Embedded graphs

S. K. Lando, A. K. Zvonkin

March 2, 2000

Alors que dans mes recherches d'avant 1970, mon attention systématiquement était dirigée vers les objets de généralité maximale, afin de dégager un language d'ensemble adéquat pour le monde de la géométrie algébrique, et que je ne m'attardais sur les courbes algébrique, que dans la stricte mesure où cela s'averais indispensable (notamment en cohomologie étale) pour développer des techniques et énoncés "pass–partout" valables en toutes dimensions et en toutes lieux (j'entends, sur tous schémas de base, voire tous topos annelés de base ...), me voici donc ramené par le truchement d'objets si simples qu'un enfant peut les connaître en jouant, aux débuts et origines de la géometrie algébrique, familiers à Riemann et à ses émules!

A. Grothendieck, *Esquisse d'un Programme*

Contents

Chapter 1

Embedded graphs, permutations and groups

1.1 Introduction

1.2 Graphs and maps

1.2.1 Two-dimensional manifolds

1.2.2 Genus and Euler characteristic

1.2.3 Hypermaps, duality and trinity

1.3 Examples and constructions

1.3.1 Plane trees

1.3.2 Icosahedron of genus 4

1.4 Cartographic group and monodromy group

1.4.1 Map automorphisms

1.4.2 Cartographic generation of Mathieu groups

1.4.3 Other examples

1.5 Map enumeration

1.5.1 Generating functions

1.5.2 Enumeration of trees

1.5.3 Inverse problem of enumeration

1.6 What this book does not contain

Bibliography

[1] **Adrianov N. M., Kochetkov Yu. Yu., Shabat G. B., Suvorov A. D.** Mathieu groups and plane trees. – Preprint, March 1995, 9 pp.

[2] **Babai L.** The probability of generating the symmetric group. – *J. of Combinat. Theory A*, 1989, vol. **52**, 148–153.

[3] **Brahana H. R.** Systems of circuits on 2-dimensional manifolds. – *Ann. Math.*, 1921, vol. **23**, 144–168.

[4] **Brahana H. R.** Pairs of generators of the known simple groups whose orders are less than one million. – *Ann. Math.*, 1930, vol. **31**, no. 2, 529–549.

[5] **Butler G.** Fundamental Algorithms for Permutation Groups. – Lecture Notes in Computer Science, vol. 559, Springer, 1991.

[6] **Butler G., McKay J.** The transitive groups of degree up to eleven. – *Commun. in Algebra*, 1983, vol. **11**, no. 8, 863–911.

[7] **Conder M.** The symmetric genus of the Mathieu groups. – *Bull. of the London Math. Soc.*, 1991, vol. **23**, no. 5, 445–453.

[8] **Conder M.** Generating the Mathieu groups and associated Steiner systems. – *Discrete Math.*, 1993, vol. **112**, 41–47.

[9] **Conder M. D. E., Wilson R. A., Woldar A. J.** The symmetric genus of sporadic groups. – *Proc. Amer. Math. Soc.*, 1992, vol. **116**, no. 3, 653–663.

[10] **Conway J. H., Curtis R. T., Norton S. P., Parker R. A., Wilson R. A.** (with computational assistance from **Thackray J. G.**) ATLAS of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups. – Clarendon Press, Oxford, 1985, xxxiv + 252 pp.

[11] **Conway J. H., Sloane N. J. A.** Sphere Packings, Lattices and Groups. – Springer, 1988.

[12] **Cori R., Machi A.** Construction of maps with prescribed automorphism group. – *Theoret. Comput. Sci.*, 1982, vol. **21**, 91–98.

[13] **Cori R., Machi A.** Maps, hypermaps and their automorphisms: a survey, I, II, III. — *Expositiones Mathematicae*, 1992, vol. **10**, 403–427, 429–447, 449-467.

[14] **Edmonds J. R.** A combinatorial representation for polyhedral surfaces. – *Notices Amer. Math. Soc.*, 1960, vol. **7**, p. 646.

[15] **Hamilton W. R.** Letter to John T. Graves "On the Icosian" (17th October 1856). – In "W. R. Hamilton, Mathematical Papers, vol. III, Algebra" (H. Halberstam, R. E. Ingram eds.), Cambridge Univ. Press, 1967, 612–625.

[16] **Harary F., Palmer E. M.** Graphical Enumeration. – Academic Press, New York and London, 1973

[17] **Heffter L.** Über das Problem der Nachbargebiete. – *Math. Ann.*, 1891, vol. **38**, 477–508.

[18] **Heffter L.** Über metacyklische Gruppen und Nachbarconfigurationen. – *Math. Ann.*, 1898, vol. **50**, 261–268.

[19] **Jones G. A.** Ree groups and Riemann surfaces. – *Journal of Algebra*, 1994, vol. **165**, no. 1, 41–62.

[20] **Jones G. A.** Characters and surfaces. – Preprint, September 1995, 16 pp.

[21] **Jones G. A., Silver S. A.** Suzuki groups and surfaces. – *J. London Math. Soc.*, 1993, vol. **48**, no. 2, 117–125.

[22] **Jones G. A., Silver S. A.** The monodromy genus of the Suzuki groups. – Contemporary Mathematics, vol. **184**, 1995, p. 215–224.

[23] **Jones G. A., Singerman D.** Maps, hypermaps and triangle groups. – Preprint, July 1993, 23 pp.

[24] **Kochetkov Yu. Yu.** Cartographic groups and genaralized Chebyshev polynomials (Kartograficheskie gruppy i obobshchennye polinomy Chebysheva). – *Uspekhi Matem. Nauk*, 1994, t. **49**, no. 6, 203–204.

[25] **Kochetkov Yu. Yu.** Plane trees, cartographical groups and Chebyshev polynomials. – Preprint, 1995, 29 pp.

[26] **Łuczak T., Pyber L.** On random generation of the symmetric group. – *Combin. Probab. Comput.*, 1993, vol. **2**, 505–512.

[27] **Malle G., Saxl J., Weigel Th.** Generation of classical groups. – *Geometriae Dedicata*, 1994, vol. **49**, 85–116.

[28] **Mathieu E.** Mémoire sur le nombre de valeurs que peut acquérir une fonction quand on y permute ses variables de toutes les manières possibles. – *J. Math. Pures et Appl.*, 1860, vol. **5**, 9–42.

[29] **Mathieu E.** Mémoire sur l'étude des fonctions de plusieurs quantités. – *J. Math. Pures et Appl.*, 1861, vol. **6**, 241–233.

[30] **Mathieu E.** Sur les fonctions cinq fois transitives de 24 quantités. – *J. Math. Pures et Appl.*, 1873, vol. **18**, 25–46.

[31] **Tutte W. T.** A census of planar maps. – *Canad. J. of Math.*, 1963, vol. **15**, 249–271.

[32] **Tutte W. T.** The enumerative theory of planar maps. – In "A Survey of Combinatorial Theory", North Holland, Amsterdam, 1973, 437–448.

[33] **Tutte W. T.** What is a map? – In *"New Directions in the Theory of Graphs"*, Academic Press, New York, 1973.

[34] **White A. T.** Graphs, Groups and Surfaces. – North Holland, Amsterdam, 1984 (2nd edition).

[35] **Zvonkin A. K.** How to draw a group. – In “Formal Power Series and Algebraic Combinatorics 95” (B. Leclerc, J.-Y. Thibon eds.), Paris, May 1995, 601–611. Submitted to *Discrete Mathematics*.

Chapter 2

Embedded graphs and algebraic number theory

2.1 Introduction

Embedded graphs, however simple they may look, are enormously rich with various mathematical structures. The notion of the cartographic group and of its various relatives may serve as an illustrations to this thesis. The present chapter provides an even more striking illustration: we describe here the relations of embedded graphs to algebraic number theory. But before we start discussing details, we invite the reader to just get fascinated in front of such a phenomenon: you make a simple sketchy drawing, and you get, as a result, a number field corresponding to it, an action of the absolute Galois group, and many other notoriously complicated objects!

In 1979 there appeared a paper of G. Belyi [6] dedicated to construction of some Galois groups. At the very end of the paper, as a kind of an appendix (a half-page of a journal text), Belyi proves a theorem which is now known as The Belyi theorem (and which is also the source of the notions of Belyi function, Belyi pair, Belyi complex, etc.).

