, in a very systematic way).

Upon removing, i.e. factoring out the submodule generated by this vector, an UIR is obtained.

The norm of the vector $|v_{ss}^{(1,1,1,1)}\rangle$ as a function of d at $s_1 = s_2 = s_3 = 0$ is $64(2d-3)(d-1)(2d-1)d$, having zeros at $d = \frac{3}{2}, 1, \frac{1}{2}$ and 0.

Between $d = \frac{3}{2}$ and $d = 1$ the norm above is negative and there are no UIR's. However, at the value $d = 1$ a new subsingular vector appears:

$$|v_{ss}^{(0,1,1,1)}\rangle = \left(a^\dagger_{3,4}a^\dagger_2 - a^\dagger_{2,4}a^\dagger_3 + a^\dagger_4a^\dagger_{2,3} - 2a^\dagger_4a^\dagger_3a^\dagger_2\right)|v_0\rangle\,. \qquad (20)$$

It can be explicitly shown that the subsingular vector $|v_{ss}^{(1,1,1,1)}\rangle$ belongs to the union of submodule generated by $|v_{ss}^{(0,1,1,1)}\rangle$ and the submodule I_{SU}^{Λ}. After factoring out submodule of the vector $|v_{ss}^{(0,1,1,1)}\rangle$ no negative or zero norm vectors remain in the factor space and an UIR is obtained for $d = 1$, $s_1 = s_2 = s_3 = 0$.

Norm of the subsingular vector (20) is $16(d-1)\ (2d-1)d$. In particular, it is negative for $1 > d > \frac{1}{2}$, precluding existence of UIR's in this range.

At $d = \frac{1}{2}$ a singular vector appears:

$$|v_s^{(0,0,1,1)}\rangle = (a_4^\dagger a_4^\dagger a_{-4,3}^\dagger - a_{3,4}^\dagger + 2a_4^\dagger a_3^\dagger)|v_0\rangle, \tag{21}$$

with norm $8(2d-1)d$.

The previous subsingular vector $|v_{ss}^{(0,1,1,1)}\rangle$ belongs to the union of submodule generated by $|v_s^{(0,0,1,1)}\rangle$ and submodule I_{SU}^{Λ}. Thus, there is UIR also at $d = 1/2$, $s_1 = s_2 = s_3 = 0$ obtained upon removing the submodule of vector $|v_s^{(0,0,1,1)}\rangle$.

Norm of $|v_s^{(0,0,1,1)}\rangle$ is negative when $\frac{1}{2} > d > 0$ and, therefore, there are no UIR's in this range.

At $d = 0$ another subsingular vector, of the norm $2d$, appears:

$$|v_s^{(0,0,0,1)}\rangle = a_4^\dagger|v_0\rangle. \tag{22}$$

This reduction point corresponds to the trivial representation of $osp(1|8)$ with representation space being spanned only by vector $|v_0\rangle$.

Proceeding in the same manner, we finally obtain the following simple scheme for $n = 4$ parabose UIR classification:

- $s_1 = s_2 = s_3 = 0$:

$$\begin{array}{l} d > 3/2; \\ d = 3/2, |v_{ss}^{(1,1,1,1)}\rangle; \\ d = 2/2, |v_{ss}^{(0,1,1,1)}\rangle; \\ d = 1/2, |v_s^{(0,0,1,1)}\rangle; \\ d = 0/2, |v_s^{(0,0,0,1)}\rangle; \end{array} \tag{23}$$

- $s_1 = s_2 = 0, s_3 > 0$:

$$\begin{array}{l} d > s_3/2 + 4/2; \\ d = s_3/2 + 4/2, |v_{ss}^{(1,1,1,0)}\rangle; \\ d = s_3/2 + 3/2, |v_s^{(0,1,1,0)}\rangle; \\ d = s_3/2 + 2/2, |v_s^{(0,0,1,0)}\rangle; \end{array} \tag{24}$$

- $s_1 = 0, s_2 > 0$:

$$\begin{aligned} &d > (s_2+s_3)/2+5/2;\\ &d = (s_2+s_3)/2+5/2, |v_s^{(1,1,0,0)}\rangle;\\ &d = (s_2+s_3)/2+4/2, |v_s^{(0,1,0,0)}\rangle; \end{aligned} \tag{25}$$

- $s_1 > 0$:

$$\begin{aligned} &d > (s_1+s_2+s_3)/2+6/2;\\ &d = (s_1+s_2+s_3)/2+6/2, |v_s^{(1,0,0,0)}\rangle. \end{aligned} \tag{26}$$

The pattern of "relative weights" of (sub)singular vectors in the above scheme is obvious, and it allows us to immediately conjecture UIR classification for $n > 4$:

- $s_1 = s_2 = \cdots = s_{n-1} = 0$:

$$\begin{array}{ll} d > (n-1)/2; & \\ d = (n-1)/2, & |v_{ss}^{(1,1,1,\ldots,1,1,1,1)}\rangle;\\ d = (n-2)/2, & |v_{ss}^{(0,1,1,\ldots,1,1,1,1)}\rangle;\\ & \ldots \\ d = 2/2, & |v_{ss}^{(0,0,0,\ldots,0,1,1,1)}\rangle;\\ d = 1/2, & |v_{s}^{(0,0,0,\ldots,0,0,1,1)}\rangle;\\ d = 0/2, & |v_{s}^{(0,0,0,\ldots,0,0,0,1)}\rangle; \end{array} \tag{27}$$

- $s_1 = s_2 = \cdots = s_{n-2} = 0, s_{n-1} > 0$:

$$\begin{array}{ll} d > s_{n-1}/2+(n-1+1)/2; & \\ d = s_{n-1}/2+(n-1)/2, & |v_{ss}^{(1,1,1,\ldots,1,1,0)}\rangle;\\ d = s_{n-1}/2+(n-1-1)/2, & |v_{ss}^{(0,1,1,\ldots,1,1,0)}\rangle;\\ & \ldots \\ d = s_{n-1}/2+4/2, & |v_{ss}^{(0,0,\ldots,1,1,1,0)}\rangle;\\ d = s_{n-1}/2+3/2, & |v_{s}^{(0,0,\ldots,0,1,1,0)}\rangle;\\ d = s_{n-1}/2+2/2, & |v_{s}^{(0,0,\ldots,0,0,1,0)}\rangle; \end{array} \tag{28}$$

- $\ldots$

- $s_1 = 0, s_2 > 0$:

$$\begin{array}{ll} d > (s_2+\cdots+s_{n-1})/2+n-3/2; & \\ d = (s_2+\cdots+s_{n-1})/2+n-3/2, & |v_s^{(1,1,0,\ldots,0,0,0)}\rangle;\\ d = (s_2+\cdots+s_{n-1})/2+n-4/2, & |v_s^{(0,1,0,\ldots,0,0,0)}\rangle; \end{array} \tag{29}$$

- $s_1 > 0$:

$$\begin{aligned} &d > (s_1 + \cdots + s_{n-1})/2 + n - 1; \\ &d = (s_1 + \cdots + s_{n-1})/2 + n - 1, \quad |v_s^{(1,0,0,\ldots,0,0,0)}\rangle. \end{aligned} \tag{30}$$

4 An explicit construction of parabose UIR's

We propose a method to explicitly construct the above classified unitary irreducible representations of parabose algebra. The method cannot be applied to UIR's from the continuous spectre, i.e. those UIR's that occur for non (half)integer values of parameter d. However, from the physical viewpoint, representations from the discrete spectre (d taking discrete (half)integer values less or equal to the first reduction point) are of far greater significance since only in these cases singular or subsingular vectors appear. And it is well known that these vectors turn into important equations of motion (e.g. see [32]). In the particular case of the parabose generalization of supersymmetry, these vectors, for example, turn into Klein-Gordon, Dirac and Maxwell equations.

In the same paper where he first introduced parabose (and parafermi) algebra [1], H.S.Green has also offered a way to construct some of the unitary representations using what is nowadays known as the Green's ansatz. We demonstrate that the ansatz, originally applicable only to "unique vacuum" representations, can also accommodate other representations of the discrete type. We also combine the ansatz with, so called, Klein transformation, so that Green operators no longer satisfy strange "mixed" commutation and anticommutation relations, but instead obey usual commutation relations of bosonic algebra.

We define a Klein transformed analogue of Green's decomposition of order p (p is known as the order of the parastatistics) as the following expression for parabose operators:

$$a_\alpha = \textstyle\sum_{a=1}^{p} I_{(1)} I_{(2)} \cdots I_{(a-1)} a_\alpha^a. \tag{31}$$

In this expression operator a_α^a and its adjoint $a_\alpha^{a\dagger}$ satisfy ordinary bosonic algebra relations. There are total of $n \cdot p$ mutually commuting pairs of annihilation-creation operators $(a_\alpha^a, a_\alpha^{a\dagger})$:

$$[a_\alpha^a, a_\beta^{b\dagger}] = \delta_{\beta\alpha}\delta^{ab}; \quad [a_\alpha^a, a_\beta^b] = 0, \tag{32}$$

where $a, b = 1, 2, \ldots p$ and $\alpha, \beta = 1, 2, \ldots n$.

In (31) we have also introduced selfadjoint unipotent Klein "inversion" operators that act on the Green's operators in the following way:

$$I_{(a)} a^b_\alpha I_{(a)} = (-)^{\delta_{ab}} a^b_\alpha. \tag{33}$$

By their introduction we avoided appearance of anticommuting relations of original Green's operators and, by this, operators a^a_α and $a^{a\dagger}_\alpha$ become familiar mathematical objects which are easier to manipulate and interpret. The easiest way to show that such inversion operators exist is by explicit construction: $I_{(a)} = \exp(i\pi \sum_\alpha \frac{1}{2}\{a^a_\alpha, a^{a\dagger}_\alpha\})$.

The overall Green's ansatz representation space of order p can be seen as tensor product of p multiples of Hilbert spaces $\mathcal{H}_{(a)}$ of ordinary linear harmonic oscillator in n-dimensions: $\mathcal{H} = \mathcal{H}_{(1)} \otimes \mathcal{H}_{(2)} \otimes \cdots \otimes \mathcal{H}_{(p)}$. A single factor Hilbert space $\mathcal{H}_{(a)}$ is the space of unitary representation of n dimensional bose algebra of operators $(a^a_\alpha, a^{a\dagger}_\alpha), \alpha = 1, 2, \ldots n$, which is, at the same time, the simplest nontrivial unitary representation of parabose algebra (i.e. the simplest positive energy UIR of $osp(1|2n)$): $\mathcal{H}_{(a)} \cong U(a^{a\dagger}_\alpha)|0\rangle_a$, where $|0\rangle_a$ is the usual bose vacuum of factor space $\mathcal{H}_{(a)}$. This picture is apropriate due to the fact that the action of even operators of $osp(1|2n)$ (and, in particular, of spacetime symmetry generators (4) in the $n = 4$ case) reduces simply to sum of actions in each of these factor spaces, by virtue of:

$$\{a_\alpha, a_\beta\} = \sum_{a=1}^{p} \{a^a_\alpha, a^a_\beta\}, \quad \{a_\alpha, a^\dagger_\beta\} = \sum_{a=1}^{p} \{a^a_\alpha, a^{a\dagger}_\beta\}. \tag{34}$$

As, from the mathematical point of view, the whole representation space exactly corresponds to Hilbert space of p particles in a n-dimensional non relativistic quantum mechanics, it is very clear that no negative or zero norm states appear. Therefore, if we can find, in this framework, a lowest weight vector $|v_0\rangle$ of a proper weight (corresponding to some UIR signature found in previous section) then the vectors of the form $\mathcal{P}(X)|v_0\rangle, \mathcal{P}(X) \in U(\mathcal{G}^+)$ will span that representation space. In addition, one can explicitly check that the corresponding (sub)singular vector vanishes, as it must.

The simplest nontrivial representation, with signature $s_1 = s_2 = s_3 = 0$, $d = 1/2$ corresponds to $p = 1$ space. The lowest weight vector is simply the vacuum of the $\mathcal{H}_{(1)}$: $|v_0^{(0,0,1,1)}\rangle = |0\rangle_1$. Space in $p = 1$ case is irreducible. Physical interpretation of the vectors in this space is that they correspond to tower of massless states with raising helicities. Other "unique vacuum states (i.e. $s_1 = s_2 = s_3 = 0$) are obtained for $p = 2$ and $p = 3$ with lowest weight vectors being $|0\rangle_1 \otimes |0\rangle_2$ and $|0\rangle_1 \otimes |0\rangle_2 \otimes |0\rangle_3$.

The simplest UIR class of non "unique vacuum" type has signature $s_1 = s_2 = 0, s_3 > 0$, $d = s_3/2 + 1$ and in these representations μ corresponds to single row Young tableaux. This class can be realized in $p = 2$ space, with

$$|v_0^{(0,0,1,0)}\rangle = \tfrac{1}{\sqrt{s_3!}}(A_4^{(1)})^{s_3}|0\rangle_1 \otimes |0\rangle_2, \tag{35}$$

where $A_\alpha^{(k)} \equiv I_{(2k)}(a_\alpha^{2k-1\dagger} + I_{(2k-1)}a_\alpha^{2k\dagger})$. We note that entire $p = 2$ space reduces w.r.t. parabose algebra action to UIR's with signatures: $s_3 = 0, 1, 2, 3, \dots$,$d = s_3/2 + 1$, $s_1 = s_2 = 0$, without any additional degeneracy. From the viewpoint of physics, this is the simplest class that contains both massless and massive states with an additional charge (related to the label s_3).

There are two more classes of "single row" discrete UIR-s: those with signatures $\{0, 0, s_3, \frac{s_3}{2} + \frac{3}{2}\}$ and $\{0, 0, s_3, \frac{s_3}{2} + 2\}$. These are constructed in a similar manner as the previously considered class with signature $\{0, 0, s_3, \frac{s_3}{2} + 1\}$, only in spaces $p = 3$ and $p = 4$, respectively, with the lowest weight states given by expressions:

$$|v_0^{(0,1,1,0)}\rangle = \tfrac{1}{\sqrt{s_3!}}(A_4^{(1)})^{s_3}|0\rangle_1 \otimes |0\rangle_2 \otimes |0\rangle_3, \tag{36}$$

$$|v_0^{(1,1,1,0)}\rangle = \tfrac{1}{\sqrt{s_3!}}(A_4^{(1)})^{s_3}|0\rangle_1 \otimes |0\rangle_2 \otimes |0\rangle_3 \otimes |0\rangle_4. \tag{37}$$

There are two "two-rows" ($s_1 = 0, s_2 > 0$) UIR classes. The class with $d = (s_2 + s_3)/2 + 2$ can be realized in $p = 4$ space, with the lowest weight state given as (up to normalization constant):

$$|v_s^{(0,1,0,0)}\rangle = (A_4^{(1)}A_3^{(2)} - A_3^{(1)}A_4^{(2)})^{s_2}(A_4^{(1)})^{s_3}|0\rangle_1 \otimes |0\rangle_2 \otimes |0\rangle_3 \otimes |0\rangle_4. \tag{38}$$

The remaining class with $d = (s_2 + s_3)/2 + 5/2$ can be realized in $p = 5$ space, with

$$|v_s^{(1,1,0,0)}\rangle = (A_4^{(1)}A_3^{(2)} - A_3^{(1)}A_4^{(2)})^{s_2}(A_4^{(1)})^{s_3}|0\rangle_1 \otimes |0\rangle_2 \otimes |0\rangle_3 \otimes |0\rangle_4 \otimes |0\rangle_5. \tag{39}$$

The only discrete class of representations that corresponds to three-rows Young tableaux ($s_1 > 0$, $d = (s_1 + s_2 + s_3)/2 + 3$) can be realized in $p = 6$ space, with the lowest weight state constructed as (up to normalization constant):

$$
\begin{aligned}
|v_0^{(1,0,0,0)}\rangle \quad = \quad & \Big(\sum_{k,l,m=1}^{3} \varepsilon_{klm} A_2^{(k)} A_3^{(l)} A_4^{(m)} \Big)^{s_1} \\
& \cdot \Big(\sum_{k,l=1}^{2} \varepsilon_{kl} A_3^{(k)} A_4^{(l)} \Big)^{s_2} (A_4^{(1)})^{s_3} \\
& |0\rangle_1 \otimes \cdots \otimes |0\rangle_6,
\end{aligned}
\tag{40}
$$

where ε denotes the Levi-Civita symbol.