Five years later a short but decisive intervention of Grothendieck took place. Alexandre Grothendieck is a legendary figure in XX-th century mathematics. After having revolutionized the whole body of algebraic geometry, he retired from any form of active research at the age of fourty two. Many years later he decided to return to mathematical research, and wrote a very informal text [19], that still remains unpublished though is widely cited. The Grothendieck's "Esquisse d'un programme" is full of brilliant though sometimes rather fuzzy mathematical ideas, mixed with memories, reflections and lyrical reminiscences. There he says, among other things, that the Belyi theorem was one of the two the most profound mathematical impression of all his life. (The first one took place in his childhood, when he was retained in a nazi concentration camp, and a woman who was giving him lessons in mathematics had told him the exact mathematical definition of a circle.) Grothendieck gave a new, more "visual" interpretation of Belyi theorem: the absolute Galois group $\Gamma = \mathrm{Gal}(\overline{\mathbf{Q}}|\mathbf{Q})$, one of the most intricate and complicated mathematical objects, acts on embedded graphs! In order to underline the simplicity of the objects under action, Grothendieck called them "dessins d'enfant" ("child's drawings" in French). This term somehow stick to the theory, and now the part of the theory of embedded graphs that studies the relations to algebraic number theory is called, even in English language literature, the theory of dessins d'enfants (see, for example, [33]).

The next important paper on the subject was that of G. Shabat and V. Voevodsky [37] in the collection of papers dedicated to the Grothendieck's 60-th anniversary. Then the usual explosion followed, with an avalanche of preprints, meetings, seminars etc., and now, only several years later, it is already impossible to cover the contents of the theory even in a book form.

For further reading we recommend the collection of papers [33]. We must also mention a pioneering work of B. Birch [10], which already contained practically the whole synthetic picture of all the connections we are planning to discuss.

The theory involves many complicated ideas, and we are going to present them one at a time. This is why the proofs of some of the basic results (the Belyi theorem is one of them) are postponed up to the end of the chapter, when the reader must already understand what we are talking about.

In the section that follows, we consider a case which is twice the simplest: that of plane trees (that is, maps of genus zero, and with only one face). We introduce a class of complex polynomials which are called *Shabat polynomials*; these are polynomials that have (at most) two critical values. In fact, they also have a third critical value at infinity, but we may temporarily forget about it. If we draw a segment connecting both critical values, and take the preimage of this segment under the polynomial in question, we get a plane tree. Every plane tree can be obtained in this way, and every Shabat polynomial corresponds to a plane tree. A very refreshing observation is that such a polynomial is essentially unique. Thus the tree acquires via this construction a canonical geometric form: it becomes a kind of rigid.

To find a Shabat polynomial corresponding to a tree is not an easy task. It requires a system of algebraic equations on its coefficients to be solved. This is how algebraic numbers comes on the scene. To every plane tree, and to every embedded graph in general, there corresponds a number field, namely, its *field of definition*. The absolute Galois group acts on this field by its automorphisms, and this action gives rise to the Galois action on trees themselves. In Section 2.3 we study this action and its various combinatorial invariants. An honorable place among these invariants belongs to cartographic group introduced in Chapter 1.

The next section deals with arbitrary planar maps. Here instead of polynomials we are forced to consider rational functions. A *Belyi function* is a rational function with (at most) three critical values, which we fix once and for all as 0, 1 and ∞. The vertices of the map are preimages of the ∞ (the multiplicity of the point being equal to the vertex degree). Inside each face we have a ponit called the "center" of the face: it is a preimage of 0, its multiplicity being equal to the face degree. Each edge contains its "middle point" which is a preimage of 1. The map itself is the full preimage of the segment $[1, \infty]$. Computationally the problem of finding a Belyi function is noticeably more difficult than that of finding a Shabat polynomial. Besides that, there appears a new phenomenon which does not exist in the case of trees: the coefficients of a Belyi function may not belong to the field of definition of the corresponding map.

The next case to study is that of maps of arbitrary genus. First, some terminology must be changed: the Riemann sphere $\overline{\mathbf{C}}$ must be replaced by a compact Riemann surface (or a complex algebraic curve) X of genus g, and a rational function from $\overline{\mathbf{C}}$ to $\overline{\mathbf{C}}$ must be replaced by a meromorphic function $f : X \to \overline{\mathbf{C}}$ with (at most) three critical values. Such a pair (X, f) is called a *Belyi pair*. Here, however, we stumble, to a conceptually very difficult point. While there is only one Riemann surface of genus 0, there are infinitely many of them for every genus $g \geq 1$. All of them are equivalent from the topological point of view (this means that they are homeomorphic to each other); but they are different as complex manifolds. Therefore *both* elements of the pair (X, f) must be computed: the function f just does not have any sense if the curve X has not been found before. Usually the computation of X is uncomparably more difficult than that of f. In the majority of cases we are unable not only to solve the system of equations, but we even don't know how to write it down! Here we come close to the famous "moduli space of complex algebraic curves", one of the highly complicated objects which turn out to lie in the nerve center of modern mathematics and theoretical physics. Up to now, no general methods of computation of Belyi pairs are known, and what we have is only a handful of examples. We hope however that their beauty at least partially outweighs their small number.

The class of Shabat polynomials and Belyi functions constitute very interesting classes of functions, rich with all sorts of connections. In Section 2.6 we try to present at least some of these. This section should have been much bigger, but the obvious conditions of finite space-and-time impose their limitations.

During the whole chapter we were trying our best to make our presentation as elementary as possible. There are, however, two places where an elementary exposition was out of our reach. The first is the use of the fundamental theorem of Galois theory in order to introduce the notion of the field of definition. The second one is the use of the Riemann existence theorem for Riemann surfaces in order to prove the existence of Belyi pair for any embedded graph. Both facts are impossible to circumvent even in the case of trees. They are, however, localized in the text, and if the reader skips them and believes us on word, this will not prevent him or her from understanding of everything else.

2.2 Plane trees and Shabat polynomials

2.2.1 Complex polynomials, their critical points and critical values

The basic fact we suppose our reader to know is that a complex polynomial of degree n usually has n complex roots. Consider a polynomial $P(z)$ of degree n with complex coefficients. It maps the complex plane of the variable z onto another one, say of the variable w. Take a point $w \in \mathbf{C}$ and consider its inverse image $P^{-1}(w) = \{z | P(z) = w\}$.

Usually, this set consists of n distinct points, i.e., solutions of the equation $P(z) = w$ for a given w (see Figure 2.1). But sometimes, for some specific

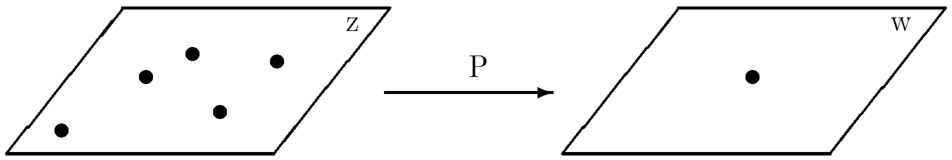


Figure 2.1: Preimage of a point

values of w, this equation can have multiple roots.

If a is a multiple root of multiplicity k of the equation $P(z) - w = 0$, where $k \geq 2$, then $P(z) - w = (z-a)^k p(z)$, $p(a) \neq 0$, and

$$P'(z) = (z-a)^{k-1}(kp(z) + (z-a)p'(z)).$$

Note that the factor $kp(z) + (z-a)p'(z)$ is not equal to zero at $z = a$. Hence a is a root of multiplicity $k-1$ of $P'(z)$.

A point $z \in \mathbf{C}$ at which $P'(z) = 0$ is called a *critical point*. A value $w = P(z)$ of the polynomial itself at a critical point z is called a *critical value*. The multiplicity k is called the *order* of the critical point. Obviously, if a critical point z has order k, then $P'(z) = 0$, $P''(z) = 0, \ldots, P^{(k-1)}(z) = 0$, $P^{(k)}(z) \neq 0$. The least possible order of a critical point is 2. However, in some situations by abuse of language we will call some *non*-critical points *critical points of order* 1. Usually it will mean "non-critical points which are preimages of critical values".

Since $P'(z)$ is a polynomial of degree $n-1$, the most usual situation is when $P(z)$ has $n-1$ distinct critical points, all of them of order 2, and $n-1$ distinct critical values. In more degenerate cases one may observe critical points of higher order, and/or several critical points may be sent by $P(z)$ to the same critical value.

What we are interested in is the most degenerate case, when the set of critical values of $P(z)$ is as small as possible.

Example 2.2.1 The polynomial $P(z) = z^n$ has only one critical point of order n (namely, $z = 0$), and only one critical value $w = 0$.

Up to obvious changes of variables, this polynomial is the only one with a single critical value. The next case, and a much more interesting one, is that of two critical values.