Thus we demonstrated a method for realization of all discrete classes of UIR's. The presented construction method can be straightforwardly generalized both to $n > 4$ and to other (half)integer values of d that belong to continuous spectrum.

5 Conclusions

We analyzed $n = 4$ parabose supersymmetry (corresponding to $D = 4$ generalized conformal supersymmetry) using a group-theoretical approach. We gave a complete classification of unitary irreducible representations of parabose algebra. These results, although obtained in the $n = 4$ case, have proved to be readily generalizable to higher values of n, that made the analysis important also in the higher dimensional context of the string theory. Apart from classifying UIR's of the symmetry, we also proposed a method for their explicit construction.

We bring a special attention to the "pairing" of factor spaces that was observable in this setup: to obtain the simplest single box UIR ($s_1 = s_2 = 0, s_3 = 1, d = 3/2$) it takes two factor spaces $p = 2$. To form the simplest UIR with two boxes in a column ($s_1 = s_3 = 0, s_2 = 1, d = 5/2$), it turns out that $p = 4$ must be taken and the vacuum is essentially obtained by antisymmetrizing two "single-box" vacuum states. Similarly, "three-box in a column" UIR ($s_2 = s_3 = 0, s_1 = 1, d = 7/2$) is obtained by antisymmetrizing tensor product of three "single box" vacua. All discrete IR classes can be realized using tensor product of up to three "single-box" $p = 2$ spaces, in a way reminiscent of forming composite particles from simpler constituent ones.

References

1. H. S. Green, *Phys. Rev.* **90**, 270 (1952).

2. A. Ch. Ganchev, T. D. Palev, *J. Math. Phys.* **21**, 797 (1980).

3. C. Fronsdal, Preprint UCLA/85/TEP/10, in "Essays on Supersymmetry", Reidel, 1986 (Mathematical Physics Studies, v. 8).

4. P. K. Townsend *Proc. of PASCOS/Hopkins* (1995) `hep-th/9507048`

5. I. Bars, *Phys. Rev. D* **54** (1996) 5203.

6. J. A. Azcárraga, J. P. Gauntlett, J. M. Izquierdo and P. K. Townsend 1989 *Phys. Rev. Lett.* **63** (1989) 2443

7. P. Townsend *Cargese Lectures* (1997) `hep-th/9712004`

8. J. Lukierski, F. Toppan, *Phys.Lett. B* **539** (2002) 266.

9. S. Ferrara, M. Porrati, *Phys. Lett. B* **423** (1998) 255.

10. E. Sezgin and I. Rudychev, `hep-th/9711128`.

11. S. Fedoruk and V. G. Zima *Mod. Phys. Lett.A* **15** (2000) 2281.

12. I. Bandos and J. Lukierski *Mod. Phys. Lett.A* **14** (1999) 1257.

13. I. Bandos, J. Lukierski, and D. Sorokin, *Phys.Rev. D* **61** (2000) 045002.

14. M. A. Vasiliev, *Phys. Rev. D* **66** (2002) 066006.

15. M. A. Vasiliev, *Nucl.Phys.* **B793** (2008) 469.

16. M. A. Vasiliev, "Relativity, Causality, Locality, Quantization and Duality in the $Sp(2M)$ Invariant Generalized Space-Time", http://arxiv.org/hep-th/01111190111119.

17. M. Plyushchay, D. Sorokin and M. Tsulaia, *JHEP* **0304** (2003) 013. `hep-th/0301067`.

18. I. Bandos et al, *JHEP05* (2005) 031, `hep-th/0501113`.

19. I. Bandos, J. Lukierski and D. Sorokin (1999) "Generalized Superconformal Symmetries and Supertwistor Dynamics" `hep-th/9912051v1`.

20. I. Salom, "Algebraic Approach to the Duality Symmetry", `hep-th/0602282v1`

21. I. Salom, *Jour. Res. Phys* **31** (2007).

22. I. Salom, *Fortschritte der Physik* **56** (2008) 505.

23. I. Salom, "Parabose algebra as generalized conformal supersymmetry", `hep-th/0707.3026v1`

24. I. Salom, Proceedings of the VII international workshop on Lie theory and its applications in physics LT-7, Varna, Bulgaria, 18-24 June 2007, Heron press, Sofia 2008

25. W. Heidenreich, *Phys. Lett.* **110B** (1982) 461/

26. J.Blank and M. Havlicek, *J.Math.Phys.* **29** (1988) 546.

27. C. Luo, "Structure of Polynomial Representations for Orthosymplectic Lie Superalgebras," http://arxiv.org/abs/1001.3474`1001.3474`.

28. M. Günaydin, *Nucl. Phys. B* **528** (1998) 432, http://arxiv.org/abs/hep-th/9803138`hep-th/9803138`.

29. S. Lievens, N. I. Stoilova, J. Van der Jeugt, *Commun. Math. Phys.* **281** (2008) 805.

30. V. K. Dobrev and R. B. Zhang, *Phys. Atomic Nuclei* **68** (2005) 1660.

31. O.W. Greenberg and A.M.L. Messiah, *Phys. Rev.B* **138** (1965) 1155.

32. V. K. Dobrev, *J. Phys. A* **28** (1995) 7135.

Josip Trampetić[25]

Noncommutative Geometry in Quantum Field Theory and its Applications for High Energy Cosmic Rays Experiments[26]

Rudjer Bošković Institute,
54, Bijenička st. HR-10.002 Zagreb, P.O.Box 180, Croatia
e-mail: josipt@rex.irb.hr

Phenomenological analysis of the covariant θ-exact noncommutative (NC) gauge field theory (GFT), inspired by high energy cosmic rays experiments, is performed in the framework of the inelastic neutrino-nucleon scatterings, plasmon and Z-boson decays into neutrino pair, the Big Bang Nucleosynthesis (BBN) and the Reheating Phase After Inflation (RPAI), respectively. Next we have have found neutrino two-point function and shows a closed form decoupling of the hard ultraviolet (UV) divergent term from softened ultraviolet/infrared

[25]This work was supported by the project 098-0982930-2900 of the Croatian Ministry of Science, Education and Sport.

[26]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

(UV/IR) mixing term and from the finite terms as well. For a certain choice of the noncommutative parameter θ which preserves unitarity, problematic UV divergent and UV/IR mixing terms vanish. Non-perturbative modifications of the neutrino dispersion relations are assymptotically independent of the scale of noncommutativity in both the low and high energy limits and may allow superluminal propagation.

Keywords: Noncommutative geometry, Quantum field theory, Neutrino physics, Cosmic ray experiments

1 Introduction

String theory indicated that noncommutative gauge field theory (NCGFT) could be one of its low-energy effective theories [1]. Studies on noncommutative particle phenomenology [2, 3] was motivated to find possible experimental signatures and/or predict/estimate bounds on space-time noncommutativity from collider physics experimental data: for example from the Standard Model (SM) invisible part of $Z \to \bar{\nu}\nu$ decays, and more important from the ultra high energy (UHE) processes occurring in the framework of the cosmic-ray neutrino physics. Constraint on the scale of the NCGFT, $\Lambda_{\rm NC}$, is possible due to a direct coupling of neutrinos to photons.

Significant progress has been obtained in the so-called Seiberg-Witten (SW) maps [1] and enveloping algebra based models where one could deform commutative gauge theories with arbitrary gauge group and representation [4–10]. In our construction the noncommutative fields are obtained via SW maps from the original commutative fields. It is commutative instead of the noncommutative gauge symmetry that is preserved as the fundamental symmetry of the theory. The constraints on the $\mathrm{U}_\star(1)$ charges, stated as “no-go theorem” [11], are also rescinded in our approach [12], and the noncommutative extensions of particle physics covariant SM (NCSM) and the noncommutative grand unified theories (NCGUT) models [10, 12–18] were constructed. These allow a minimal deformation with no new particle content and with the sacrifice that interactions include infinitely many terms defined through recursion over the NC parameter $\theta^{\mu\nu}$; in practice cut-off at certain θ-order.

In a simple model of NC spacetime local coordinates x^μ are promoted to hermitian operators $\hat{x}^\mu$ satisfying spacetime NC and implying uncertainty relations

$$[\hat{x}^\mu, \hat{x}^\nu] = i\theta^{\mu\nu} \longrightarrow |\Delta x^\mu \Delta x^\nu| \geq \frac{1}{2}|\theta^{\mu\nu}|, \tag{1}$$

where $\theta^{\mu\nu}$ is real, antisymmetric matrix. The Moyal-Weyl $\star$-product, relevant for the case of a constant $\theta^{\mu\nu}$, is defined as follows:

$$(f \star g)(x) = e^{\frac{i}{2}h\frac{\partial}{\partial x^\mu}\theta^{\mu\nu}\frac{\partial}{\partial y^\nu}} f(x)g(y)\Big|_{y\to x}. \tag{2}$$

The operator commutation relation (1) is then realized by the so-called $\star$-commutator

$$[\hat{x}^\mu, \hat{x}^\nu] = [x^\mu \stackrel{\star}{,} x^\nu] = i\theta^{\mu\nu}. \tag{3}$$

The perturbative quantization of noncommutative field theories was first proposed in a pioneering paper by Filk [19]. Other famous examples are the running of the coupling constant of NC QED [20] and the UV/IR mixing [21, 22]. Later well behaving one-loop quantum corrections to noncommutative scalar ϕ^4 theories [23–25] and the NC QED [26] have been found. Also the SW expanded NCSM [10, 13, 15, 17] at first order in θ, albeit breaking Lorentz symmetry is anomaly free [27, 28], and has well-behaved one-loop quantum corrections [20–22, 29–37]. However, despite of some significant progress in the models [23–37], a better understanding of various models quantum loop corrections still remains in general a challenging open question. This fact is particularly true for the models constructed by using SW map expansion in the NC parameter θ, [5, 10, 16, 38, 39]. Resulting models are very useful as effective field theories including their one-loop quantum properties [27–37] and relevant phenomenology [40–47].

Discussions on the C,P,T, and CP properties of the noncommutative interactions are given in [44], and in particular in [46]. For example, fixing θ spontaneously breaks C, P, and/or CP discrete symmetries [16]. A breaking of C symmetry occurs in $Z \to \gamma\gamma$ process. One common approximation in those existing works is that only the vertices linear in terms of the NC parameter θ were used.

Quite recently, θ-exact SW map and enveloping algebra based theoretical models were constructed in the framework of covariant noncommutative quantum gauge field theory [4], and applied in loop computation [48–51] and to the phenomenology, as well [52, 53].

At θ-order there are two important interactions that are suppressed and/or forbidden in the SM, the triple neutral gauge boson [13, 15, 17], and the tree level coupling of neutrinos with photons [38, 39], respectively. Here an expansion and cut-off in powers of the NC parameters $\theta^{\mu\nu}$ corresponds to an expansion in momenta and restrict the range of validity to energies well below the NC scale $\Lambda_{\rm NC}$. Usually, this is no problem for experimental predictions because the lower bound on the NC parameters $\theta^{\mu\nu} = c^{\mu\nu}/\Lambda^2_{\rm NC}$ (the coefficients $c^{\mu\nu}$ running between zero and one) runs higher than typical momenta involved in a particular process. However, there are exotic processes in the early universe as well as those involving ultra high energy cosmic rays [47, 52–54] in which the typical energy involved is higher than the current experimental bound on the NC scale $\Lambda_{\rm NC}$. Thus, the previous θ-cut-off approximate results are inapplicable. To cure the cut-off approximation, we are using θ-exact expressions, inspired by exact formulas for the SW map [8, 55, 56], and expand in powers of gauge fields, as we did in [53]. In θ-exact models we have studied the UV/IR mixing [48, 49], the neutrino propagation [50] and also some NC photon-neutrino phenomenology [47, 52–54], respectively. Due to the presence of the UV/IR mixing the θ-exact model is not perturbatively renormalizable, thus the relations of quantum corrections to the observations [57] are not entirely clear.

In this work we present NCSM extended neutrino gauge bosons actions to all orders of θ. Finally we discuss the decay width $\Gamma(Z \to \nu\nu)$ as functions of the NC scale $\Lambda_{\rm NC}$ for light-like noncommutativity which are allowed by unitarity condition [58, 59].

2 UHE cosmic ray motivation

Direct coupling of gauge bosons to neutral and "chiral" fermion particles [38,52,53], via $\star$-commutator in the NC background, which plays the role of an external field in the theory, allow us to estimate a constraint on the scale of the noncommutative guge field theory, $\Lambda_{\rm NC}$, arising from ultra-high energy cosmic ray experiments involving ν-nucleon inelastic cross section, see i.e. Fig. 27. The observation

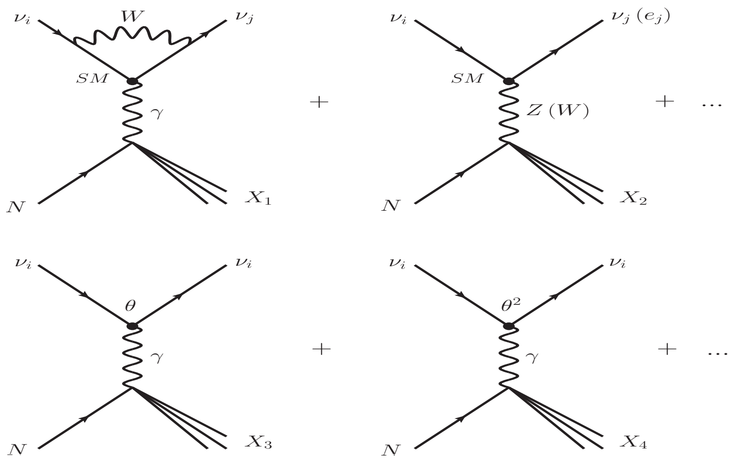


Figure 27: Diagrams contributing to $\nu N \to \nu + X$ processes.

of ultra-high energy (UHE) ν's from extraterrestrial sources would open a new window to look to the cosmos, as such ν's may easily escape very dense material backgrounds around local astrophysical objects, giving thereby information on regions that are otherwise hidden to any other means of exploration. In addition, ν's are not deflected on their way to the earth by various magnetic fields, pointing thus back to the direction of distant UHE cosmic-ray source candidates. This could also help resolving the underlying acceleration in astrophysical sources.

In the energy spectrum of UHE cosmic rays at $\sim 4 \times 10^{19}$ eV the GZK-structure has been observed recently with high statistical accuracy [60]. Thus the flux of the so-called cosmogenic ν's, arising from photo-pion production on the cosmic microwave background $p\gamma_{CMB} \to \Delta^* \to N\pi$ and subsequent pion decay, is now guaranteed to exist. Possible ranges for the size of the flux of cosmogenic ν's can be obtained from separate analysis of the data from various large-scale observatories [61, 62].

Note that there is the uncertainty in the flux of cosmogenic ν's regarding the chemical composition of UHE cosmic rays (for details see [52]). Usinging the upper bound on the νN cross section derived from the RICE Collaboration search results [63] at $E_\nu = 10^{11}$ GeV (4×10^{-3} mb for the FKRT ν-flux [61])), one can infer from θ-truncated model on the NC scale $\Lambda_{\rm NC}$ to be greater than 455 TeV, a really strong bound. Here we have $\theta^{\mu\nu} \equiv c^{\mu\nu}/\Lambda_{\rm NC}^2$ such that the matrix elements of c are of order one. One should however be careful and suspect this result as it has been obtained from the conjecture that the θ-expansion stays well-defined in the kinematical region of interest. Although a heuristic criterion for the validity of the perturbative θ-expansion, $\sqrt{s}/\Lambda_{\rm NC} \lesssim 1$, with $s = 2E_\nu M_N$, would underpin our result on $\Lambda_{\rm NC}$, a more thorough inspection on the kinematics of the process does reveal a more stronger energy dependence $E_\nu^{1/2} s^{1/4}/\Lambda_{\rm NC} \lesssim 1$. In spite of an additional phase-space suppression for small x's in the θ^2-contribution [40] of the cross section relative to the θ-contribution, we find an unacceptably large ratio $\sigma(\theta^2)/\sigma(\theta) \simeq 10^4$, at $\Lambda_{\rm NC} = 455$ TeV. Hence, the bound on $\Lambda_{\rm NC}$ obtained this way is incorrect, and our last resort is to modify the model adequately to include the full-θ resummation, thereby allowing us to compute nonperturbatively in θ.

Total cross section, as a function of the NC scale at fixed $E_\nu = 10^{10}$ GeV and $E_\nu = 10^{11}$ GeV, together with the upper bounds depending on the actual size of the cosmogenic ν-flux (FKRT [61] and PJ [62]) as well as the total SM cross sections at these energies, are depicted in our Figure 28. In order to maximize the NC θ-exact effect we choose $c_{01} - c_{13} = c_{02} - c_{23} = c_{03} = 1$. Even if the future data confirm that UHE cosmic rays are composed mainly of Fe nuclei, as indicated by the PAO data, then still valuable information on $\Lambda_{\rm NC}$ can be obtained with our method, as seen in Fig.29. Here we see the intersections of our curves with the RICE results (cf. Fig.28) as a function of the fraction α of Fe nuclei in the UHE cosmic rays. On top of results, presented in Figs.28 and 29, we also have the NC scale given as a function of the plasmon frequency, from the plasmon decay into neutrino pairs $\gamma_{\rm pl} \to \bar{\nu}\nu$ (Fig.30), and as a function of the T_{dec} from BBN (Fig.31), respectively. All results depicted in Figs.28-31, shows convergent behavior. In our opinion those were the strong signs to continue research towards quantum properties and phenomenology of such θ-exact noncommutative gauge field theory model.

3 Consistency of the SW map and enveloping algebra approach to NCGFT

The choice of gauge group appears to be severely restricted in a noncommutative setting [1]: The star commutator of two Lie algebra valued gauge fields will involve

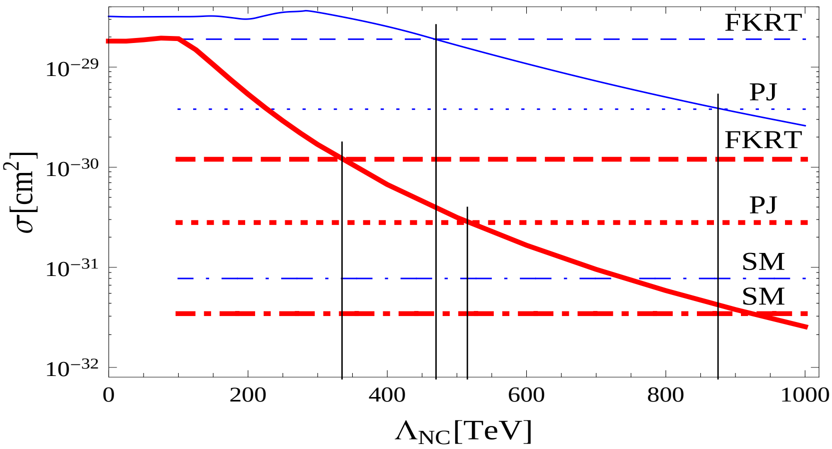


Figure 28: $\nu N \to \nu + anything$ cross sections vs. $\Lambda_{\rm NC}$ for $E_\nu = 10^{10}$ GeV (thick lines) and $E_\nu = 10^{11}$ GeV (thin lines). FKRT and PJ lines are the upper bounds on the ν-nucleon inelastic cross section, denoting different estimates for the cosmogenic ν-flux. SM denotes the SM total (charged current plus neutral current) ν-nucleon inelastic cross section. The vertical lines denote the intersections of our curves with the RICE results.

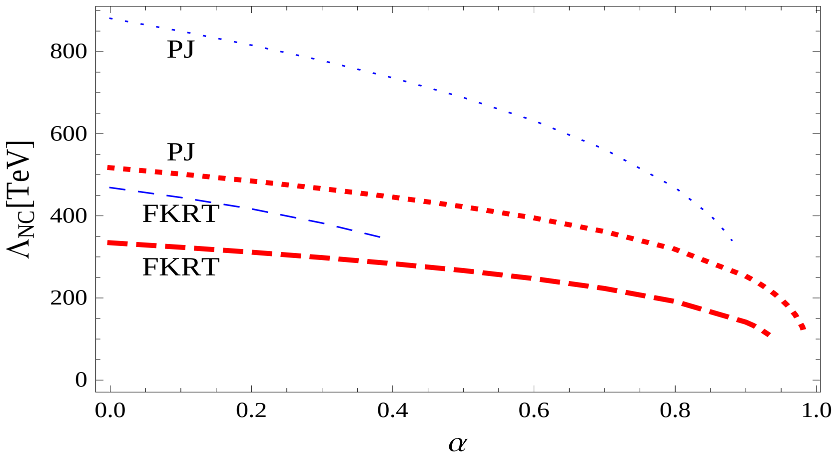


Figure 29: The intersections of our curves with the RICE results (cf. Fig.2) as a function of the fraction of Fe nuclei in the UHE cosmic rays. The terminal point on each curve represents the highest fraction of Fe nuclei above which no useful information on $\Lambda_{\rm NC}$ can be inferred with our method.

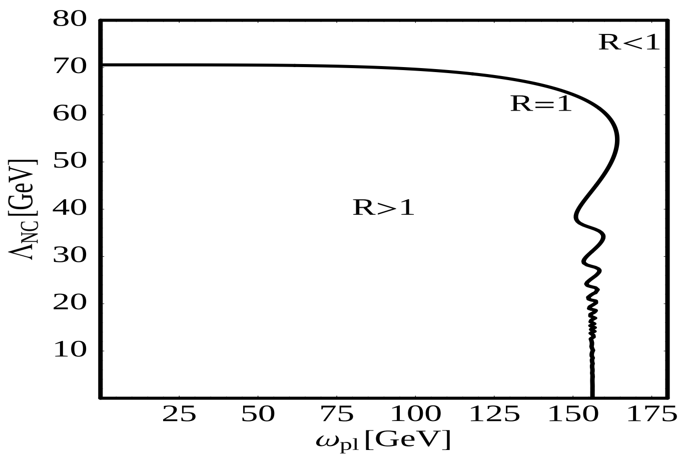


Figure 30: The plot of scale $\Lambda_{\rm NC}$ versus the plasmon frequency $\omega_{\rm pl}$ with $R = 1$, from the plasmon decay into neutrino pairs $\gamma_{\rm pl} \to \bar{\nu}\nu$.

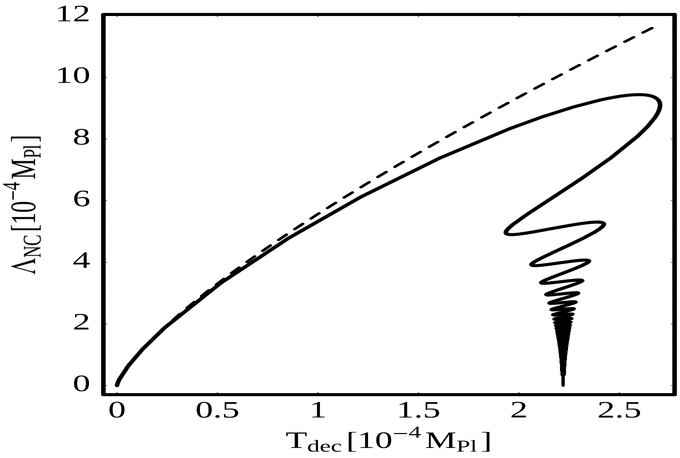


Figure 31: The plot of the scale $\Lambda_{\rm NC}$ versus T_{dec} for perturbative/exact solution (dashed/full curve).

the anti-commutator as well as the commutator of the Lie algebra generators. The algebra still closes for Hermitian matrices, but it is for instance not possible to impose the trace to be zero. This observation can be interpreted in two ways:
(a) The choice of gauge group is restricted to U(N) in the fundamental, anti-fundamental or adjoint representation; or
(b) the gauge fields are valued in the enveloping algebra of a Lie algebra and then any (unitary) representation is possible.
The case (a) applies also to the U(1) case and imposes severe restrictions on the allowed charges; it has been studied carefully and has led to "theorems" [64, 65]. The second case avoids the restrictions on the gauge group and choice of representation, but needs to address the potential problem of too many degrees of freedom, since all coefficient functions of the monomials in the generators could a priori be physical fields. The solution to this problem is that the coefficient fields are not all independent. They are rather functions of the correct number of ordinary gauge fields via Seiberg-Witten maps and their generalizations. The situation is reminiscent of the construction of superfields and supersymmetric actions in terms of ordinary fields in supersymmetry. This method, referred as Seiberg-Witten map or enveloping algebra approach avoids both the gauge group

and the U(1) charge issues. It was shown mathematically rigorously that any U(1) gauge theory on an arbitrary Poisson manifold can be deformation-quantized to a noncommutative gauge theory via the the enveloping algebra approach [66] and later extended to the non-Abelian gauge groups [67, 68]. The important step that has been missed in a paper [11] opposing above conclusions, is the use of reducible representations [12].

Following [12] we introduce a consistent noncommutative, Seiberg-Witten map and enveloping algebra based theory: Let $\hat{\Phi}[\Phi, A_\mu]$, $\hat{A}_\mu[A_\mu]$, $\hat{\Lambda}[\Lambda, A_\mu]$ be the SW map expanded fields (consider for example the well-known non-abelian maps for the Moyal-Weyl case [1]). Under an ordinary gauge transformation δ of the underlying fields $\phi_i(x)$, $i = 1, 2, 3$ and a_μ the SW expanded fields transform like it is expected for noncommutative fields.

Since in the noncommutative case the order of fields matters, there are in fact more choices than the one given in (4). In general all fields carry left and right charges that combines into the total commutative charge. Gauge invariance requires that the respective charges of neighboring fields must match with opposite signs. In the notation of (2) and (4), we have:

$$\delta\hat{\Phi} = i\hat{\Lambda}^L \star \hat{\Phi} - i\hat{\Phi} \star \hat{\Lambda}^R . \tag{4}$$

Using the associativity of the star product one can easily verify the formal consistency relation

$$[\delta_{\hat{\Lambda}}, \delta_{\hat{\Sigma}}]\hat{\Phi} = [i\hat{\Lambda}^L \overset{\star}{,} i\hat{\Sigma}^L] \star \hat{\Phi} - \hat{\Phi} \star [i\hat{\Lambda}^R \overset{\star}{,} i\hat{\Sigma}^R]. \tag{5}$$

Therefore the noncommutative gauge transformations $\hat{\Lambda}^{L/R}$ can be constructed from the classical fields and parameters $A_\mu^{L/R} = a_\mu(x)Q^{L/R}$ and $\Lambda^{L/R} = \lambda(x)Q^{L/R}$ with $Q^{L/R} = \mathrm{diag}(q_1^{L/R}, q_2^{L/R}, q_3^{L/R})$ and $q_i = q_i^L - q_i^R$ by so-called hybrid Seiberg-Witten maps [10, 69]. The hybrid covariant derivative is given by $\hat{D}_\mu\hat{\Phi} = \partial_\mu\hat{\Phi} - i\hat{A}_\mu^L \star \hat{\Phi} + i\hat{\Phi} \star \hat{A}_\mu^R$. Thanks to (5) the left and right NC gauge fields $\hat{A}_\mu^{L/R}$ are constructed from $A_\mu^{L/R}$ only, respectively. The gauge field action could be written as

$$\mathcal{L}_{gauge} = -\frac{1}{4g^2}\mathrm{tr}\left(\hat{F}_{\mu\nu}^L \star \hat{F}^{\mu\nu L} + \hat{F}_{\mu\nu}^R \star \hat{F}^{\mu\nu R}\right) , \tag{6}$$

with $g := e\sqrt{\mathrm{tr}(Q^L)^2 + \mathrm{tr}(Q^R)^2}$. In [12] we have employed this constructon on deformed Yukawa couplings. Namely, in the Yukawa terms, a star product deformation would prevent the charge summation. The hybrid SW map [10, 69] is introduced to recover gauge invariance. Thus the classical charge q is split into left and right charges $q = q^L - q^R$, as we have seen above.

4 Covariant θ-exact $\mathrm{U}_\star(1)$ model

We start with the following SW type of NC $\mathrm{U}_\star(1)$ gauge model:

$$S = \int -\frac{1}{4} F^{\mu\nu} \star F_{\mu\nu} + i\bar{\Psi} \star \not{D}\Psi \,, \tag{7}$$

with the NC definitions of the nonabelian field strength and the covariant derivative, respectively:

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu \stackrel{\star}{,} A_\nu], \\ D_\mu \Psi &= \partial_\mu \Psi - i[A_\mu \stackrel{\star}{,} \Psi]. \end{aligned} \tag{8}$$

All noncommutative fields in this action (A_μ, Ψ) are images under (hybrid) Seiberg-Witten maps of the corresponding commutative fields (a_μ, ψ). Here we shall interpret the NC fields as valued in the enveloping algebra of the underlying gauge group. This naturally corresponds to an expansion in powers of the gauge field a_μ and hence in powers of the coupling constant e. At each order in a_μ we shall determine θ-exact expressions.

In the next step we expand the action in terms of the commutative gauge parameter λ and fields a_μ and ψ using the SW map solution [48] up to the $\mathcal{O}(a^3)$ order:

$$\begin{aligned} \Lambda &= \lambda - \frac{1}{2}\theta^{ij} a_i \star_2 \partial_j \lambda \,, \\ A_\mu &= a_\mu - \frac{1}{2}\theta^{\nu\rho} a_\nu \star_2 (\partial_\rho a_\mu + f_{\rho\mu}), \\ \Psi &= \psi - \theta^{\mu\nu} a_\mu \star_2 \partial_\nu \psi \\ &+ \frac{1}{2}\theta^{\mu\nu}\theta^{\rho\sigma}\Big[(a_\rho \star_2 (\partial_\sigma a_\mu + f_{\sigma\mu})) \star_2 \partial_\nu \psi \\ &+ 2 a_\mu \star_2 (\partial_\nu (a_\rho \star_2 \partial_\sigma \psi)) - a_\mu \star_2 (\partial_\rho a_\nu \star_2 \partial_\sigma \psi) \\ &- \big(a_\rho \partial_\mu \psi (\partial_\nu a_\sigma + f_{\nu\sigma}) - \partial_\rho \partial_\mu \psi a_\nu a_\sigma\big)_{\star_3} \Big], \end{aligned} \tag{9}$$

with Λ being the NC gauge parameter and $f_{\mu\nu}$ is the abelian commutative field strength $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$.