Example 2.2.2 *Chebyshev polynomials.* It is well known that $\cos n\varphi$ can be expressed as a polynomial of degree n in $\cos\varphi$. For example,

$$\cos 2\varphi = 2\cos^2\varphi - 1, \quad \cos 3\varphi = 4\cos^3\varphi - 3\cos\varphi, \quad \ldots,$$

and in general

$$\cos n\varphi = T_n(\cos\varphi),$$

where $T_n(z)$ is the n-th *Chebyshev polynomial*,

$$T_0(z) = 1, \quad T_1(z) = z, \quad T_2(z) = 2z^2 - 1, \quad T_3(z) = 4z^3 - 3z, \quad \ldots,$$

and

$$T_n(z) = 2zT_{n-1}(z) - T_{n-2}(z) \quad \text{for} \quad n \geq 2.$$

The graph of the polynomial $T_n(z)$ on the segment $[-1, 1]$ resembles that of $\cos n\varphi$ on the segment $[-\pi, 0]$: all its maxima are equal to 1, and all its minima are equal to -1. Thus, $T_n(z)$ has $n-1$ critical points of order 2 (namely, $z = \cos\frac{k\pi}{n}$), but (for $n \geq 3$) *only two critical values*: $w = \pm 1$.

Chebyshev polynomials constitute a very important class of polynomials rich with various structures and connections; see, for example, [32]. They will play an important role in this chapter. Besides them, there are many other polynomials haveing only two critical values. Our main goal in this paper is to study the class of such polynomials.

Definition 2.2.3 We call a polynomial $P(z)$ a *Shabat polynomial* if it has at most two critical values.

In other words, for any such polynomial there exist two distinct complex numbers c_0, c_1 such that

$$P'(z) = 0 \quad \Rightarrow \quad P(z) \in \{c_0, c_1\}.$$

(For $P(z) = z^n$ one of the values (say, c_0) should be taken equal to $c_0 = 0$, and the other one may be chosen arbitrarily; by convention, we take $c_1 = 1$.)

Remark 2.2.4 These polynomials were introduced in [37]. One may also find in the literature the following terminology: *generalized Chebyshev polynomials* (see [38]) and *Chebyshev–Grothendieck–Belyi polynomials*, or, in abbreviated form, *ChGB-polynomials* (see [34]).

2.2.2 A tree as the preimage of a segment

Take two arbitrary complex numbers c_0, c_1 on the w-plane, and join them by a straight line segment. In what follows we will denote this segment by $[c_0, c_1]$, even in the case when one or both of its endpoints do not belong to $\mathbf{R}$, and also when they are both real but $c_1 < c_0$. Suppose there are no critical values on this segment; then its preimage $P^{-1}([c_0, c_1])$ is a disjoint union of n separate sets on z-plane, each one being homeomorphic to a segment though not necessarily straight (see Figure 2.2). We color the ends of $[c_0, c_1]$ black and white, in order to distinguish among their preimages on the z-plane.

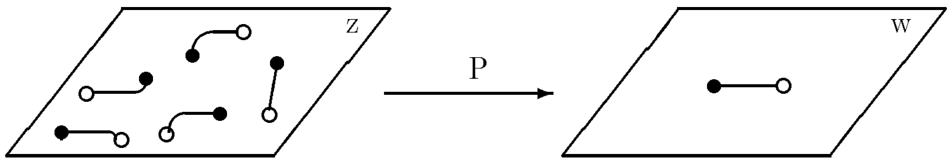


Figure 2.2: Preimage of a segment

Now suppose there are still no critical values *inside* the segment $[c_0, c_1]$, but one or both of its ends become critical. This leads to the fact that on the z-plane some "curvilinear segments" are glued one to another (see Figure 2.3).

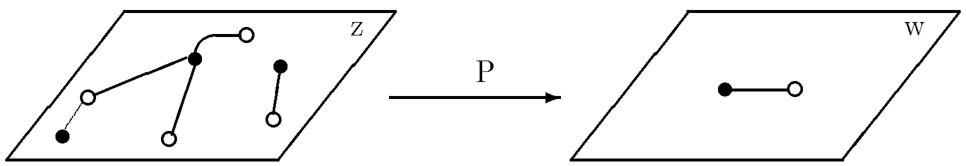


Figure 2.3: Ends of some segments become critical points

Now, what happens if the polynomial $P(z)$ in question is a Shabat polynomial, and c_1, c_2 are its (only) critical values? The answer is that we will see on the z-plane a *bicolored plane tree.*

Example 2.2.5 For $P(z) = z^n$ the preimage of the segment $[c_0, c_1] = [0, 1]$ is a "star-tree"; for $P(z) = T_n(z)$, a Chebyshev polynomial, the preimage of the segment $[c_0, c_1] = [-1, 1]$ is a "chain-tree" (see Figure 2.4).

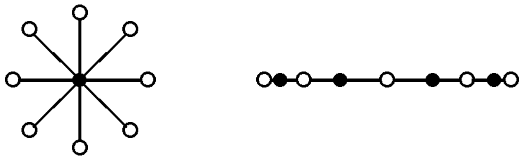


Figure 2.4: Trees that correspond to z^8 and to $T_8(z)$

It is obvious from the definition that affine transformations $z \mapsto az + b$ of the z-plane and $w \mapsto Aw + B$ of the w-plane do not destroy the property of a polynomial to be of "Shabat" type. This fact is reflected in the following definition.

Definition 2.2.6 Let $P(z)$ and $Q(z)$ be two Shabat polynomials, and let c_0, c_1 and d_0, d_1 be their critical values. We call the two *pairs* $(P, [c_0, c_1])$ and $(Q, [d_0, d_1])$ *equivalent*, if there exist constants $A, B, a, b \in \mathbf{C}$, $A, a \neq 0$, such that

$$Q(z) = AP(az + b) + B, \quad \text{and} \quad d_0 = Ac_0 + B, \quad d_1 = Ac_1 + B.$$

By abuse of language, instead of saying that the pairs $(P, [c_0, c_1])$ and $(Q, [d_0, d_1])$ are equivalent, we shall often say that the polynomials P and Q are equivalent.

The following is the main theorem of the theory of Shabat polynomials.

Theorem 2.2.7 *There is a bijection between the set of* (combinatorial bicolored plane) *trees and the set of equivalence classes of Shabat polynomials.*

It is very difficult to attribute this theorem to a specific author. We have learned it from Shabat, though it is not explicitly stated in his early papers, and Shabat himself considered this proposition as a particular case of the Belyi theorem. But this part of the Belyi theorem follows immediately from the Riemann existence theorem. Maybe the most proper attribution of the theorem would be to ascribe it to Riemann. But don't take it too seriously: don't look for the theorem in Riemann's works.

The theorem asserts three things:

(1) The preimage of the segment connecting critical values under the action of a Shabat polynomial is a tree (i.e., a connected graph without cycles).

(2) For any combinatorial bicolored plane tree there exists a Shabat polynomial that produces a tree possessing the same combinatorial structure.

(3) This polynomial is unique, up to the equivalence relation introduced in Definition 2.2.6.

We postpone the proof of statements (2) and (3) until Section 2.5, because it is essentially the same for the general case of arbitrary maps of arbitrary genus. Here we prove the first statement.

Proof of the statement (1) Our previous explanations amount to saying that the preimage of the segment connecting two critical values of a Shabat polynomial is a *graph*. Its black vertices are solutions of the equation $P(z) = c_0$, and its white vertices are solutions of $P(z) = c_1$. Vertices of valency greater than 1 are critical points of order equal to the valency. Vertices of valency 1 are ordinary (non-critical) points, i.e. non-multiple roots of the equations $P(z) = c_0$ and $P(z) = c_1$.

The first question to answer is why this graph has no circuits. Suppose it has one (see Figure 2.5).

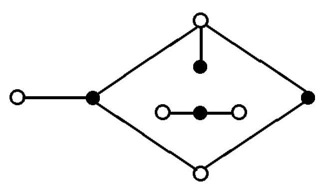


Figure 2.5: A graph with a circuit

Using, if necessary, an affine transformation of the w-plane, assume the points c_0 and c_1 to be real. Then our polynomial takes only real values on the boundary of the domain bounded by our circuit. This means that its imaginary part $\mathrm{Im}P(z)$, while being a harmonic function of variables x, y, is identically equal to zero on the boundary. Therefore, it is identically equal to zero inside, which is impossible.

A graph without circuits is either a tree, or a forest, i.e. a disconnected union of several trees. Let us prove that it is in fact connected. To do that, let us count the number of its vertices. Two equations $P(z) = c_0$ and $P(z) = c_1$ have in total $2n$ roots (with multiplicity taken into account). But we need to know their number *without multiplicities*. The multiplicity of a root is equal to the valency $\mathrm{Val}(v)$ of the corresponding vertex v. Thus

$$\#(\text{vertices}) = 2n - \sum_v (\mathrm{Val}(v) - 1),$$

where the sum is taken over all the vertices. Recall that a vertex of valency k is also a root of $P'(z)$, its multiplicity being equal to $k-1$. The number of roots of $P'(z)$ (with multiplicity taken into account) is equal to $n-1$. Hence,

$$\sum_v (\mathrm{Val}(v) - 1) = n - 1,$$

and the number of vertices is equal to $2n - (n-1) = n+1$. $\square$

A reader better familiar with topology may also apply a more straightforward topological argument: The preimage of the complement to $[c_0, c_1]$ in the *Riemann sphere* is a disc: indeed, P maps it to this complement with just infinity as a ramification point of the maximal order. Hence, the preimage of the segment itself cannot have either cycles, or more than one component.