The generalized Mojal-Weyl star products $\star_2$ and $\star_3$, appearing in (9), are defined, respectively, as

$$\begin{aligned} f(x) \star_2 g(x) &= [f(x) \stackrel{\star}{,} g(x)] \\ &= \frac{\sin\frac{\partial_1\theta\partial_2}{2}}{\frac{\partial_1\theta\partial_2}{2}} f(x_1)g(x_2)\bigg|_{x_1=x_2=x}, \end{aligned} \tag{10}$$

$$\begin{aligned} (f(x)g(x)h(x))_{\star_3} &= \Bigg(\frac{\sin(\frac{\partial_2\theta\partial_3}{2})\sin(\frac{\partial_1\theta(\partial_2+\partial_3)}{2})}{\frac{(\partial_1+\partial_2)\theta\partial_3}{2}\frac{\partial_1\theta(\partial_2+\partial_3)}{2}} \\ &+ \{1\leftrightarrow 2\}\Bigg) f(x_1)g(x_2)h(x_3)\bigg|_{x_i=x}, \end{aligned} \tag{11}$$

where $\star$ is associative but noncommutative, while $\star_2$ and $\star_3$ are both commutative but nonassociative.

The resulting expansion defines θ-exact neutrino-photon $\mathrm{U}_\star(1)$ actions, for a gauge and a matter sectors respectively. Pure gauge field (3-photon) action reads:

$$\begin{aligned} S_g &= \int i\partial_\mu a_\nu \star [a^\mu \stackrel{\star}{,} a^\nu] \\ &+ \frac{1}{2}\partial_\mu\Big(\theta^{\rho\sigma} a_\rho \star_2 (\partial_\sigma a_\nu + f_{\sigma\nu})\Big) \star f^{\mu\nu}. \end{aligned} \tag{12}$$

The photon-fermion action up to 2-photon 2-neutrino fields can be derived by using the first order gauge field and the second order neutrino field expansions,

$$
\begin{aligned}
S_f &= \int \Big(\bar\psi + (\theta^{ij}\partial_i\bar\psi \star_2 a_j)\Big)\gamma^\mu[a_\mu \overset{\star}{,} \psi] \\
&+ i(\theta^{ij}\partial_i\bar\psi \star_2 a_j)\partial\!\!\!/\psi - i\bar\psi \star \partial\!\!\!/(\theta^{ij}a_i \star_2 \partial_j\psi) \\
&- \bar\psi\gamma^\mu[a_\mu \overset{\star}{,} \theta^{ij}a_i\star_2\partial_j\psi] \\
&- \bar\psi\gamma^\mu\left[\frac{1}{2}\theta^{ij}a_i\star_2(\partial_j a_\mu + f_{j\mu})\overset{\star}{,}\psi\right] \\
&- i(\theta^{ij}\partial_i\bar\psi\star_2 a_j)\partial\!\!\!/(\theta^{kl}a_k\star_2\partial_l\psi) \\
&+ \frac{i}{2}\theta^{ij}\theta^{kl}\Big[(a_k \star_2 (\partial_l a_i + f_{li})) \star_2 \partial_j\bar\psi \\
&+ 2a_i \star_2 (\partial_j(a_k \star_2 \partial_l\bar\psi)) - a_i \star_2 (\partial_k a_j \star_2 \partial_l\bar\psi) \\
&+ \big(a_i\partial_k\bar\psi(\partial_j a_l + f_{jl}) - \partial_k\partial_i\bar\psi a_j a_l\big)_{\star_3}\Big]\partial\!\!\!/\psi \\
&+ \frac{i}{2}\theta^{ij}\theta^{kl}\bar\psi\partial\!\!\!/\Big[(a_k \star_2 (\partial_l a_i + f_{li})) \star_2 \partial_j\psi \\
&+ 2a_i\star_2(\partial_j(a_k\star_2\partial_l\psi)) - a_i\star_2(\partial_k a_j\star_2\partial_l\psi) \\
&+ \big(a_i\partial_k\psi(\partial_j a_l + f_{jl}) - \partial_k\partial_i\psi a_j a_l\big)_{\star_3}\Big]. \qquad (13)
\end{aligned}
$$

Note that actions for gauge and matter fields obtained above, (12) and (13) respectively, are nonlocal objects due to the presence of the star products: $\star$, $\star_2$ and $\star_3$. Feynman rules from above actions, represented in Fig.32, are given explicitly in [50].

5 Quantum properties: neutrino two-point function

As depicted in Fig. 33, there are four Feynman diagrams contributing to the ν-self-energy at one-loop. With the aid of (13), we have verified by explicit calculation that the 4-field tadpole (Σ_2) does vanish. The 3-fields tadpoles (Σ_3 and Σ_4) can be ruled out by invoking the NC charge conjugation symmetry [16]. Thus only the Σ_1 diagram needs to be evaluated. In spacetime of the dimensionality D we

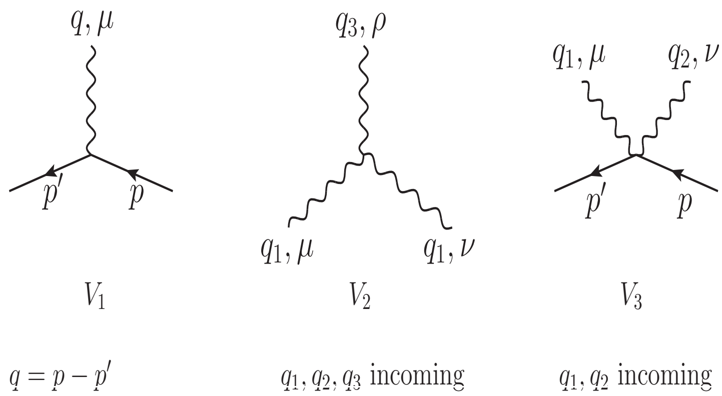


Figure 32: Three- and your-field vertices

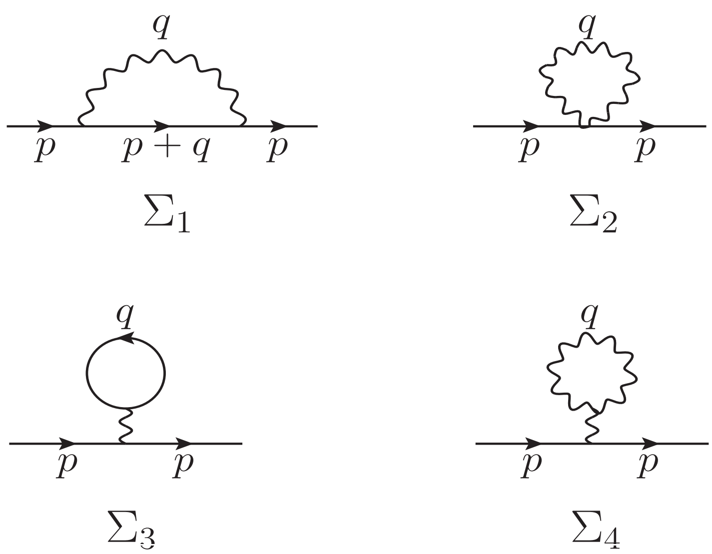


Figure 33: One-loop self-energy of a massless neutrino

obtain

$$
\begin{aligned}
\Sigma_1 &= \mu^{4-D}\int \frac{d^Dq}{(2\pi)^D}\left(\frac{\sin\frac{q\theta p}{2}}{\frac{q\theta p}{2}}\right)^2 \frac{1}{q^2}\frac{1}{(p+q)^2} \\
&\cdot \Bigg[(q\theta p)^2(4-D)(\not p+\not q) \\
&+ (q\theta p)\bigg(\not{\tilde q}(2p^2+2p\cdot q)-\not{\tilde p}(2q^2+2p\cdot q)\bigg) \\
&+ \bigg(\not p(\tilde q^2(p^2+2p\cdot q)-q^2(\tilde p^2+2\tilde p\cdot\tilde q)) \\
&+ \not q(\tilde p^2(q^2+2p\cdot q)-p^2(\tilde q^2+2\tilde p\cdot\tilde q))\bigg)\Bigg]\,,
\end{aligned}
\tag{14}
$$

where $\tilde p^\mu = (\theta p)^\mu = \theta^{\mu\nu}p_\nu$, and in addition $\tilde{\tilde p}^\mu = (\theta\theta p)^\mu = \theta^{\mu\nu}\theta_{\nu\rho}p^\rho$. To perform computations of those integrals using the dimensional regularization method, we first use the Feynman parametrization on the quadratic denominators, then the Heavy Quark Effective theory (HQET) parametrization [70] is used to combine the quadratic and linear denominators. In the next stage we use the Schwinger parametrization to turn the denominators into Gaussian integrals. Evaluating the relevant integrals for $D = 4-\epsilon$ in the limit $\epsilon \to 0$, we obtain the closed form expression for the self-energy

$$
\Sigma_1 = \gamma_\mu\left[p^\mu\, A + (\theta\theta p)^\mu\, \frac{p^2}{(\theta p)^2}\, B\right], \tag{15}
$$

$$
\begin{aligned}
A &= \frac{-1}{(4\pi)^2}\Bigg[p^2\left(\frac{\mathrm{tr}\theta\theta}{(\theta p)^2}+2\frac{(\theta\theta p)^2}{(\theta p)^4}\right)A_1 \\
&+ \left(1+p^2\left(\frac{\mathrm{tr}\theta\theta}{(\theta p)^2}+\frac{(\theta\theta p)^2}{(\theta p)^4}\right)\right)A_2\Bigg],
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
A_1 &= \frac{2}{\epsilon} + \ln(\mu^2(\theta p)^2) + \ln(\pi e^{\gamma_{\mathrm{E}}}) \\
&+ \sum_{k=1}^{\infty} \frac{\left(p^2(\theta p)^2/4\right)^k}{\Gamma(2k+2)} \left(\ln \frac{p^2(\theta p)^2}{4} + 2\psi_0(2k+2) \right) , \qquad (17) \\
A_2 &= -\frac{(4\pi)^2}{2} B = -2 \\
&+ \sum_{k=0}^{\infty} \frac{\left(p^2(\theta p)^2/4\right)^{k+1}}{(2k+1)(2k+3)\Gamma(2k+2)} \Bigg(\ln \frac{p^2(\theta p)^2}{4} \\
&- 2\psi_0(2k+2) - \frac{8(k+1)}{(2k+1)(2k+3)} \Bigg), \qquad (18)
\end{aligned}
$$

with $\gamma_{\mathrm{E}} \simeq 0.577216$ being Euler's constant.

The $1/\epsilon$ UV divergence could in principle be removed by a properly chosen counterterm. However due to the specific momentum-dependent coefficient in front of it, a nonlocal form for it is required.

5.1 UV/IR mixing

Turning to the UV/IR mixing problem, we recognize a soft UV/IR mixing term represented by a logarithm,

$$
\Sigma_{\mathrm{UV/IR}} = \not{p} \frac{p^2}{(4\pi)^2} \left(\ln \frac{1}{|\mu(\theta p)|^2} \right) \left(\frac{\mathrm{tr}\theta\theta}{(\theta p)^2} + 2\frac{(\theta\theta p)^2}{(\theta p)^4} \right). \qquad (19)
$$

Instead of dealing with nonlocal counterterms, we take a different route here to cope with various divergences besetting (15). Since $\theta^{0i} \neq 0$ makes a NC theory nonunitary [58], we can, without loss of generality, chose θ to lie in the (1, 2) plane

$$
\theta^{\mu\nu} = \frac{1}{\Lambda_{\mathrm{NC}}^2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \qquad (20)
$$

Automatically, this produces

$$
\frac{\mathrm{tr}\theta\theta}{(\theta p)^2} + 2\frac{(\theta\theta p)^2}{(\theta p)^4} = 0, \ \forall p. \qquad (21)
$$

With (21), Σ_1, in terms of Euclidean momenta, receives the following form:

$$\Sigma_1 = \frac{-1}{(4\pi)^2}\gamma_\mu \left[p^\mu\left(1+\frac{\mathrm{tr}\theta\theta}{2}\frac{p^2}{(\theta p)^2}\right) - 2(\theta\theta p)^\mu \frac{p^2}{(\theta p)^2}\right] A_2. \tag{22}$$

By inspecting (18) one can be easily convinced that A_2 is free from the $1/\epsilon$ divergence and the UV/IR mixing term, being also well-behaved in the infrared, in the $\theta \to 0$ as well as $\theta p \to 0$ limit. We see, however, that the two terms in (22), one being proportional to $\not{p}$ and the other proportional to $\tilde{\not{p}}$, are still ill-behaved in the $\theta p \to 0$ limit. If, for the choice (20), P denotes the momentum in the (1, 2) plane, then $\theta p = \theta P$. For instance, a particle moving inside the NC plane with momentum P along the one axis, has a spatial extension of size $|\theta P|$ along the other. For the choice (20), $\theta p \to 0$ corresponds to a zero momentum projection onto the (1, 2) plane. Thus, albeit in our approach the commutative limit ($\theta \to 0$) is smooth at the quantum level, the limit when an extended object (arising due to the fuzziness of space) shrinks to zero, is not. We could surely claim that in our approach the UV/IR mixing problem is considerably softened; on the other hand, we have witnessed how the problem strikes back in an unexpected way. This is, at the same time, the first example where this two limits are not degenerate.

5.2 Neutrino dispersion relations

In order to probe physical consequence of the 1-loop quantum correction, with Σ_{1-loop_M} from Eq. (3.25) in [50], we consider the modified propagator

$$\frac{1}{\not{\Sigma}} = \frac{1}{\not{p} - \Sigma_{1-loop_M}} = \frac{\not{\Sigma}}{\Sigma^2}. \tag{23}$$

We further choose the NC parameter to be (20) so that the denominator is finite and can be expressed explicitly:

$$\Sigma^2 = p^2\left[\hat{A}_2^2\left(\frac{p^4}{p_r^4} + 2\frac{p^2}{p_r^2} + 5\right) - \hat{A}_2\left(6 + 2\frac{p^2}{p_r^2}\right) + 1\right], \tag{23}$$

where p_r represents r-component of the momentum p in a cylindrical spatial coordinate system and $\hat{A}_2 = e^2 A_2/(4\pi)^2 = -B/2$.

From above one see that $p^2 = 0$ defines one set of the dispersion relation, corresponding to the dispersion for the massless neutrino mode, however the denominator Σ^2 has one more coefficient Σ' which could also induce certain zero-points. Since the $\hat{A}_2$ is a function of a single variable $p^2 p_r^2$, with $p^2 = p_0^2 - p_1^2 - p_2^2 - p_3^2$ and $\mathrm{p}_\mathrm{r}^2 = \mathrm{p}_1^2 + \mathrm{p}_2^2$, the condition $\Sigma' = 0$ can be expressed as a simple algebraic equation

$$\hat{A}_2^2 z^2 - 2\left(A_2 - \hat{A}_2^2\right) z + \left(1 - 6\hat{A}_2 + 5\hat{A}_2^2\right) = 0\,, \tag{24}$$

of new variables $z := p^2/p_r^2$, in which the coefficients are all functions of $y := p^2 p_r^2/\Lambda_{\mathrm{NC}}^4$.

The two formal solutions of the equation (24)

$$z = \frac{1}{\hat{A}_2}\left[\left(1 - \hat{A}_2\right) \pm 2\left(\hat{A}_2 - \hat{A}_2^2\right)^{\frac{1}{2}}\right]\,, \tag{25}$$

are birefringent. The behavior of solutions (25), is next analyzed at two limits $y \to 0$, and $y \to \infty$.