Maybe one of the most surprising consequences of Theorem 2.2.7 is the following fact: *Every plane tree has a unique and canonical geometric form.*

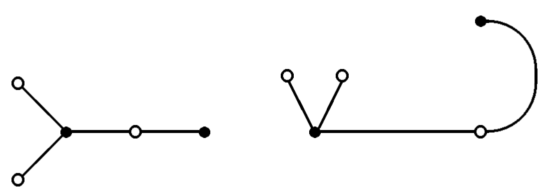


Figure 2.6: Two "geometric forms" of the same combinatorial tree

Any tree can be drawn on the plane in infinitely many different ways (cf. Figure 2.6). But one of its geometric forms is distinguished. Indeed, the affine transformation $z \mapsto az + b$, may change the size of a tree and its position on the z-plane, but it does not change its geometric form. In catalog [7] the computer-made "true" geometric forms are given for all the plane trees up to 8 edges.

Below one of these canonical forms is shown. Let us repeat once more: this picture is obtained as a preimage of a segment via a Shabat polynomial mapping. The picture may be enlarged or diminished, and it also may be rotated: all that amounts to an affine transformation of the z-plane. But all the ratios of lengths of different edges, all angles and curvatures are uniquely determined, and all this geometric information is somehow encoded in the combinatorial structure of the tree.

Figure 2.7

Figure 2.7: "True form" of a tree

2.2.3 Examples

At this stage it is very useful to see how actually to compute Shabat polynomials. We will show in this section several examples of various levels of difficlty. It goes without saying that this section is not the only one containing examples: much more will be scattered all along the text.

A very simple example

The first tree which is not a star and not a chain is the tree with the set of valencies $<3,1;2,1,1>$ (see Figure 2.8).

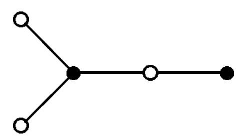


Figure 2.8: A simple example

Using an affine transformation of the w-plane, we may fix the critical values c_0 and c_1 at our own convenience. Let us take $c_0 = 0$, so that the coordinates of the black vertices are preimages of zero, i.e., the roots of the polynomial we are looking for, the vertex of valency 3 being the root of multiplicity 3.

Using an affine transformation of the z-plane, we may fix the positions of two arbitrary points of the tree. Let us place the vertex of valency 3 at the point $z = 0$, and the other black vertex, at the point $z = 1$. Then we may take

$$P(z) = z^3(z-1),$$

and we are finished: this polynomial is a Shabat polynomial of our tree.

If we are also interested in computing the coordinates of the other (white) vertices, we may proceed as follows. First, compute the derivative

$$P'(z) = z^2(4z-3).$$

The roots of the derivative are all critical points, i.e., all the vertices of valency strictly greater than 1, and the multiplicity of a root of $P'(z)$ is equal to the vetrex valency minus one. In our case the derivative has a double root at $z = 0$, and one more root $z = \frac{3}{4}$; it is exactly the position of the white vertex of valency 2.

Now we may compute the other critical value of $P(z)$: namely,

$$c_1 = P\left(\frac{3}{4}\right) = -\frac{27}{256}.$$

Finally, to get the coordinates of two other white vertices, we solve the equation

$$z^3(z-1) = -\frac{27}{256}.$$

It has a double root $z = \frac{3}{4}$, and two simple roots, $\frac{-1\pm\sqrt{-2}}{4}$.

Remark 2.2.8 Looking at the coordinate $\frac{3}{4}$ of the white vertex and at the negative value of c_1, we may conclude that it would have been more clever to take from the very beginning the position of the black vertex of valency 1 at $z = 4$, and to change the sign in front of the polynomial, thus having $P(z) = z^3(4 - z)$. Then the white vertex of valency 2 finds itself at $z = 3$, the second critical value is $c_1 = P(3) = 27$, and the coordinates of the other two white vertices become equal to $-1 \pm \sqrt{-2}$.

A new phenomenon: conjugate trees

Consider the following plane tree of type $<3, 2, 2; 2, 2, 1, 1, 1>$, see Figure 2.9.

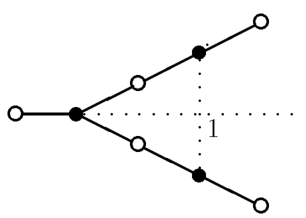


Figure 2.9: Another tree

Take, as in previous example, $c_0 = 0$. Let us place the black vertex of valency 3 at the point $z = 0$. Place the two black vertices of valency 2 so that the midpoint of the segment joining them is $z = 1$, or, in other words, the sum of the corresponding complex numbers is equal to 2. Then these two points are the roots of a quadratic polynomial $z^2 - 2z + a$, with a yet unknown value of the parameter a. Thus, the corresponding Shabat polynomial is equal to

$$P(z) = z^3(z^2 - 2z + a)^2.$$

What are the additional conditions that enable us to find a? To answer the question, let us compute the derivative:

$$P'(z) = z^2(z^2 - 2z + a)(7z^2 - 10z + 3a).$$

The derivative has (as it should have) a double root at $z = 0$, two simple roots, those of $z^2 - 2z + a$, and two more simple roots, namely the roots of the polynomial

$$Q(z) = 7z^2 - 10z + 3a.$$

Obviously, these latter roots are the coordinates of the two white vertices of valency 2. Thus, we have two conditions to insure:

(1) the roots of $Q(z)$ must be distinct, hence its discriminant $21a - 25 \neq 0$;

(2) (more important condition) *the values of* $P(z)$ *at the roots of* $Q(z)$ *must be equal to each other.*

Our first intention is to compute both roots of $Q(z)$ and to substitute them into $P(z)$. But this way is far too cumbersome. The following one is better: let us divide P by Q,

$$P(z) = S(z)Q(z) + R(z),$$

where $R(z)$ is the remainder, and, hence, is linear (since Q is quadratic): $R(z) = Az + B$, with coefficients A and B depending on the parameter a. Whenever $Q = 0$, the values of P coincide with those of R, hence the values of $R(z)$ must be equal to each other at the roots of $Q(z)$. This means that we must have $A = 0$.

Now, to actually divide $P(z)$ by $Q(z)$, it would be a good idea to use a computer algebra system, such as MAPLE. The results of the computation are as follows:

$$A = -\frac{16}{7^6}(21a - 25)(49a^2 - 476a + 400),$$

$$B = -\frac{196}{7^6}a(28a - 25)(7a - 10).$$

Therefore the conditions $A = 0$ and $21a - 25 \neq 0$ lead to the equation

$$S(a) = 49a^2 - 476a + 400 = 0,$$

hence

$$a = \frac{1}{7}(34 \pm 6\sqrt{21}),$$

while the second critical value is equal to $c_1 = B = B(a)$.

A priori it is not obvious which one of the two values of a we must take, so let us take both. Then, besides the tree in Figure 2.9, we also obtain the one of Figure 2.10.

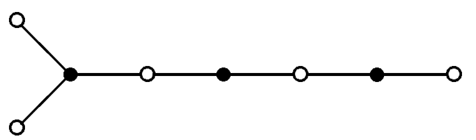


Figure 2.10: The tree conjugate to that of Figure 2.9

What do these two trees have in common? The answer is, they have the same set of black-and-white valencies: $<3, 2, 2; 2, 2, 1, 1, 1>$. If we look more closely at the process of computations, we will see that the set of valencies is the only information we used about the trees in question.

Proceeding from the above example, we would like to introduce a few terms.

- The number field $\mathbb{Q}(\sqrt{21})$ is the *field of definition* of the two trees.
- The polynomial $S(a) = 49a^2 - 476a + 400$, whose roots generate the field of definition, is called a *defining polynomial*.
- The trees themselves are called *conjugate* (to allude to the fact that the two values of the parameter a are algebraically conjugate).

- The set that consists of these two trees is a *Galois orbit*, i.e. an orbit under the action of the absolute Galois group $\Gamma = \mathrm{Gal}(\overline{\mathbb{Q}} \mid \mathbb{Q})$ on the set of bicolored plane trees.

- The set of all the trees having the same set of black-and-white valencies is called a *family* of trees of the same *type*. (In the present example the family coincides with the Galois orbit, which is not always the case.)

More precise meaning of each notion will be explained in Section 2.3. The most important, but rather difficult to grasp, is the notion of the field of definition. To introduce it properly, we need some knowledge of Galois theory. However, the computational part of it is not so conceptually difficult. For some time we may amuse ourselves computing fields of definition, without completely understanding what we are talking about. The following quasi-definition would do for the time being:

Definition 2.2.9 The *field of definition* of a bicolored plane tree is the simplest number field containing the coefficients of its Shabat polynomial.

The problem with this definiton is that the Shabat polynomial of a given tree is not unique. Therefore, the questions of exitence and unicity arise. They will be addressed in the next section.

A cubic orbit

In the same vein, we may consider a *cubic orbit*, i.e., a set of three conjugate trees with a cubic field of definition. Let us take the family of trees of the type $<3,2,1;2,2,1,1>$ (see Figure 2.11).

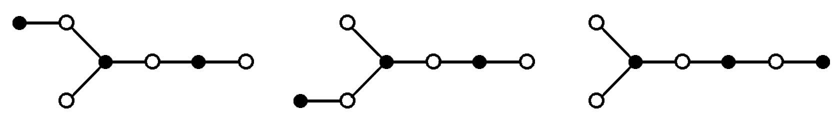


Figure 2.11: A cubic orbit

Place the black vertex of valency 3 at $z=0$, the black vertex of valency 2 at $z=1$, and the black vertex of valency 1 at $z=a$. Then we get a Shabat polynomial of the form

$$P(z) = z^3(z-1)^2(z-a),$$

whose derivative is equal to

$$P'(z) = z^2(z-1)Q(z),$$

where

$$Q(z) = 6z^2 - (5a+4)z + 3a.$$

The two white vertices of valency 2 are the roots of $Q(z)$. They must be different, hence the discriminant of $Q(z)$ must be non-zero:

$$25a^2 - 32a + 16 \neq 0.$$

The remainder after division of $P(z)$ by $Q(z)$ is $Az + B$, where

$$A = -\frac{1}{6^5}(25a^2 - 32a + 16)(25a^3 - 12a^2 - 24a - 16),$$

$$B = \frac{1}{2^5 3^4} a(5a-8)(25a^3 - 6a^2 + 8).$$

Hence the condition $A = 0$ (together with the condition on the discriminant of $Q(z)$) gives us a cubic equation

$$S(a) = 25a^3 - 12a^2 - 24a - 16 = 0.$$

The roots of this cubic equation generate a cubic number field, which is the field of definition of this set of trees.

By the way, the equation $S(a) = 0$ can be solved explicitly:

$$a_i = \frac{4}{25} + \frac{6}{25}\varepsilon_i(3\sqrt[3]{2} + \sqrt[3]{4}), \quad i = 1, 2, 3,$$

where ε_i are the cubic roots of 1. This gives a more explicit presentation of the field of definition: it is $\mathbb{Q}(\sqrt[3]{2})$ (we mean by this notation the field generated by *all the three* cubic roots of 2).

Exercise 2.2.10 Compute a Shabat polynomial for the tree of the type $< 2, 1, 1, \ldots, 1; m, n, 1, 1, \ldots, 1 >$ (see Figure 2.12). What is the position of the black vertex of valency 2?

Figure 2.12

Figure 2.12: A two-branch tree

Exercise 2.2.11 Compute a Shabat polynomial for the tree of the type $<m, 1, 1, \ldots, 1; n, 1, 1, \ldots, 1>$ (see Figure 2.13).
Hint: Compute the derivative $P'(z)$ first.

Figure 2.13

Figure 2.13: A tree of diameter 3

Exercise 2.2.12 Verify that the polynomial

$$P(z) = ((z^3 - 1)^3 + 1)^3$$

is of Shabat type. Draw the corresponding tree. Find its set of valencies.

2.2.4 Computational generalities and shortcuts

The algorithmic part of dealing with Shabat polynomials and trees is very rich. In this section we provide only a general discussion and number of recipes which would be useful for the reader if he or she would like to make computations.

A general case

Let the type

$$<\alpha, \beta> = <\alpha_1, \alpha_2, \ldots, \alpha_p; \beta_1, \beta_2, \ldots, \beta_q>$$

of a tree be given. This means that there are p black vertices (resp., q white vertices) of velencies $\alpha_1, \ldots, \alpha_p$ (resp., $\beta_1, \ldots, \beta_q$). Here $\alpha = <\alpha_1, \ldots, \alpha_p>$ and $\beta = <\beta_1, \ldots, \beta_q>$ are partitions of the number n of edges (because every edge has one black and one white end), and the numbers p and q satisfy the relation

$$p + q = n + 1,$$

which means that the total number of vertices is $n + 1$.

Exercise 2.2.13 Prove that for any pair of partitions $\alpha \vdash n$, $\beta \vdash n$ satisfying $p + q = n + 1$ there exists at least one bicolored plane tree of type $<\alpha, \beta>$.

Let us fix $c_0 = 0, c_1 = 1$. Denote $a_1, a_2, \ldots, a_p$ and $b_1, b_2, \ldots, b_q$ coordinates of the black and white vertices respectively. Then we have

$$P(z) = \lambda(z - a_1)^{\alpha_1}(z - a_2)^{\alpha_2} \ldots (z - a_p)^{\alpha_p}$$

and

$$P(z) - 1 = \lambda(z - b_1)^{\beta_1}(z - b_2)^{\beta_2} \ldots (z - b_q)^{\beta_q}.$$

The equalities between the coefficients provide us with n algebraic equations in $n + 2$ unknowns $\lambda, a_1, \ldots, a_p, b_1, \ldots, b_q$. Two additional "degrees of freedom" correspond to the possibility of making a linear transformation of the z-plane. We may use it in the way we find convenient. We may, for example, fix the positions of any two vertices, or make some other choice that places the tree in in an unambiguous position on the complex plane. In addition to the above set of equations, we must impose the condition that all the coordinates a_i, b_j are distinct.

2.3 Galois group action on plane trees

2.3.1 The field of definition of a plane tree

2.3.2 Combinatorial invariants of the Galois group action

2.3.3 A Shabat polynomial realizing a tree over the field of definition

2.3.4 More about fields of definition

2.4 Belyi functions: graphs on the sphere

In this section we pass from plane trees to arbitrary plane maps, and from polynomials to rational finctions. We cannot ignore the infinity anymore and must now work on the complex sphere rather than on the complex plane. A *Belyi function* is a rational function with at most three critical values, which are conventionally fixed at 0, 1 and ∞. A map is, as before, obtained as the preimage of a segment. Compared to the tree case, this more general setting has three new features: (1) There is a new wealth of examples; (2) The theory becomes computationally more difficult; (3) A new phenomenon appears: the coefficients of a Belyi function may not belong to the field of definition of the corresponding map. This makes the Galois action on maps much more intricate than in the case of trees.

2.4.1 Critical points of rational functions

Let $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the Riemann complex sphere, i.e., the complex plane $\mathbb{C}$ compactified by adding to it a point at ∞. Let us say a few words concerning this notion for a reader unfamiliar with it. On $\mathbb{C}$ every point is represented by its coordinate z. On $\overline{\mathbb{C}}$ there are two coordinates, z and s. But they do not represent all the points of the sphere: the coordinate z works everywhere outside infinity, and the coordinate s works everywhere outside zero. In between they agree one to another according to the formula $s = 1/z$. To say "$z = \infty$" is just a manner of speaking, when one really means to say "$s = 0$" (in the same way $s = \infty$ means $z = 0$). Any rational function

$$f(z) = \frac{P(z)}{Q(z)}$$

may be extended to a function from $\overline{\mathbb{C}}$ to $\overline{\mathbb{C}}$ (we will denote it by the same letter f) in an obvious way: if $Q(z) = 0$, then $f(z) = \infty$; if $\deg P < \deg Q$, then $f(\infty) = 0$, etc. The reader has noticed, of course, that the second complex plane, that of the variable w, had also been transformed to the sphere, so it now possesses two coordinates, w and t, and $t = 1/w$.

Exercise 2.4.1 Explain the meaning of $f(\infty)$, when $\deg P = \deg Q$; when $\deg P > \deg Q$.

For a function f thus defined it makes sense to speak about its derivative at any point $z \in \overline{\mathbb{C}}$. Only, one must choose correct variables. Namely:

- When z and w are both finite, there is no problem, we may take the ordinary derivative $f'(z)$.
- When $z = \infty$ and w is finite, we must substitute $1/s$ for z, differentiate $f(1/s)$ (don't forget to use the chain rule!), and then substitute $s = 0$.
- When z is finite and $w = \infty$, we must differentiate t as a function of z, i.e., $t = 1/w = 1/f(z)$.
- Finally, when both z and w are infinite, we must differentiate the function $t = 1/f(1/s)$.