5.2.1 The low-energy regime: $p^2 p_r^2 \ll \Lambda_{\mathrm{NC}}^4$

For $y \ll 1$ we set $\hat{A}_2$ to its zeroth order value $e^2/8\pi^2$,

$$\begin{aligned} p^2 &\sim \left(\left(\frac{8\pi^2}{e^2} - 1\right) \pm 2\left(\frac{8\pi^2}{e^2} - 1\right)^{\frac{1}{2}}\right) \cdot p_r^2 \\ &\simeq (859 \pm 59) \cdot p_r^2\,, \end{aligned} \tag{26}$$

obtaining two (approximate) zero points. From the definition of p^2 and p_r^2 we see that both solutions are real and positive. Taking into account the higher order (in y) correction these poles will locate nearby the real axis of the complex p_0 plane thus correspond to some metastable modes with the above defined dispersion relations. As we can see, the modified dispersion relation (26) does not depend on the noncommutative scale, therefore it introduces a discontinuity in the $\Lambda_{\mathrm{NC}} \to \infty$ limit, which is not unfamiliar in noncommutative theories.

5.2.2 The high-energy regime: $p^2 p_r^2 \gg \Lambda_{\mathrm{NC}}^4$

At $y \gg 1$ we analyze the asymptotic behavior of

$$A_2 \sim \frac{i\pi^2}{8}\sqrt{y}\left(1 - \frac{16i}{\pi y} e^{-\frac{i}{2}\sqrt{y}}\right) + \mathcal{O}\left(y^{-1}\right)\,, \tag{27}$$

from [50], therefore (25) can be reduced to

$$z \sim -1 \pm 2i \quad \rightarrow \quad p_0^2 \sim p_3^2 \pm 2ip_r^2. \tag{28}$$

We thus reach two unstable deformed modes besides the usual mode $p^2 = 0$ in the high energy regime. Here again the leading order deformed dispersion relation does not depend on the noncommutative scale $\Lambda_{\rm NC}$.

5.3 The alternative action self-energy

Using the Feynman rule of the alternative action (2.15) from Ref [50], which is a consequence of the SW freedom, we find the following contribution to the neutrino self-energy from diagram Σ_1

$$\Sigma_{1_{alt}} = \not p\,\frac{8}{3}\frac{1}{(4\pi)^2}\frac{1}{|\theta p|^2}\left(\frac{{\rm tr}\theta\theta}{(\theta p)^2} + 4\frac{(\theta\theta p)^2}{(\theta p)^4}\right). \tag{29}$$

The detailed computation is presented in Appendix B of Ref. [50]. We notice that there are no hard $1/\epsilon$ UV divergent and no logarithmic UV/IR mixing terms, and the finite terms like in A_1 and A_2 are also absent. Thus the subgraph Σ_1 for the alternative action (2.15) in [50] does not require any counter-term. However, the result (29), does express powerful UV/IR mixing effect, that is in terms of scales terms, the $\Sigma_{1_{alt}}$ experience the forth-power of the *NC-scale/momentum-scale* ratios $\sim |p|^{-2}|\theta p|^{-2}$ in (29), i.e. we are dealing with the $\Sigma_{1_{alt}} \sim \not p\,(\Lambda_{\rm NC}/p)^4$ within the ultraviolet and infrared limits for $\Lambda_{\rm NC}$ and p, respectively.

6 Phenomenology: $Z \to \nu\bar\nu$ decay rate

To illustrate another phenomenologicall effects of our θ-exact construction, we present a computation the $Z \to \nu\bar\nu$ decay rate in the Z–boson rest frame, which is then readily to be compared with the precision Z resonance measurements, where Z is almost at rest. Since the complete $Z\nu\nu$ interaction on noncommutative spaces was discussed in details in [12,49,50,53], we shall not repeat it here. We only give the *almost complete* $Z\nu\bar{\nu}$ vertex from [12]

$$
\begin{aligned}
\Gamma^\mu(p',p) &= i\frac{g}{2\cos\theta_W}\Bigg(\gamma^\mu + \frac{i}{2}F_\bullet(p',p) \\
&\cdot\Big[(p'\theta p)\gamma^\mu + (\theta p')^\mu \not{p} - (\theta p)^\mu \not{p}'\Big]\Bigg)\frac{1-\gamma_5}{2} \\
&+ \frac{\kappa e}{2}\tan\theta_W F_{\star 2}(p',p)\Big[(p'\theta p)\gamma^\mu + (\theta p')^\mu \not{p} - (\theta p)^\mu \not{p}'\Big] ,
\end{aligned}
\tag{30}
$$

where κ is an arbitrary constant[27], and

$$
\begin{aligned}
(p'\theta p)F_\bullet(p',p) &= -2i\left(1-\exp\left(i\frac{M_Z\,p}{2\Lambda^2_{\rm NC}}\cos\vartheta\right)\right) , \\
(p'\theta p)F_{\star 2}(p',p) &= -2\sin\left(\frac{M_Z\,p}{2\Lambda^2_{\rm NC}}\cos\vartheta\right) .
\end{aligned}
\tag{31}
$$

Note here that due to the equations of motions, for massless on-shell neutrinos the terms $[(\theta p')^\mu \not{p} - (\theta p)^\mu \not{p}']\,(1-\gamma_5)$ in the vertex (30) do not contribute to the $Z \to \nu\bar{\nu}$ amplitude. Thus the vertex (30) has the same form as the SM vertex $\frac{ig}{2\cos\theta_W}\,\gamma^\mu\,(g_V - g_A\gamma_5)$ [71, 72] with

$$
\begin{aligned}
g_V &= 1-\frac{1}{2}\exp\left(\frac{iM_Z\,p\cos\vartheta}{2\Lambda^2_{\rm NC}}\right) \\
&+ 2i\kappa\sin^2\theta_W\sin\left(\frac{M_Z\,p\cos\vartheta}{2\Lambda^2_{\rm NC}}\right) ,
\end{aligned}
\tag{32}
$$

$$
g_A = 1-\frac{1}{2}\exp\left(\frac{iM_Z\,p\cos\vartheta}{2\Lambda^2_{\rm NC}}\right) .
\tag{33}
$$

The temporary component $\vec{E}_\theta$ of θ is reduced from equations above since for the Z–boson at rest we have

$$
p'\theta\,p = -M_Z\,\vec{p}\cdot\vec{E}_\theta = -\frac{M_Z\;p\cos\vartheta}{\Lambda^2_{\rm NC}} .
\tag{34}
$$

[27]The constant κ measures a correction from the $\star$-commutator coupling of the right handed neutrino ν_R to the noncommutative hypercharge $\mathrm{U}_\star(1)_\mathrm{Y}$ gauge field $B^0_\mu[\kappa]$. Coupling is chiral blind and it vanishes in the commutative limit. The non-κ-proportional term, on the other hand, is the noncommutative deformation of standard model Z-neutrino coupling, which involves the left handed neutrinos only. Details can be found in section four of [12].

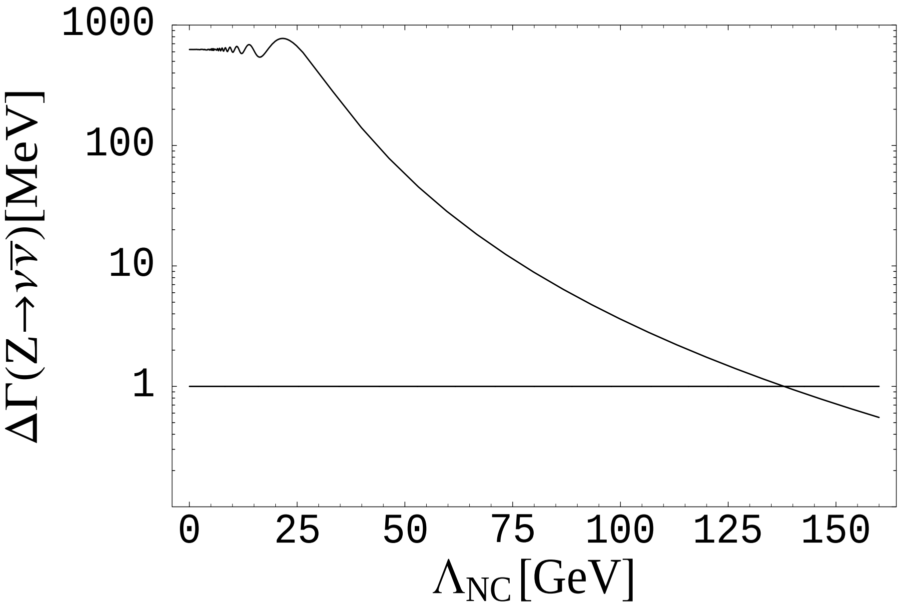


Figure 34: $\Delta\Gamma(Z \to \nu\bar{\nu})$ decay width vs. $\Lambda_{\rm NC}$.

with $|\vec{E}_\theta| = 1/\Lambda^2_{\rm NC}$ and ϑ the angle between $\vec{p}$ and $\vec{E}_\theta$ respectively.

Using $Z\nu\bar{\nu}$ vertex (30), we obtain the following $Z \to \nu\bar{\nu}$ partial width [73]

$$\begin{aligned}
\Gamma(Z \to \nu\bar{\nu}) &= \Gamma_{\rm SM}(Z \to \nu\bar{\nu}) \\
&+ \frac{\alpha}{3M_Z|\vec{E}_\theta|}\Bigg[\kappa\,(1-\kappa+\kappa\cos 2\theta_W)\sec^2\theta_W \cos\left(\frac{M_Z^2|\vec{E}_\theta|}{4}\right) \\
&- 8\csc^2 2\theta_W\Bigg]\cdot\sin\left(\frac{M_Z^2|\vec{E}_\theta|}{4}\right) \\
&+ \frac{\alpha M_Z}{12}\Bigg[-2\kappa^2 + (\kappa(2\kappa-1)+2)\sec^2\theta_W + 2\csc^2\theta_W\Bigg],
\end{aligned} \tag{35}$$

whose NC part vanishes when $\vec{E}_\theta \to 0$, i.e. for vanishing θ or space-like noncommutativity, but not light-like [58, 59].

A comparison of the experimental Z decay width $\Gamma_{\rm invisible} = (499.0 \pm 1.5)$ MeV [74] with its SM theoretical counterpart, allows us to set a constraint $\Gamma(Z \to \nu\bar{\nu})-$

$\Gamma_{\rm SM}(Z \to \nu\bar{\nu}) \lesssim 1$ MeV, from where a bound on the scale of noncommutativity $\Lambda_{\rm NC} = |\vec{E}_\theta|^{-1/2} \gtrsim 140$ GeV is obtained (see Fig. 34), for the choice $\kappa = 1$.

7 Discussion and conclusions

We have presented the tree level cosmogenic neutrinos (ν's) scatterings: $\nu N \to \nu$ + *anything* and particle decays: $((\gamma_{\rm pl}, Z) \to \nu\bar{\nu})$ in the covariant θ-exact noncommutative quantum gauge theory based on Seiberg-Witten maps and enveloping algebra formalism.

In the energy range of interest, 10^{10} to 10^{11} GeV, where there is always energy of the system (E) larger than the NC scale ($E/\Lambda_{\rm NC} > 1$), the perturbative expansion in terms of $\Lambda_{\rm NC}$ retains no longer its meaningful character, thus it is forcing us to resort to those NC field-theoretical frameworks involving the full θ-resummation. Our numerical estimates of the contribution to the processes coming from the photon exchange, pins impeccably down a lower bound on $\Lambda_{\rm NC}$ to be as high as around up to $\mathcal{O}(10^6)$ GeV, depending on the cosmogenic ν-flux.

For above analysis it was necessary to use results of [12] which shows explicitly that the "no-go theorem" [11] is certainly not applicable to our SW-map based θ-exact models of the NCGFT. Namely, it is known to be impossible in noncommutative geometry to directly form tensor products from the NC fields as long as there is no additional underlaying mathematical structure. The SW-map based models do however have an additional underlying mathematical structure: They can be understood as the deformation quantization of ordinary fiber bundles over a Poisson manifold. With this additional structure, tensor products are possible and survive the quantization procedure [66]. However, the authors in [11] failed to directly form tensor products of noncommutative fields. The proof of this failure is given in [12].

Now we first discuss θ-exact computation of the one-loop quantum correction to the ν-propagator. We in particular evaluate the neutrino two-point function, and demonstrate how quantum effects in the θ-exact SW map approach to NCGFT's, together with a combination of Schwinger, Feynman, and HQET parameterization, reveal a much richer structure yielding the one-loop quantum correction in a closed form.

General expression for the neutrino self-energy (15) contains in (17) both a hard $1/\epsilon$ UV term and the UV/IR mixing term with a logarithmic infrared singularity $\ln|\theta p|$. Results shows complete decoupling of the UV divergent term from softened UV/IR mixing term and from the finite terms as well. Our deformed dispersion relations at both the low and high energies and at the leading order do not depend on the noncommutative scale $\Lambda_{\rm NC}$. The low energy dispersion relation (26) is, in principle, capable of generating a direction dependent superluminal velocity, this can be seen clearly from the maximal attainable velocity of the neutrinos

$$\frac{\mathrm{v}_{max}}{c} = \frac{dE}{d|\vec{p}|} \sim \sqrt{1 + (859 \pm 59)\sin^2\vartheta}\,, \tag{36}$$

where ϑ is the angle with respect to the direction perpendicular to the NC plane. This gives one more example how such spontaneous θ-background breaking of Lorentz symmetry could affect the particle kinematics through quantum corrections, even without divergent behavior like UV/IR mixing. On the other hand one can also see that the magnitude of superluminosity is in general very large in our model as a quantum effect, thus seems contradicting various observations which suggests much smaller values [75–77]. On the other hand, note that the large superluminal velocity issue may also be reduced/removed by taking into account several considerations and/or properties:

(1.) Selection of a constant nonzero θ background in this paper is due to the computational simplicity. The results will, however, still hold for a NC background that is varying sufficiently slowly with respect to the scale of noncommutativity. There is no physics reason to expect θ to be a globally constant background *ether*. In fact, if the θ background is only nonzero in tiny regions (NC bubbles) the effects of the modified dispersion relation will be suppressed macroscopically. Certainly a better understanding of possible sources of NC is needed.

(2.) We have considered only the purely noncommutative neutrino-photon coupling. However, it has been pointed out that modified neutrino dispersion relation could open decay channels within the commutative standard model framework [78]. In our case this would further provide decay channel(s) which can bring superluminal neutrinos to normal ones.

(3.) Note that the model 1 is not the only allowed deformed model with noncommutative neutrino-photon coupling. And as we have shown for our model 2, there could be no modified dispersion relation(s) for deformation(s) other than

1, therefore it is reasonable to conjecture that Seiberg-Witten map freedom may also serve as one possible remedy to this issue.
(4.) Our results differs with respect to [64] since in our case both terms are proportional to the spacetime noncommutativity dependent θ-ratio (the scale-independent structure!) factor in (21), which arise from the natural non-locality of our actions. Besides the divergent terms, a new spinor structure $(\theta\theta p)$ with finite coefficients emerges in our computation, see (15)-(18). All these structures are proportional to p^2, therefore if appropriate renormalization conditions are imposed, the commutative dispersion relation $p^2 = 0$ can still hold, as a part of the full set of solutions obtained in (23).
(5.) Finally, we mention that our approach to UV/IR mixing should not be confused with the one based on a theory with UV completion ($\Lambda_{\rm UV} < \infty$), where a theory becomes an effective QFT, and the UV/IR mixing manifests itself via a specific relationship between the UV and the IR cutoffs [79, 80].

From the same actions (12, 13), but for three different cosmological laboratories, that is from UHE cosmic ray neutrino scatterings on nuclei [52], from the BBN and from the RPAI [47], we obtain very similar, a quite strong bounds on the NC scale, of the order of 10^6 GeV. Note in particular that all results depicted in Figs.28-31, and 34 show closed-convergent forms.