Exercise 2.4.2 Carry out all the above mentioned differentiations.

The *degree* $\deg f$ of a rational function $f = P/Q$ is defined as

$$d = \deg f := \max\{\deg P, \deg Q\}.$$

This definition is coherent with the fact that the equation

$$f(z) = w,$$

considered as an equation on z for a given w, is a polynomial equation of degree d, namely,

$$P(z) = wQ(z), \quad \text{or} \quad P(z) - wQ(z) = 0.$$

Exercise 2.4.3 Explain the meaning of the equation $f(z) = w$ when one or both of the variables z and w are infinite.

Thus, the preimage of a point $w \in \overline{\mathbb{C}}$ usually consists of d points. But there are exceptions.

Definition 2.4.4 A multiple root of the equation $f(z) = w$ is called a *critical point* of the function f. Its multiplicity is called the *order* of the critical point. The value of the function f at a critical point z is called a *critical value* of f.

Exercise 2.4.5 Show that a point $z \in \overline{\mathbb{C}}$ is critical point of f if and only if $f'(z) = 0$, where the derivative is understood as is explained above. The order of a critical point z is equal to k, $k \geq 2$, if and only if $f'(z) = 0$, $f''(z) = 0$, ..., $f^{(k-1)}(z) = 0$, and $f^{(k)}(z) \neq 0$.

Remark 2.4.6 If we take into account the multiplicities, then the number of points in the preimage of a critical value is also equal to d. In fact, to any $w \in \overline{\mathbb{C}}$ there corresponds a prtition $d_1, d_2, \ldots, d_k$ of the number d, where d_i, $i = 1, \ldots, k$ are the multiplicities of the roots of the equation $f(z) = w$. For a non critical w this partition is $(1, 1, \ldots, 1)$, while for a critical value of w at least one of d_i is greater than 1.

There is a very convenient though somewhat old-fashioned terminology: we call an expression

$$A(z) = a_0 + a_1 z + \ldots + a_n z^n, \quad a_i \in \mathbb{C}, \quad i = 1, \ldots, n$$

a polynomial of *degree* n even if its leading coefficient a_n does *not* necessarily satisfy the condition $a_n \neq 0$. If it does, then n is called the *exact degree* of A. Then for a polynomial of degree n and of exact degree m we consider as its roots (1) all its finite roots, with their multiplicities; (2) the point ∞, with multiplicity $n - m$.

For a rational function $f = P/Q$ of degree $\deg f = d$ one of the polynomials P and Q is of exact degree d, while the exact degree of the other one may be less than d or equal to d. Let us consider both of them as polynomials of degree d. Then $f^{-1}(0)$ *consists of all the roots of* P *(with their multiplicities), and* $f^{-1}(\infty)$ *consists of all the roots of* Q *(also with their multiplicities).*

Now comes the main definition of the section.

Definition 2.4.7 A rational function $f : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ is called *Belyi function* if its only possible critical values are 0, 1 and ∞.

We have already explained how to find the preimages $f^{-1}(0)$ and $f^{-1}(\infty)$. In order to find $f^{-1}(1)$ we must take the rational function

$$f(z) - 1 = \frac{P(z) - Q(z)}{Q(z)}$$

and find the roots of its numerator $P - Q$, considering this polynomial as a polynomial of degree d.

We will use the following notation:

- $f^{-1}(0)$ consists of points $v_1, \ldots, v_p$ with multiplicities $\alpha_1, \ldots, \alpha_p$;
- $f^{-1}(\infty)$ consists of pointd $c_1, \ldots, c_q$ with multiplicities $\beta_1, \ldots, \beta_q$;
- $f^{-1}(1)$ consists of points $m_1, \ldots, m_n$ with multiplicities $\gamma_1, \ldots, \gamma_n$.

All the three tuples $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_q)$ and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ are partitions of the number d:

$$\sum_{i=1}^{p} \alpha_i = \sum_{j=1}^{q} \beta_j = \sum_{k=1}^{n} \gamma_k = d.$$

For a Belyi function f, we call the triplet $< \alpha, \beta, \gamma >$ of partitions of the number $d = \deg f$ the *set of valencies* corresponding to f (it will soon become clear why we have chosen the word "valency"). Sometimes the word *passport* is used to denote the same object (see [24]).

Exercise 2.4.8 Prove that for a Belyi function of degree d the total number of *distinct* points in the preimage of the set $\{0, 1, \infty\}$ is equal to

$$p + q + n = d + 2.$$

Hint: Use the derivative. Don't forget to consider finite and infinite points separately.

2.4.2 A map as the inverse image of a segment

The class of the affine transformations $z \mapsto az+b$, $a \neq 0$ of the complex plane $\mathbb{C}$ is the class of its complex *automorphisms*, where an automorphism of $\mathbb{C}$ is understood as a function $f : \mathbb{C} \to \mathbb{C}$ which is at the same time bijective and complex analytic everywhere. In the same way, the class of automorphisms of the Riemann sphere $\overline{\mathbb{C}}$ consists of all the fractional linear transformations

$$z \mapsto \frac{az+b}{cz+d}, \quad ad - bc \neq 0$$

(see, for example, [31]). A fractional linear transformation does not change if we multiply its numerator and denominator by the same non-zero constant. To get rid of this ambiguity, sometimes the condition $ad - bc \neq 0$ is replaced by $ad - bc = 1$, though computationally it is not always convenient.

An affine transformation can move any two given points $z_1, z_2 \in \mathbb{C}$ to any two given positions $z_1', z_2' \in \mathbb{C}$, and it is uniquely detemined once z_1, z_2 and z_1', z_2' are given. In the same way, a fractional linear transformation can move any *three* given points $z_1, z_2, z_3 \in \overline{\mathbb{C}}$ to any three given positions $z_1', z_2', z_3' \in \overline{\mathbb{C}}$, and it is uniquely determined once both triplets of points are given. In fact, the class of affine transformations of $\mathbb{C}$ may be considered as a proper subclass of fractional linear transformations of $\overline{\mathbb{C}}$, namely, of those of them that send ∞ to ∞.

Exercise 2.4.9 Prove that if $f(z)$ is a Belyi function, then

$$F(z) = f(\frac{az+b}{cz+d}), \quad ad - bc \neq 0$$

is also a Belyi function with the same set of valencies.

Definition 2.4.10 Two Belyi functions f_1 and f_2 are called *equivalent* if there exists a fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$, $ad - bc \neq 0$ such that

$$f_2(z) = f_1(\frac{az+b}{cz+d}).$$

The main theorem of our theory for the planar case is the following

Theorem 2.4.11 *There is a bijection between the equivalence classes of Belyi functions and the plane hypermaps.*

As in Section 2.2, we postpone the proof until later. Here we only explain the connection between Belyi functions and hypermaps.

On the sphere of the variable w we take the segment $[0,1]$. In order to distinguish visually the ends of this segment (and also their preimages) we mark the point 0 by $\bullet$, and the point 1, by $\times$. Thus, we get the following picture:

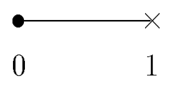


Figure 2.14: The "primary" hypermap

We may consider this segment as the primary hypermap: it has one vertex of valency 1, positioned at $w = 0$; one half-edge (the segment itself), with its "middle point" positioned at $w = 1$; and one face of valency 1 whose center is positioned at $w = \infty$.

Now, let f be a Belyi function of degree d. The corresponding hypermap, which we denote by H_f, is constructed as follows:

1. The hypermap H_f itself is "drawn" on the sphere of the variable z as the preimage $f^{-1}([0,1])$. The interior of the segment $[0,1]$ does not contain critical values; hence, it is repeated d times on the z-sphere. Each of these d intervals serves as a *half-edge* of H_f.

2. The *vertices* of H_f are the points of $f^{-1}(0)$, i.e., the points $v_1, \ldots, v_p$; the vertex valencies are $\alpha_1, \ldots, \alpha_p$ respectively.

3. The *middle points* of edges are the points of $f^{-1}(1)$, i.e., the points $m_1, \ldots, m_n$; the edge valencies are $\gamma_1, \ldots, \gamma_n$ respectively.

4. The faces of H_f are the preimages of the open disc $\overline{\mathbb{C}} \setminus [0,1]$. The *centers* of faces are the points $c_1, \ldots, c_q$, and their multiplicities $\beta_1, \ldots, \beta_q$ are equal to the valencies of the corresponding faces.

The reader would easily recognize the geometric representation of hypermaps of Chapter 1.

Remark 2.4.12 It goes without saying that the convention about which points to call vertices, centers of faces and middle points of edges, is a matter of an rabitrary choice. In different publications different conventions are used; quite often, for example, it is the preimage of infinity which is considered as the set of vertices (see [37] and others). The conventions used here in this book seem, however, to be more natural, at least for a biginner. Also, when one tries to make a computer picture of a map, the first and the most natural intention is to divide the segment [0,1] into, say, k of equal parts in order to solve the equations $f(z) = i/k$, $i = 0, \ldots, k$. For the segment $[1, \infty]$ this operation of dividing it into k "equal" parts seems to be less immediate.

The canonical triangulation of the hypermap H_f is obtained as the preimage of the canonical triangulation of the primary hypermap. The latter consists of two triangles drawn on the w-sphere: their vertices are placed at 0, 1 and ∞; their edges are the segments $[0, 1]$, $[1, \infty]$ and $[\infty, 0]$; and the interiors of the triangles correspond to the upper and lower half-planes of the complex plane w.

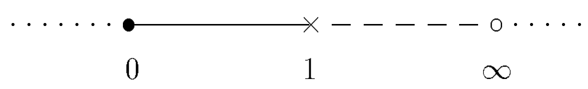


Figure 2.15: Primary triangulation

Note that our markings of vertices and edges of a canonical triangulation correspond to those of Chapter 1.

Remark 2.4.13 The equality $p+q+n = d+2$ of Exercise 2.4.8, obtained by counting the root multiplicities for a rational function, corresponds exactly to the Euler formula for hypermaps given in Chapter 1.

Now we may repeat once more what Theorem 2.4.11 actually sais. As in the case of Theorem 2.2.7, it essentially consists of three statements:

(1) The preimage of the segment $[0, 1]$ under a Belyi function mapping is always a plane hypermap.

(2) For any plane hypermap their exists a corresponding Belyi function.

(3) This function is unique, up to a fractional linear transformation of the z-sphere.

Let us remark that the liberty of a fractional linear transformation of the w-sphere, which also exists a priori, is eliminated by fixing once and for all the positions of all the three critical values. It would be the same for trees, had we fixed the critical values of a Shabat polynomial, say, at 0 and 1.

Definition 2.4.14 A Belyi function f is called *pure Belyi function* if all the points of $f^{-1}(1)$ are critical points of order 2:

$$d = 2n, \quad \text{and} \quad \gamma_1 = \gamma_2 = \ldots = \gamma_n = 2.$$

It is clear that *pure Belyi functions correspond to maps.*

Concerning the geometric form of a map or of a hypermap, it is not fixed to the same extent as it was in the case of trees. Indeed, a fractional linear transformations involve not only shifts and homotheties but also inversion. But in practice one usually (not always!) places the center of one of the faces at ∞; then, the "outer" face being fixed, the form of the (hyper)map becomes as rigid as it was for trees. Figure 2.16 below presents the "true form" of a map. The reader will better understand the details after reading the next section containing several examples.

Figure 2.16

Figure 2.16: The true form of a map

2.4.3 Examples

A very simple example

Let us consider the map with two edges, with two vertices of valencies 3 and 1 respectively, and with two faces, also of valencies 3 and 1, see Figure 2.17.

Figure 2.17

Figure 2.17: A map

Using the liberty of a fractional linear transformations of the variable z, we may choose the positions of any three points of our map at our own convenience. Let us put the vertex of valency 3 at $z = 0$, and the vertex of valency 1, at $z = 1$. This implies that the numerator of the corresponding Belyi function would, up to a constant factor, be equal to $z^3(1 - z)$.

2.5 Belyi functions: arbitrary genus

2.5.1 Riemann surfaces and Belyi pairs

2.5.2 Belyi theorem

2.5.3 Examples

2.6 More about Shabat polynomials and Belyi functions

2.6.1 Compositions

2.6.2 Julia sets

2.6.3 Davenport bound

2.6.4 Pell equation for polynomials

Bibliography

[1] **Adrianov N. M., Shabat G. B.** Plane trees and classical math. – Preprint, 1994, 13 pp.

[2] **Adrianov N. M., Zvonkin A. K.** Composition of plane trees. – Preprint, July 1994, 6 pp. Submitted to *Acta Applicandae Mathematicae.*

[3] **Bauer M., Itzykson C.** Triangulations. – In "Formal Power Series and Algebraic Combinatorics 93" (A. Barlotti, M. Delest, R. Pinzani eds.), Florence, June 1993, 1–45. Also in "The Grothendieck Theory of Dessins d'Enfant" (L. Schneps ed.), London Math. Soc. Lecture Notes series, vol. 200, Cambridge Univ. Press, 1994, 179–236.

[4] **Beazley Cohen P., Itzykson C., Wolfart J.** Fuchsian triangle groups and Grothendieck dessins: Variations on a theme of Belyi. – Preprint, 1993, 27 pp.

[5] **Beazley Cohen P., Wolfart J.** Dessins de Grothendieck et variétés de Shimura. – *C. R. Acad. Sci. Paris, sér. I*, 1992, t. **315**, 1025–1028.

[6] **Belyi G. V.** On Galois extensions of a maximal cyclotomic field. – *Math. USSR Izvestija*, 1980, vol. **14**, no. 2, 247–256. (Original Russian paper: *Izv. Acad. Nauk SSSR, cer. mat.*, 1979, vol. **43**, no. 2, 269–276.)

[7] **Bétréma J., Péré D., Zvonkin A.** Plane trees and their Shabat polynomials. Catalog. – Rapport interne du LaBRI, Bordeaux, 1992, 119 pp.

[8] **Bétréma J., Zvonkin A.** La vraie forme d'un arbre. – In: "TAPSOFT '93: Theory and Practice of Software Development. 4-th International

Joint Conference CAAP/FASE (Orsay, April 1993). Proceedings" (M.-C. Gaudel, J.-P. Jouannaud eds.), Lecture Notes in Computer Science, vol. 668, Springer, 1993, 599–612.

[9] **Bétréma J., Zvonkin A.** Plane trees and Shabat polynomials. – In "Formal Power Series and Algebraic Combinatorics 93" (A. Barlotti, M. Delest, R. Pinzani eds.), Florence, 1993, 83–95. (To appear in *Discrete Mathematics.*)

[10] **Birch B. J.** ?????

[11] **Birch B. J.** Shabat trees of diameter 4: Appendix to a paper of Zvonkin. – Preprint, 1995, 5 pp.

[12] **Cohen H.** A Course in Computational Algebraic Number Theory. – Springer (Graduate Texts in Mathematics, vol. 138), 1993.

[13] **Couveignes J.-M.** Calcul et rationnalité de fonctions de Belyi en genre 0. – *Ann. de l'Inst. Fourier*, 1994, vol. **44**, no. 1, ??–??.

[14] **Couveignes J.-M.** A propos du théorème de Belyi. – Prépublication, Mars 1995, 5 pp.

[15] **Couveignes J.-M., Granboulan L.** Dessins from a geometric point of view. – In "The Grothendieck Theory of Dessins d'Enfant" (L. Schneps, ed.), Cambridge Univ. Press, 1994 (see [33]), 79–113.

[16] **El Marraki M., Zvonkin A.** Composition des cartes et hypercartes. – *Séminaire Lotharingien de Combinatoire*, 1995, vol. **34**, 18 pp. (`http://cartan.u-strasbg.fr/~slc`)

[17] **Filimonenkov V. O., Shabat G. B.** Fields of definition of Belyi function and Galois cohomology. – Preprint, 1995, 14 pp. (in Russian)

[18] **Granboulan L.** Construction d'une extension régulière de $\mathbf{Q}(T)$ de groupe de Galois M_{24}. – Prépublication, Juillet 1994, 9 pp.

[19] **Grothendieck A.** Esquisse d'un programme. – Manuscrit non publié, 1984, 57 pp.

[20] **Jones G. A.** Maps on surfaces and Galois groups. – Preprint, August 1995, 25 pp. Submitted to *Mathematica SLOVACA*.

[21] **Jones G. A., Singerman D.** Maps, hypermaps and triangle groups. – Preprint, July 1993, 23 pp.

[22] **Jones G. A., Singerman D.** Belyi functions, hypermaps and Galois groups. – Preprint, December 1993, 25 pp.

[23] **Jones G. A., Streit M.** Galois groups and plane trees. – Preprint, 1994, 19 pp.

[24] **Khovanskii A. G., Zdravkovska S.** Branched covers of S^2 and braid groups. – Preprint, 1994, 21 pp.

[25] **Kochetkov Yu. Yu.** Cartographic groups and genaralized Chebyshev polynomials (Kartograficheskie gruppy i obobshchennye polinomy Chebysheva). – *Uspekhi Matem. Nauk*, 1994, t. **49**, no. 6, 203–204.

[26] **Kochetkov Yu. Yu.** Plane trees, cartographical groups and Chebyshev polynomials. – Preprint, 1995, 29 pp.