All above summarized properties are previously unknown features of θ-exact NC gauge field theory. They appear in the model with the action presented in section 4. The alternative action, and the corresponding ν-self-energy (29), has less striking features, but it does have it's own advantages due to the absence of a hard UV divergences, and the absence of complicated finite terms. The structure in (29) is different (it is *NC-scale/energy* dependent) with respect to the NC scale-independent structure from (21), as well as to the structure arising from fermion self-energy computation in the case of $\star$-product only unexpanded theories [64,81]. However, (29) does posses powerful UV/IR mixing effect. This is fortunate with regard to the use of low-energy NCQFT as an important window to holography [57] and quantum gravity [82].

Acknowledgement

I would like to thank Jiangyang You for many valuable comments/remarks.

References

1. N. Seiberg and E. Witten, String theory and noncommutative geometry, JHEP **9909** (1999) 032.
2. I. Hinchliffe, N. Kersting, and Y. L. Ma, Review of the phenomenology of noncommutative geometry, Int. J. Mod. Phys. **A19** (2004), 179–204.
3. J. Trampetic, Renormalizability and Phenomenology of theta-expanded Noncommutative Gauge Field Theory, Fortsch. Phys. **56** (2008) 521.
4. R. Jackiw, S. Y. Pi, Covariant coordinate transformations on noncommutative space, Phys. Rev. Lett. **88** (2002) 111603.
5. J. Madore, S. Schraml, P. Schupp, and J. Wess, Gauge theory on noncommutative spaces, Eur. Phys. J. **C16** (2000) 161.
6. B. Jurco, S. Schraml, P. Schupp, and J. Wess, Enveloping algebra valued gauge transformations for non-Abelian gauge groups on non-commutative spaces, Eur. Phys. J. **C17** (2000) 521.
7. B. Jurco and P. Schupp, Noncommutative Yang-Mills from equivalence of star products, Eur. Phys. J. **C14** (2000) 367.
8. B. Jurco, P. Schupp, and J. Wess, Nonabelian noncommutative gauge theory via noncommutative extra dimensions, Nucl. Phys. **B604** (2001) 148.
9. B. Jurco, L. Moller, S. Schraml, P. Schupp, and J. Wess, Construction of non-Abelian gauge theories on noncommutative spaces, Eur. Phys. J. **C21** (2001) 383.
10. X. Calmet, B. Jurco, P. Schupp, J. Wess, and M. Wohlgenannt, The standard model on non-commutative space-time, Eur. Phys. J. **C23** (2002) 363.
11. M. Chaichian, P. Presnajder, M. M. Sheikh-Jabbari, A. Tureanu, Can Seiberg-Witten Map Bypass Noncommutative Gauge Theory No-Go Theorem?, Phys. Lett. **B683**, 55-61 (2010).
12. R. Horvat, A. Ilakovac, P. Schupp, J. Trampetić, and J. You, Yukawa couplings and seesaw neutrino masses in noncommutative gauge theory, Phys. Lett. **B715**, 340-347 (2012).
13. W. Behr, N. Deshpande, G. Duplančić, P. Schupp, J. Trampetić, and J. Wess, The $Z \to \gamma\gamma, gg$ decays in the noncommutative standard model, Eur. Phys. J. **C29** (2003) 441.
14. N. Deshpande and X. He, Triple neutral gauge boson couplings in noncommutative standard model, Phys. Lett. **B533** (2002) 116.
15. G. Duplancic, P. Schupp and J. Trampetic, Comment on triple gauge boson interactions in the non-commutative electroweak sector, Eur. Phys. J. **C32** (2003) 141.

16. P. Aschieri, B. Jurco, P. Schupp, and J. Wess, Non-commutative GUTs, standard model and C, P, T, Nucl. Phys. **B651** (2003) 45.

17. B. Melic, K. Passek-Kumericki, J. Trampetic, P. Schupp, and M. Wohlgenannt, The standard model on non-commutative space-time: Electroweak currents and Higgs sector, Eur. Phys. J. **C42** (2005) 483–499.

18. B. Melic, K. Passek-Kumericki, J. Trampetic, P. Schupp, and M. Wohlgenannt, The standard model on non-commutative space-time: Strong interactions included, Eur. Phys. J. **C42** (2005) 499–504.

19. T. Filk, Divergencies in a field theory on quantum space, Phys. Lett. **B376** (1996) 53.

20. C.P. Martin, D. Sanchez-Ruiz, The One-loop UV Divergent Structure of $U(1)$ Yang-Mills Theory on Noncommutative R^4, Phys. Rev. Lett. **83** (1999) 476–479.

21. S. Minwalla, M. Van Raamsdonk and N. Seiberg, Noncommutative perturbative dynamics, JHEP **0002**, 020 (2000).

22. A. Matusis, L. Susskind, and N. Toumbas, The IR/UV connection in the non-commutative gauge theories, JHEP **0012** (2000) 002.

23. H. Grosse and R. Wulkenhaar, Renormalisation of ϕ^4 theory on noncommutative $\mathbb{R}^4$ in the matrix base, *Commun. Math. Phys.* **256** (2005) 305–374.

24. S. Meljanac, A. Samsarov, J. Trampetic and M. Wohlgenannt, Scalar field propagation in the ϕ^4 κ-Minkowski model, JHEP **1112** (2011) 010.

25. J. Magnen, V. Rivasseau and A. Tanasa, Commutative limit of a renormalizable noncommutative model, Europhys. Lett. **86** (2009) 11001.

26. L. C. Q. Vilar, O. S. Ventura, D. G. Tedesco, and V. E. R. Lemes, Renormalizable Noncommutative U(1) Gauge Theory Without IR/UV Mixing, **PoS ISFTG (2009) 071**.

27. C. P. Martin, The gauge anomaly and the Seiberg-Witten map, Nucl. Phys. **B652** (2003) 72.

28. F. Brandt, C. P. Martin, and F. R. Ruiz, Anomaly freedom in Seiberg-Witten noncommutative gauge theories, JHEP **0307**, 068 (2003).

29. A. Bichl, J. Grimstrup, H. Grosse, L. Popp, M. Schweda, and R. Wulkenhaar, Renormalization of the noncommutative photon self-energy to all orders via Seiberg-Witten map, JHEP **0106** (2001) 013.

30. M. Buric, V. Radovanovic, and J. Trampetic, The one-loop renormalization of the gauge sector in the noncommutative standard model, JHEP **0703** (2007) 030.

31. D. Latas, V. Radovanovic, and J. Trampetic, Non-commutative SU(N) gauge theories and asymptotic freedom, Phys. Rev. **D76** (2007) 085006.

32. C. P. Martin, C. Tamarit, Renormalisability of the matter determinants in noncommutative gauge theory in the enveloping-algebra formalism, Phys. Lett. **B658**, 170 (2008).

33. M. Buric, D. Latas, V. Radovanovic and J. Trampetic, The absence of the 4ψ divergence in noncommutative chiral models, Phys. Rev. **D77** (2008) 045031.

34. C. Martin, C. Tamarit, Noncommutative GUT inspired theories and the UV finiteness of the fermionic four point functions, Phys. Rev. **D80**, 065023 (2009).

35. C. P. Martin, C. Tamarit, Renormalisability of noncommutative GUT inspired field theories with anomaly safe groups, JHEP **0912** (2009) 042.

36. C. Tamarit, Noncommutative GUT inspired theories with U(1), SU(N) groups and their renormalisability, Phys. Rev. **D81** (2010) 025006.

37. M. Buric, D. Latas, V. Radovanovic and J. Trampetic, Chiral fermions in noncommutative electrodynamics: renormalizability and dispersion, Phys. Rev. **D83** (2011) 045023.

38. P. Schupp,J. Trampetic, J. Wess, and G. Raffelt, The photon neutrino interaction in non-commutative gauge field theory and astrophysical bounds, Eur. Phys. J. **C36** (2004) 405.

39. P. Minkowski, P. Schupp, and J. Trampetic, Neutrino dipole moments and charge radii in non- commutative space-time, Eur. Phys. J. **C37** (2004) 123,

40. T. Ohl and J. Reuter, Testing the noncommutative standard model at a future photon collider, Phys. Rev. **D70** (2004) 076007

41. A. Alboteanu, T. Ohl, and R. Ruckl, Probing the noncommutative standard model at hadron colliders, Phys. Rev. **D74** (2006) 096004, ibid *Acta Phys. Polon.* **B38** (2007) 3647.

42. M. Buric, D. Latas, V. Radovanovic, and J. Trampetic, Nonzero $Z \to \gamma\gamma$ decays in the renormalizable gauge sector of the noncommutative standard model, Phys. Rev. **D75** (2007) 097701

43. T. Ohl, C. Speckner, The Noncommutative Standard Model and Polarization in Charged Gauge Boson Production at the LHC, Phys. Rev. **D82** (2010) 116011.

44. B. Melic, K. Passek-Kumericki, and J. Trampetic, Quarkonia decays into two photons induced by the space-time noncommutativity, Phys. Rev. **D72**, 054004 (2005).

45. B. Melic, K. Passek-Kumericki and J. Trampetic, $K \to \pi\gamma$ decay and space-time noncommutativity, Phys. Rev. **D72** (2005) 057502.

46. C. Tamarit and J. Trampetic, Noncommutative fermions and quarkonia decays, Phys. Rev. **D79**, 025020 (2009).

47. R. Horvat, J. Trampetic, Constraining spacetime noncommutativity with primordial nucleosynthesis, Phys. Rev.**D79** (2009) 087701.

48. P. Schupp and J. You, UV/IR mixing in noncommutative QED defined by Seiberg-Witten map, JHEP **0808** (2008) 107.

49. R. Horvat, A. Ilakovac, J. Trampetic and J. You, On UV/IR mixing in noncommutative gauge field theories, JHEP **1112** (2011) 050.

50. R. Horvat, A. Ilakovac, P. Schupp, J. Trampetic and J. You, Neutrino propagation in noncommutative spacetimes, JHEP **1204** (2012) 108.

51. J. Zeiner, PhD thesis.

52. R. Horvat, D. Kekez and J. Trampetic, Spacetime noncommutativity and ultra-high energy cosmic ray experiments, Phys. Rev. **D83** (2011) 065013.

53. R. Horvat, D. Kekez, P. Schupp, J. Trampetic, J. You, Photon-neutrino interaction in theta-exact covariant noncommutative field theory, Phys. Rev. **D84** (2011) 045004.

54. R. Horvat, J. Trampetic, A bound on the scale of spacetime noncommutativity from the reheating phase after inflation, Phys. Lett. **B710** (2012) 219.

55. Y. Okawa and H. Ooguri, An exact solution to Seiberg-Witten equation of noncommutative gauge theory, Phys. Rev. **D64** (2001) 046009.

56. C. P. Martin, arXiv:1206.2814 [hep-th].

57. R. Horvat and J. Trampetic, Constraining noncommutative field theories with holography, JHEP **1101** (2011) 112.

58. J. Gomis and T. Mehen, Space-time noncommutative field theories and unitarity, Nucl. Phys. **B591**, 265 (2000).

59. O. Aharony, J. Gomis and T. Mehen, On theories with lightlike noncommutativity, JHEP **0009**, 023 (2000).

60. J. Abraham *et al.* [Pierre Auger Collaboration], Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory, Astropart. Phys. **29**, 243 (2008).

61. Z. Fodor, S. D. Katz, A. Ringwald and H. Tu, Bounds on the cosmogenic neutrino flux, JCAP **0311**, 015 (2003).

62. R. J. Protheroe and P. A. Johnson, Propagation of ultrahigh-energy protons over cosmological distances and implications for topological defect models, Astropart. Phys. **4**, 253 (1996).

63. I. Kravchenko *et al.*, Limits on the ultra-high energy electron neutrino flux from the RICE experiment, Astropart. Phys. **20**, 195 (2003).

64. M. Hayakawa, Perturbative analysis on infrared aspects of noncommutative QED on R**4, Phys. Lett. **B478**, 394 (2000).

65. M. Chaichian, P. Presnajder, M. M. Sheikh-Jabbari, A. Tureanu, Noncommutative gauge field theories: A No go theorem, Phys. Lett. **B526**, 132-136 (2002).

66. B. Jurco, P. Schupp, J. Wess, Noncommutative line bundle and Morita equivalence, Lett. Math. Phys. **61**, 171-186 (2002).

67. M. Bordemann, N. Neumaier, S. Waldmann, and S. Weiss, arXiv:0711.2965.

68. H. Bursztyn, V. Dolgushev, and S. Waldmann, arXiv:0909.4259.

69. P. Schupp, Non-Abelian gauge theory on noncommutative spaces, PrHEP-hep2001/238 (2001).

70. A. G. Grozin, *Lectures on perturbative HQET. I*, hep-ph/0008300.

71. C. Giunti and C. W. Kim, *Fundamentals of Neutrino Physics and Astrophysics*, Oxford, UK:U.Pr.(2007)710 p.

72. V. A. Novikov, L. B. Okun, A. N. Rozanov, and M. I. Vysotsky, Theory of Z boson decays, Rept. Prog. Phys. **62** (1999) 1275–1332.

73. R. Horvat, A. Ilakovac, D. Kekez, J. Trampetic and J. You, Forbidden and Invisible Z Boson Decays in Covariant theta-exact Noncommutative Standard Model, arXiv:1204.6201 [hep-ph].

74. K. Nakamura et al. (Particle Data Group), Journal of Physics **G37**, 075021 (2010) and 2011 partial update for the 2012 edition.

75. K. Hirata *et al.* [KAMIOKANDE-II Collaboration], Observation of a Neutrino Burst from the Supernova SN 1987a, Phys. Rev. Lett. **58** (1987) 1490.

76. R. M. Bionta, G. Blewitt, C. B. Bratton, D. Casper, A. Ciocio, R. Claus, B. Cortez, M. Crouch, S.T. Dye, S. Errede *et al.*, Observation of a Neutrino Burst in Coincidence with Supernova SN 1987a in the Large Magellanic Cloud, Phys. Rev. Lett. **58** (1987) 1494.

77. M. J. Longo, Tests of Relativity from SN1987a, Phys. Rev. **D36** (1987) 3276.

78. A. G. Cohen and S. L. Glashow, Pair Creation Constrains Superluminal Neutrino Propagation, Phys. Rev. Lett. **107** (2011) 181803.

79. L. Alvarez-Gaume and M. A. Vazquez-Mozo, General properties of noncommutative field theories, Nucl. Phys. **B668**, 293 (2003).

80. S. A. Abel, J. Jaeckel, V. V. Khoze and A. Ringwald, *Vacuum birefringence as a probe of Planck scale noncommutativity,* JHEP **0609**, 074 (2006)

81. F. T. Brandt, A. K. Das, J. Frenkel, Dispersion relations for the selfenergy in noncommutative field theories, Phys. Rev. **D66** (2002) 065017.

82. R. J. Szabo, Quantum Gravity, Field Theory and Signatures of Noncommutative Spacetime, Gen. Rel. Grav. **42** (2010) 1-29.

Marko Vojinović[28]

Category Theory in Spincube Model of Quantum Gravity[29]

Institute of Physics, University of Belgrade,
118, Pregrevica, Belgrade, Serbia
e-mail: vmarko@ipb.ac.rs

We give a brief review of the problem of quantum gravity. After the discussion of the nonrenormalizability of general relativity, we briefly mention the main research directions which aim to resolve this problem. Our attention then focuses on the approach of Loop Quantum Gravity, specifically spinfoam models. These models have some issues concerning the semiclassical limit and coupling of matter fields. The recent developments in category theory provide us with the necessary formalism to introduce a new action for general relativity and perform covariant quantization so that the issues of spinfoam models are successfully resolved.

[28]This work was supported in parts by the Project-OI171031 of the Serbian Ministry of Education, Science and Technological Development.

[29]Based on invited talks given at the 5th Petrov International Symposium on *High Energy Physics, Cosmology and Gravity* (April 29–May 5, 2012, Kyiv, Ukraine), which were partially supported by the Project No. 1202.094-12 of the Central European Initiative Cooperation Fund.

1 Introduction

It is well known that Einstein's theory of General Relativity is not straightforward to quantize. This is easily seen from the fact that GR is not perturbatively renormalizable. Simply put, one can attempt to qunatize GR as an ordinary spin-two field in flat Minkowski spacetime, in the following way (for a nice review see [1]). Starting from the usual Einstein-Hilbert action

$$S_{EH} = \int d^4x \sqrt{-g} R,$$

one rewrites the metric tensor $g_{\mu\nu}$ as the flat Minkowski metric $\eta_{\mu\nu}$ and the spin-two field $h_{\mu\nu}$ as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu},$$

and substitutes it into the action, rewriting it in terms of the new variable $h_{\mu\nu}$. Thereby one obtains

$$S_{EH} = \int d^4x \, h_{\mu\nu} \Box h^{\mu\nu} + (\text{gauge fixing terms}) + \\ + (\text{self} - \text{interaction terms}).$$

The D'Alambertian operator is defined in flat Minkowski space, $\Box \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu$. From here one can proceed to perform the standard field theory quantization in the naive way — first formulate the free quantum field theory, and then introduce interactions perturbatively.

However, very soon one is bound to face the difficulty of nonrenormalizability of this theory. The tree-level Feynman diagrams are finite, the one-loop divergences can be removed by wavefunction renormalization, but at the two-loop level a Lagrangian counterterm of the form

$$\mathcal{L}_2 = \frac{const}{\varepsilon^2} R^{\alpha\beta}{}_{\mu\nu} R^{\mu\nu}{}_{\rho\sigma} R^{\rho\sigma}{}_{\alpha\beta} \qquad (\varepsilon \to 0)$$

appears [2], which is nonzero on-shell. Here $\varepsilon = 4 - D$ is the cutoff parameter from dimensional regularization scheme. At higher loop levels similar terms involving R^4, R^5, etc. terms are also expected to appear, rendering the theory perturbatively nonrenormalizable. This means that in order to remove all divergences one needs to introduce at least one additional coupling constant for each loop level. The infinite number of these coupling constants implies the loss of predictive power of the theory, since all experiments doable in principle can only ever fix a finite number of coupling constants. This property of General Relativity has been known for quite some time, and there are various research directions which attempt to address this issue. They can be broadly separated into two classes, by the methodology.

The first class of approaches considers modifying or substituting GR by another theory, which should preferably be renormalizable. Such attempts have evolved into vast research directions such as supergravity, string field theory, noncommutative geometry, and so on. The goal of each proposed model is to have a renormalizable theory that looks like GR at least on the length scales which can be tested experimentally, while at the same time have only a finite number of coupling constants. These coupling constants could then in principle be used to predict the values of the infinite set of coupling constants appearing in the perturbative quantum gravity approach.

The second class of approaches is based on the point of view that abandons the renormalization paradigm, and essentially gives physical meaning to the cutoff parameters of some particular regularization scheme. In other words, the assumption is that at some scale (typically expected to be near the Planck scale) expectation values of the physical observables will start to depend explicitly on cutoff parameters. This dependence is assumed to be measurable (in principle), rather than being removed by renormalization. These attempts have also evolved into vast research directions such as loop quantum gravity, causal dynamical triangulations, causal set theory, etc. The goal of all proposed models is exactly the same as before — predict some definite values for the infinite number of coupling constants present in the perturbative quantum gravity. All these research directions have had limited success, and in the absence of any experimental data relevant at the Planck scale, none of these directions can be preferred over the others.
In what follows, we shall be mainly concerned with the approach of loop quantum gravity (for a review see [3]), more specifically spin foam models, and we shall propose one novel particular model that addresses some serious issues present in all other spin foam models so far.

In section 2 we shall give a short overview of the status of LQG in general and spin foam models in particular. We will argue that the main drawbacks of all 4D spin foam models stem from the fact that tetrad fields are not basic variables of the theory. Section 3 deals with the categorical generalization of the Poincaré group, called the Poincaré 2-group. This will give us the necessary mathematical tools to reformulate the GR action in a convenient way which includes tetrad fields as basic variables. The analysis of this new action is then given in section 4, with a sketch of a quantization procedure giving rise to the so-called spincube model. Section 5 contains conclusions and discussion of the results.

2 Loop Quantum Gravity and Spin Foam Models

A detailed review of the Loop Quantum Gravity approach can be found in [3]. Here we just give some basic properties at an informal level.

The basic idea of LQG is to choose diffeomorphism-invariant quantities as basic degrees of freedom for the gravitational field, and then perform a canonical nonperturbative quantization of gravity in terms of these quantities. The natural candidates for basic variables turned out to be Wilson loops, and subsequently their generalizations called spin networks. This choice of variables introduces a natural diffeomorphism-invariant cutoff at the Planck length scale l_P, thereby rendering the theory UV-finite. The quantization is performed in the Schrödinger picture, and provides one with a mathematically well-defined constructions of the kinematical Hilbert space for the theory and some basic operators for geometric observables such as lengths, areas and volumes of space. Evolution in time is embodied in the Hamiltonian constraint, corresponding to the Wheeler-de Witt equation in the LQG setting.

The main features of such canonical approach to quantization are as follows. The theory represents a nonperturbative quantization of GR, and can in principle be applied to the study of physical systems where gravity is the dominant factor at short distances — such systems include the black hole and cosmological singularities. It gives one a mathematical handle on a well-defined Hilbert space of states for the gravitational field, thereby giving some insight into the quantum mechanical features of gravity. The natural basis for the Hilbert space is the set of the *spin network states*, combinatorial graphs colored by the irreducible representations of the $SU(2)$ group, and corresponding intertwiners. Finally, the study of the geometric observables — the length, area and volume operators — reveals that each of them has a discrete spectrum, giving rise to the geometric interpretation of the gravitational field wavefunctional, as well as the discrete character of space.

The theory also has some drawbacks. First, the Hamiltonian constraint is not uniquely defined, due to the usual ordering problems present in quantum mechanics. Second, even if one chooses some particular ordering, the Hamiltonian constraint is extremely complicated and impossible to solve in practice. This severely limits the possibility for any practical calculations and the study of the dynamics of the theory. As the main obstacle, the proof of the correct semiclassical limit of the theory is still missing, as well as any attempt to predict the coupling constants from the perturbative gravity approach.

A way to resolve these drawbacks has been found in the spin foam approach [4]. The idea is to give up canonical quantization, but instead attempt a covariant, path-integral quantization of the theory. Building on the results of the canonical approach, one wants to define the gravitational path-integral

$$Z = \int \mathcal{D}g_{\mu\nu} \exp\left(iS_{EH}[g_{\mu\nu}]\right)$$

in some way, in order to be able to calculate expectation values of observables, both in deep quantum regime and the semiclassical regime. This approach tends to give one a good handle on the dynamics of the theory, in addition to all features of the canonical approach.

The basic procedure of defining Z goes as follows. One starts from the Plebanski action for General Relativity,

$$S = \int B_{ab} \wedge R^{ab} + \phi^{abcd} B_{ab} \wedge B_{cd}.$$

The first part of this action represents the topological BF theory for the $SO(3,1)$ group. The R^{ab} is the curvature 2-form, a field strength "F" for the $SO(3,1)$ connection 1-

form ω^{ab}. The B_{ab} is the Lagrange multiplier 2-form. The second part of the action is the Plebanski constraint, featuring B_{ab} and the 0-form Lagrange multiplier ϕ^{abcd}. The purpose of the constraint is to enforce the B_{ab} to be a simple 2-form (i.e. an exterior product of two 1-forms). This constraint is therefore called "simplicity constraint", and it can be shown that the simplicity requirement of the B_{ab} field is enough to convert the topological BF theory into General Relativity. The fact that B_{ab} is simple gives rise to nontrivial degrees of freedom in the theory, reducing the equation of motion for ω^{ab} from Riemann-flat to Ricci-flat.

The second step is the quantization of the topological BF theory. This can be done in a rigorous way by employing the methods of topological quantum field theory. One first discretizes spacetime into 4-simplices, motivated by the structure of space in the canonical LQG, and rewrites the BF action in the form

$$\int B_{ab} \wedge R^{ab} \stackrel{\text{discr.}}{\longrightarrow} \sum_{\triangle} B_{\triangle} R_{\triangle},$$

where the sum goes over all triangles in the triangulation. Then one defines a topological invariant

$$\begin{aligned} Z &\equiv \int \mathcal{D}\omega \int \mathcal{D}B \exp\Big(i \sum_{\triangle} B_{\triangle} R_{\triangle} \Big) = \\ &= \sum_{\Lambda} \prod_{f} A_2(\Lambda_f) \prod_{v} A_4(\Lambda_v). \end{aligned}$$

Here Λ are the irreducible representations of $SO(3,1)$, labelling the faces f, edges e and vertices v of the Poincaré dual lattice corresponding to the triangulation. The colored 2-complex dual to the spacetime triangulation is called a *spin foam*. The amplitudes $A_2(\Lambda)$ and $A_4(\Lambda)$ are determined such that Z is in fact a topological invariant — the total expression must not depend on the particular choice of the spacetime triangulation. In that way one arrives at the TQFT corresponding to the BF theory for the $SO(3,1)$ group, commonly called the *Ooguri spin foam model*. Of course, the invariant Z may be (and actually is) badly divergent, but that is not important at this stage, since we are only interested in the structure of the path integral.

The last step in the quantization procedure is to enforce the simplicity constraint on the BF path integral at the quantum level. The exact technique for this is quite involved [5, 6], but the bottomline is that one projects the $SO(3,1)$ irreducible representations Λ to the $SU(2)$ representations present in the canonical LQG formalism, in order to obtain the same structure of the Hilbert space on the spin foam boundary. The resulting theory is not topologically invariant, but represents one possible rigorous definition for the theory of quantum gravity. The most advanced spin foam model in this respect is the EPRL/FK model, developed independently by two research groups [5, 6].

The main feature of spin foam models is that they correct some drawbacks of the canonical theory, primarily the dynamical sector is more under control. In addition, there remains a certain ambiguity in the choice of the amplitudes A_2 and A_4. This can be conveniently utilised to redefine the model such that it becomes IR-finite and to have a correct semiclassical limit [7, 8]. One can also employ standard QFT methods to calculate the effective action for the model in the semiclassical limit, which opens a possibility to explicitly determine the coupling constants from perturbative quantum gravity. Unfortunately, the spin foam models introduce their own set of problems. Aside from the "unusual" properties like fuziness of geometry at the Planck scale, all spin foam models suffer from two major handicaps. The first is related to the fact that, in addition to the good semiclassical limit, all models have *additional semiclassical limits*, which do not give rise to the standard GR, but to the so-called *area-Regge geometry*. Since these different classical limits are not observed in experiments, one needs some additional mechanism to supress such solutions. However, so far no mechanism could be constructed to deal with this problem.

The second handicap is related to the inability of the spin foam models to couple matter fields to gravity. Namely, the basic geometric variables which are employed in description of spacetime geometry are areas and volumes of space, but not lengths. This situation makes it extremely complicated (and in the case of massive fermionic matter even impossible) to incorporate matter fields into the spin foam model. Even if doable (see [9] for the massless fermion coupling), the resulting theory would be too complicated to be useful for any calculation.

As it turns out, both of these handicaps have a common origin — the edge lengths in the triangulation are not well-defined at the quantum level. This is itself a consequence of the choice of spin network states as basic degrees of freedom in the canonical LQG — the choice which emphasizes the spin connection ω^{ab}, while entirely ignoring the tetrad fields e^a. At the level of spin foam models, it is easy to see that the Plebanski constraint was purposefully designed to require the simplicity of B_{ab}, while avoiding to explicitly state that (the dual of) B_{ab} is the product of two tetrad 1-forms. The reason for this is that the tetrad fields do not appear as variables in the topological BF sector of the theory, which is being used for the definition of the path integral.

In the remainder of this paper we will present a novel way to address this main difficulty, and to introduce tetrad fields explicitly in the topological sector of the theory. However, in order to do this, it is important to introduce some mathematical concepts which provide the background formalism for the new model.

3 Poincaré 2-group

We begin by giving a very brief review of the so-called *categorification ladder*, an important and active research topic in category theory. We shall not attempt at any rigor, leaving out most of the details, which can be found for example in [10] and references therein.

In the branch of mathematics called *category theory*, one defines a structure called a *category* as a set of *objects* and a set of *morphisms* between those objects, satisfying some basic axioms. Such a structure is fairly general and does not have many interesting properties itself. However, this generality allows one to use it for all sorts of purposes. For example, one can define the usual structure of a *group* as a category which has only one object, while all morphisms (mapping the object onto itself) are invertible. The composition rules for the morphisms can be chosen to be the group multiplication, thereby providing an isomorphism between a given group and the corresponding category with one element.

The first step in the categorification ladder is to introduce the concept of a *2-category*. A 2-category consists of a set of objects, a set of morphisms and a set of *2-morphisms*, maps between morphisms. Intuitively, if a category can be represented by a linear graph of dots (objects) and arrows connecting them (morphisms), a 2-category can be represented by a planar graph, consisting of dots (objeects), arrows connecting them (morphisms) and "surface arrows" mapping one arrow into another (see [10] for details and pictures). The main point is that the dimensionality of the graph has been raised by one. The categorification ladder can continue by introducing a 3-category (or in general an n-category) by a similar process, leading to 3-dimensional (in general n-dimensional) graphs.

In analogy with a group, one can then define a 2-group, as a 2-category which has only one element, while all morphisms and 2-morphisms are invertible. A 2-group is a categorical generalization of a group, and is not a group itself. One can prove that any 2-group is equivalent to a *crossed module*, a structure that has been studied independently by mathematicians before the idea of the categorification ladder has even been introduced. A crossed module is a quadruple $(G, H, \partial, \triangleright)$. This is a pair of groups G and H, such that $\partial : H \to G$ is a homomorphism and $\triangleright : G \times H \to H$ is an action of G on H such that certain axioms are satisfied, which turn out to be directly related to the structure of a 2-category, see [10]. The elements of G represent the 1-morphisms, while the elements of the semidirect product $G \ltimes H$ represent the 2-morphisms. The canonical example of a 2-group relevant for physics is the Poincaré 2-group, where $G = SO(3,1)$, $H = \boldsymbol{r}^4$, ∂ is a trivial homomorphism and $\triangleright$ is the usual action of the Lorentz transformations on the $\boldsymbol{r}^4$ space. The Lorentz group is the group of morphisms, while the usual Poincaré group is the group of 2-morphisms.

The main feature of the whole 2-group formalism is that one can generalize the concept of a *holonomy* along a line to its two-dimensional analog — a *surface holonomy*. The initial interest in this came from string theory. A point-particle travels along a world line in spacetime, and one is naturally led to the concept of a parallel transport along a given line. String theory promotes the point particle into a one-dimensional object — a string — which then travels along a world surface in spacetime. Thus one would like to have a concept of a *parallel transport along a given surface*.

One of the main aims of the 2-category and 2-group formalism is to introduce and formalize this concept.

Given the strong categorical relationship between groups and 2-groups, one can construct a gauge theory on a 4-manifold $\mathcal{M}$ based on a crossed module $(G, H, \partial, \triangleright)$ of Lie groups by using 1-forms A, which take values in the Lie algebra $\mathfrak{g}$ of G, and 2-forms β, which take values in the Lie algebra $\mathfrak{h}$ of H [11,12]. The forms A and β transform under the usual gauge transformations $g : \mathcal{M} \to G$ as

$$A \to g^{-1}Ag \,+\, g^{-1}dg\,, \quad \beta \to g^{-1} \rhd \beta\,,$$

while the gauge transformations generated by H are given by

$$A \to A \,+\, \partial\eta\,, \quad \beta \to \beta + d\eta + A \wedge^{\rhd} \eta + \eta \wedge \eta\,,$$

where η is a one-form taking values in $\mathfrak{h}$, see [12]. When the group H is Abelian, which happens in the Poincaré 2-group case, then the $\eta \wedge \eta$ term vanishes, and one obtains the gauge transformations given in [11].

The pair (A, β) represents a 2-connection on a 2-fiber bundle associated to the 2-Lie group (G, H) and the manifold $\mathcal{M}$. The corresponding curvature forms are given by

$$\mathcal{F} = dA + A \wedge A - \partial\beta\,, \quad \mathcal{G} = d\beta + A \wedge^{\rhd} \beta\,,$$

and they transform as

$$\mathcal{F} \to g^{-1}\mathcal{F}g\,, \quad \mathcal{G} \to g^{-1} \rhd \mathcal{G}\,,$$

under the usual gauge transformations, while

$$\mathcal{F} \to \mathcal{F}\,, \quad \mathcal{G} \to \mathcal{G} + \mathcal{F} \wedge^{\rhd} \eta\,,$$

under the H-gauge transformations.

One can introduce a natural topological gauge theory determined by the vanishing of the 2-curvature

$$\mathcal{F} = 0\,, \quad \mathcal{G} = 0\,.$$

These equations can be obtained from the action

$$S = \int \langle B \wedge \mathcal{F} \rangle_{\mathfrak{g}} + \langle C \wedge \mathcal{G} \rangle_{\mathfrak{h}}\,,$$

where B is a Lagrange multiplier 2-form taking values in $\mathfrak{g}$, C is a Lagrange multiplier 1-form taking values in $\mathfrak{h}$, $\langle\ ,\ \rangle_{\mathfrak{g}}$ is a G-invariant nondegenerate bilinear form in $\mathfrak{g}$ and $\langle\ ,\ \rangle_{\mathfrak{h}}$ is a G-invariant nondegenerate bilinear form in $\mathbf{h}$. This action is called $BFCG$ action, in analogy with the BF theory action. The gauge transformations of the Lagrange multiplier fields are given by

$$B \to g^{-1}Bg\,, \quad C \mapsto g^{-1} \rhd C\,,$$

for the usual gauge transformations, while

$$B \to B - [C, \eta]\,, \quad C \mapsto C\,,$$

for the H-gauge transformations.

Let us now examine the case of the Poincaré 2-group. In this case $A = \omega^{ab} J_{ab}$, $\beta = \beta^a P_a$, where $a, b \in \{0,1,2,3\}$, J_{ab} are the generators of the Lorentz group while P_a are the generators of the translation group $\boldsymbol{r}^4$. Consequently

$$\begin{array}{rclcl} \mathcal{F} & = & (d\omega^{ab} + \omega^a{}_c \wedge \omega^{cb}) J_{ab} & = & R^{ab} J_{ab}, \\ \mathcal{G} & = & \left(d\beta^a + \omega^a{}_b \wedge \beta^b\right) P_a & = & (\nabla \beta^a) P_a. \end{array}$$

The G-gauge transformations are the local Lorentz rotations

$$\omega \to g^{-1} \omega g + g^{-1} dg \,, \quad \beta \to g^{-1} \triangleright \beta \,,$$

while the H-gauge transformations are the local translations

$$\delta_\varepsilon \omega^{ab} = 0 \,, \quad \delta_\varepsilon \beta^a = d\varepsilon^a + \omega^a{}_b \wedge \varepsilon^b \,,$$

where $\eta = \varepsilon^a P_a$.

The $BFCG$ action then becomes

$$S = \int_{\mathcal{M}} \left(B^{ab} \wedge R_{ab} + C_a \wedge \nabla \beta^a\right) \,,$$

where

$$\delta_\varepsilon B = 0 \,, \quad \delta_\varepsilon C = 0 \,.$$

At this point a very important observation is in order. The transformation properties of the 1-form C^a are the same as the transformation properties of the tetrad 1-form e^a under the local Lorentz and the diffeomorphism transformations. In addition, the equation of motion for C^a is $\nabla C^a = 0$, just like the no-torsion equation for the tetrad, $\nabla e^a = 0$. Based on this, *we identify the Lagrange multiplier C^a with the tetrad field e^a*, and write the action in the form

$$S = \int_{\mathcal{M}} \left(B^{ab} \wedge R_{ab} + e^a \wedge \nabla \beta_a\right) \,.$$

In this way one can construct a categorical generalization of the topological BF action. The new action is again topological, but more rich in structure, since the tetrad fields are explicitly present. In addition, the 2-group formalism provides a framework to construct a topological quantum field theory from this action, in analogy with the BF case. This provides us with the necessary tools to construct a categorical generalization of a spin foam model, based on the $BFCG$ action instead of the BF action. The explicit presence of the tetrads should help us resolve the two handicaps of spin foam models discussed in section 2.

4 The Spincube Model

The first step in the construction of the new model is to write the action for General Relativity, starting from the $BFCG$ action. In order to do this, all we need is the simplicity constraint,

$$B_{ab} = \varepsilon_{abcd}\, e^c \wedge e^d \,,$$

which can now be added into the action as it stands, as opposed to the BF case where the Plebanski constraint had to be introduced due to the absence of the tetrads e^a in the BF action. Therefore, one can write the *constrained* $BFCG$ *action* in the form

$$\begin{aligned} S \;=\; \int_{\mathcal{M}} \Big[& B^{ab} \wedge R_{ab} + e^a \wedge \nabla\beta_a - \\ & - \phi_{ab} \wedge \left(B^{ab} - \varepsilon^{abcd} e_c \wedge e_d \right) \Big] \,, \end{aligned} \tag{1}$$

where ϕ_{ab} is an additional Lagrange multiplier 2-form field, introduced in order to enforce the simplicity constraint.

The equations of motion are obtained by varying S with respect to B, e, ω, β and ϕ, respectively, to give:

$$\begin{aligned} & R_{ab} - \phi_{ab} = 0 \,, \\ & \nabla\beta_a + 2\varepsilon_{abcd}\phi^{bc} \wedge e^d = 0 \,, \\ & \nabla B_{ab} - e_{[a} \wedge \beta_{b]} = 0 \,, \\ & \nabla e_a = 0 \,, \\ & B_{ab} - \varepsilon_{abcd} e^c \wedge e^d = 0 \,. \end{aligned}$$

With the usual assumption that the tetrad fields are nondegenerate, these equations can be reworked into an equivalent form:

$$\phi^{ab} = R^{ab}, \qquad B_{ab} = \varepsilon_{abcd} e^c \wedge e^d, \qquad \beta^a = 0,$$

$$\nabla e^a = 0 \,, \qquad \varepsilon_{abcd} R^{bc} \wedge e^d = 0 \,.$$

The first three equations determine β^a and the multipliers B_{ab} and ϕ_{ab} in terms of e^a and ω^{ab}. The fourth equation is the no-torsion equation, which determines the connection ω^{ab} to be the Levi-Civita connection (a function of the tetrads e^a). The last equation is nothing but the Einstein field equation for the only remaining field e^a. Thus we see that the action (1) is classically equivalent to General Relativity. More precisely, it is equivalent to the Einstein-Cartan theory,

$$S_{EC} = \int_{\mathcal{M}} \varepsilon_{abcd} e^a \wedge e^b \wedge R^{cd} \,,$$

since the torsion is equal to zero as an equation of motion rather than by definition.

Given the new action for General Relativity, we can proceed with the covariant quantization in analogy with the spin foam models. The action has the topological term and the constraint term, so as a first step we construct a topological quantum field theory by defining the path integral for the $BFCG$ part of the action. In the second step, we enforce the constraint term by requiring a suitable restriction in the path integral of the topological theory.

One begins by triangulating spacetime into 4-simplices, and rewriting the topological part of the action in the form

$$\sum_{\triangle} B_{\triangle} R_{\triangle} + \sum_{l} e_l (\nabla\beta)_l,$$

where the first sum goes over all triangles and the second goes over all edges in the triangulation of the spacetime manifold. Then one constructs a topologically invariant path integral in the form (see [13] for the details of the construction)

$$
\begin{aligned}
Z &\equiv \int \mathcal{D}\omega \int \mathcal{D}B \int \mathcal{D}e \int \mathcal{D}\beta \\
&\quad \exp\Big(i \sum_{\triangle} B_{\triangle} R_{\triangle} + i \sum_{l} e_l (\nabla\beta)_l \Big) = \\
&= \sum_{\Lambda} \prod_{p} A_1(\Lambda_p) \prod_{f} A_2(\Lambda_f) \prod_{v} A_4(\Lambda_v).
\end{aligned} \tag{2}
$$

The labels $\Lambda = (L_p, m_f)$, where $L_p \in \boldsymbol{r}_0^+$ and $m_f \in \mathbb{Z}$, are now irreducible representations of the Poincaré 2-group, and in addition to vertices v and faces f of the Poincaré dual lattice, we also take the product over all the polyhedra p, since they are dual to the edges of the triangulation and naturally appear in the construction due to the presence of the $e \wedge \nabla\beta$ term in the $BFCG$ action. The amplitudes $A_1(\Lambda)$, $A_2(\Lambda)$ and $A_4(\Lambda)$ are chosen so that Z does not change under the action of the Pachner moves, which guarantees its independence of the triangulation. The polyhedra are colored with L_p, which have the interpretation as lengths of triangulation edges, while faces are colored with m_f, which have the interpretation as areas of the triangles in the triangulation. In the topological theory, edge lengths and triangle areas are independent of each other.

Note that the path integral is not defined over a colored 2-complex (the spinfoam), but rather over a colored 3-complex (called *spincube*).

Finally, we can impose the simplicity constraint, in order to turn the topological path integral into a realistic model for quantum gravity. Based on the geometric interpretation of the variables, the constraint actually says that a very natural requirement should be enforced — the triangle areas must be compatible with the corresponding edge lengths. This can be formalized in the requirement

$$
|m_f| l_P^2 = A_f(L), \quad \forall f
$$

where $A_f(L)$ is the Heron formula for the triangle area in terms of its edges. The Planck length appears naturally in order to balance the dimensions of the two sides of the equation. As a last step, one redefines the amplitudes A_1, A_2 and A_4 in order to render the theory IR-finite, as well as to enforce the correct semiclassical limit, in a way similar to the spinfoam models.

Note that imposing this constraint leaves only edge lengths as independent variables in the theory, so that the “area-Regge” problem present in spinfoam models is resolved automatically. In addition, the edge length variables allow for a completely straightforward coupling of matter fields to the spincube model. Namely, at the level of the classical theory, one can introduce fermions via the action

$$
\begin{aligned}
S \;=\; & \int \Big[B^{ab}\wedge R_{ab} + e^a\wedge\nabla\beta_a - \phi_{ab}\wedge\left(B^{ab}-\varepsilon^{abcd}e_c\wedge e_d\right)\Big]+ \\
& + i\kappa_1 \int \varepsilon_{abcd}\, e^a\wedge e^b\wedge e^c\wedge\bar\psi\left[\gamma^d \overset{\leftrightarrow}{d} + \{\omega,\gamma^d\} + \frac{im}{2}\,e^d\right]\psi + \\
& + i\kappa_2 \int \varepsilon_{abcd} e^a\wedge e^b\wedge\beta^c\,\bar\psi\gamma_5\gamma^d\psi\,,
\end{aligned} \tag{3}
$$

where $\omega=\omega_{ab}[\gamma^a,\gamma^b]/8$, $\kappa_1 = 8\pi l_P^2/3$ and $\kappa_2=-2\pi l_P^2$. The first term is the constrained $BFCG$ action, while the second and third terms introduce fermion coupling which results in the same equations of motion as in the ordinary Einstein-Cartan theory with fermions.

The quantization procedure of the action (3) is essentially the same as the one without fermions. The only difference is in the fact that the vertex amplitude A_4 will change to reflect the presence of the fermionic matter, as

$$A_4 \to A_4 \exp\left[iS_R^{(\mathrm{ferm})}(L,\psi)\right]\,,$$

where $S_R^{(\mathrm{ferm})}$ is the Regge discretized action of a fermion field ψ coupled to gravity. The expressions which appear in $S_R^{(\mathrm{ferm})}$ can be easily obtained, in contrast to the EPRL/FK model case, where the expression for the 4-simplex volume is impossible to define uniquely in terms of the spin foam variables [9].

Similarly to (3), one can also couple other matter fields to (1) in a completely straightforward way, including gauge and scalar fields, the cosmological constant, the Holst term, and so on.

5 Conclusions

The proposed 2-group reformulation of GR can be used to obtain a categorical ladder generalization of Loop Quantum Gravity. The advantage of this generalization is that the edge lengths of a triangulation become the basic dynamical variables. This will facilitate the construction of the path integral such that the classical limit of the corresponding quantum theory is GR and the coupling of matter will be much easier to accomplish.

The categorical nature of the theory implies that the edge labels of a spacetime triangulation should be the 2-group irreducible representations on a 2-Hilbert space. Note that this is not unique, since one can also use the category of chain complexes of vector spaces in order to define the representations, see [12, 14]. The structure of the chain-complex representations is different from the 2-Hilbert space representations, which means that chain-complex representation theory defines an alternative quantization of GR. Hence it would be interesting to develop the chain-complex representation theory of the Poincaré 2-group.

The physical significance of 2-Hilbert space representations could be better understood by performing a canonical quantization of the action (1).

As far as the construction of 4-manifold invariants based on the $BFCG$ state sum is concerned, one would have to regularize the topological state sum/integral based on the amplitude (2) such that the triangulation independence is preserved. One way to do it is to try to implement a gauge-fixing procedure, see [15]. Another way is to find a quantum group regularization, since there are strong indications that categorified quantum groups and their representations will be important for the construction of 4-manifold invariants [16]. Hence one can try to find a crossed module of Hopf algebras which is a deformation of the Poincaré 2-group, and then try to find an appropriate 2-category of representations which will give a finite topological state sum.

References

1. H. W. Hamber, Discrete and Continuum Quantum Gravity, `arXiv:0704.2895`.
2. M. H. Goroff and A. Sagnotti, *Phys. Lett. B* **160**, 81 (1985); *Nucl. Phys.* **B266**, 709 (1986).
3. C. Rovelli, *Quantum gravity* (Cambridge University Press, 2004).
4. C. Rovelli, Zakopane lectures on loop gravity, `arXiv:1102.3660`.
5. J. Engle, E. R. Livine, R. Pereira and C. Rovelli, *Nucl. Phys.* **B799**, 136 (2008).
6. L. Freidel and K. Krasnov, *Class. Quant. Grav.* **25**, 125018 (2008).
7. A. Miković and M. Vojinović, *Class. Quant. Grav.* **28**, 225004 (2011).
8. A. Miković and M. Vojinović, Finiteness of the ELPR/FK spin foam model, `arXiv:1101.3294`.
9. E. Bianchi, M. Han, E. Magliaro, C. Perini, C. Rovelli and W. Wieland, Spinfoam fermions, `arXiv:1012.4719`.
10. J. C. Baez and J. Huerta, An invitation to higher gauge theory, `arXiv:1003.4485`.
11. F. Girelli, H. Pfeiffer and E. M. Popescu, *J. Math. Phys.* **49**, 032503 (2008).
12. J. Faria Martins and A. Miković, *Adv. Theor. Math. Phys.* **15**, 1059 (2011), `arXiv:1006.0903`.
13. Class. Quant. Grav. **29** 165003 (2012), `arXiv:1110.4694`.
14. L. S. Cirio and J. Faria Martins, *Diff. Geom. App.* **30** 238 (2012), `arXiv:1106.0042`.
15. A. Baratin and L. Friedel, *Class. Quant. Grav.* **24**, 2027 (2007).
16. L. Crane and I. B. Frenkel, *J. Math. Phys.* **35**, 5136 (1994).