[27] **Kralina V., Shabat G.** Belyi pairs corresponding to centrally symmetric torus dessins. – Preprint, September 1993, 6 pp.

[28] **Matzat B. H.** Konstruktive Galoistheorie. – Lecture Notes in Math., vol. **1284**, 1987.

[29] **Pakovich F. B.** Elliptic polynomials. – Preprint, 1995, 14 pp. (in Russian)

[30] **Präve A.** Belyifunktionen und Weierstrasspunkte. – Diplomarbeit, January 1994, 24 pp.

[31] **Reyssat E.** Quelques Aspects des Surfaces de Riemann. – Birkhäuser, 19?? ("Progress in Mathematics", vol. **??**).

[32] **Rivlin T. J.** Chebyshev Polynomials. – Wiley, 1990 (2nd edition).

[33] **Schneps L. (ed.)** The Grothendieck Theory of Dessins d'Enfant. – London Math. Soc. Lecture Notes Series, vol. 200, Cambridge Univ. Press, 1994, 368 pp.

[34] **Shabat G. B.** Notes of the seminar on the "dessins d'enfant". Moscow, 1990, unpublished (in Russian), 70 pp.

[35] **Shabat G. B.** The arithmetic of 1-, 2- and 3-edged Grothendieck dessins. – Preprint IHES, November 1991, 16 pp.

[36] **Shabat G. B.** On the cartographical normalization of Grothendieck dessins. – Preprint, January 1993, 13 pp.

[37] **Shabat G. B., Voevodsky V. A.** Drawing curves over number fields. – In "The Grothendieck Festschrift" (P. Cartier, L. Illusie, N. M. Katz, G. Laumon, Y. Manin, K. A. Ribet eds.), Birkhäuser, 1990, vol. 3, 199–227.

[38] **Shabat G. B., Zvonkin A. K.** Plane trees and algebraic numbers. – In "Jerusalem Combinatorics '93" (H. Barcelo, G. Kalai eds.), AMS, Contemporary Mathematics series, vol. **178**, 1994, 233–275.

[39] **Singerman D., Watson P. D.** Weierstrass points on regular maps. – Preprint, August 1995, 22 pp.

[40] **Smit D.-J.** Summation over equilaterally triangulated surfaces and the critical string measure. – Preprint, August 1990, 45 pp.

[41] **Streit M.** Homology, Belyi functions and canonical curves. – Preprint, 1994, 8 pp.

[42] **Wolfart J.** Mirror-invariant triangulations of Riemann surfaces, triangle groups and Grothedieck dessins: Variations on a thema of Belyi. – Preprint, June 1992, 27 pp.

[43] **Wolfart J.** The 'obvious' part of Belyi's theorem and Riemann surfaces with many automorphisms. – Preprint, August 1995, 7 pp.

[44] **Zannier U.** On Davenport's bound for the degree of $f^3 - g^2$ and Riemann's Existence Theorem. – *Acta Arithmetica*, 1995, vol. **71**, no. 2, 107–137.

[45] **Zvonkin A. K.** How to draw a group. – In "Formal Power Series and Algebraic Combinatorics 95" (B. Leclerc, J.-Y. Thibon eds.), Paris, May 1995, 601–611. Submitted to *Discrete Mathematics*.

Chapter 3

Two–dimensional quantum gravity models

3.1 Introduction

Among all exciting new developments in the theory of embedded graphs one of the most interesting is the fact that physicists (working in two-dimensional quantum gravity) became interested in map enumeration. This is how very classic problems of the enumerative combinatorics became a hot point in theoretical physics. But, as is often the case with physicists, they developed their own, highly original techniques of map enumeration based on the computation of integrals over the space of Hermitian matrices. The foundation of the method were developed in [7],[23]. In physical literature they also cite [22] as a source of the main idea.

Unfortunately, our background in physics is insufficient to present a clear and complete picture of the exciting world of quantum gravity. We restrict ourselves with some general remarks.

In string theory a *particle* is presented as a *string*. A string is simply a (topological) circle. During its life time (starting with its birth and ending with its death) a particle interacts with other particles. It can collide with another particle or split into two particles. This conglomerate of interacting particles forms thus a two-dimensional surface in the space-time (see Figure 3.1). The total life time of the conglomerate is usually supposed to be finite, what means that the surface is compact.

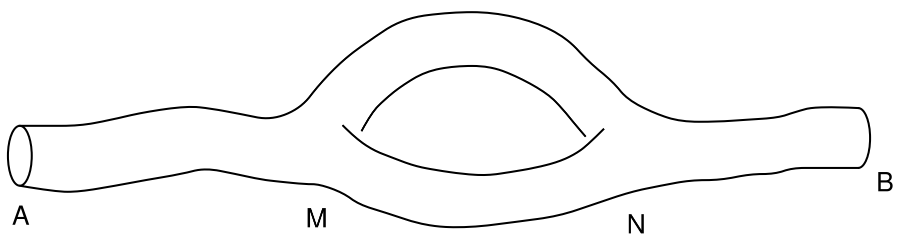


Figure 3.1: Evolution of a string in time

A "quantum particle" moves all in possible ways with the same probability. The "way of moving" corresponds to a metric on the surface. Thus, all possible (smooth) metrics have the same probability. A physical "observable", i. e., any quantity characterizing the particle, which can be measured, is thus understood as an integral over the space of all metrics on the surface.

However, integration over the space of all smooth metrics is a cumbersome procedure, if at all defined. A natural idea is thus to *discretize* the surface, and to substitute the integration over the space of metrics with summing over the discrete set of discretized surfaces.

Here is the very place where embedded graphs appear on the scene. The most natural way to discretize a surface is to substitute the surface with its triangulation. A triangulation of a surface determines a piecewise flat metric on the surface: we simply suppose each triangle to be a regular plane triangle of given area. Knowing the total area of the surface and the number of triangles one can easily calculate the area of a triangle, and therefore its side.

Let us fix the number of triangles in the triangulation. Suppose that all the triangulations are equally probable. Physicists usually suppose that "for the number of triangles large enough summation over the (finite) set of all triangulations is equivalent to integration over the space of all metrics". This statement is very difficult to formalize. However, since there is no consistent definition of integration over the space of metrics, summing over triangulations can be understood as such a definition: *a physical observable is the limit of the mean value of some quantity over the set of all triangulations of the given surface as the number of triangles tends to infinity.*

In fact, we are not able to restrict ourselves with only one surface since the topology of the life-time surface is unpredictable. One must therefore integrate over the space of all metrics over all surfaces. The last procedure can be substituted with summing over triangulations of surfaces of arbitrary genus.

A triangulation of a surface may be considered as a "generic" embedded graph: any graph close (in some natural sense) to a triangulation is itself a triangulation. In fact, one may introduce a cell decomposition of moduli spaces of curves associated with embedded graphs structure. In these cell decompositions triangulations naturally correspond to open cells, while other embedded graphs numerate cells of higher codimension. Thus, summing up over triangulations produces the most essential part of the integral.

In order to sum over embedded graphs we need a procedure to enumerate them. This problem proved to be closely related with integration over the Gaussian measure in the space of Hermitian matrices. The method of Gaussian integrals is based on the fact that the number of ways to glue a set of polygons (or, in the dual setting, a set of stars) into a surface of given genus may be represented as the integral of a polynomial in traces of Hermitian matrices over the Gaussian measure on the space of Hermitian matrices. A model problem solved with the help of this machinery is the following:

Considering a $2n$-gon, what is the number of ways to obtain an oriented surface of genus g by gluing the edges of this polygon pairwise?

Solving this problem Harrer and Zagier [20] rediscovered the method of Gaussian integrals invented earlier by physicists [4]. (In fact, the aim of their paper [20] was to calculate the virtual Euler characteristic of the moduli space of complex curves, and solving the gluing problem was the crucial technical point.) Although there exist different ways to obtain the same result without using integration over the space of Hermitian matrices, we present the original proof as a very instructive example of the general method.

Another example is given by the enumeration problem for gluings of a finite set of 4-stars. Describing a solution of this problem we present a general method of generalized Hermitian polynomials, applicable to calculation of any Gaussian integral. However, a direct application usually leads to very complicated computational problems, and the final results are obtained only in few cases. It is worth mentioning the enumeration of 4-stars gluings for small genera.

A variant of the Gaussian integral method suggested by Kazakov allows one to write down integrals enumerating more specific objects, like generic unicursal plane curves (the Gauss problem). However, these integrals are even more complicated and no good estimates are obtained up to now.

The Gaussian integral enumerating all possible gluings of any number of arbitrary stars forms the *universal one-matrix model* of the two-dimensional quantum gravity. Another model, suggested by Witten? describes the two-dimensional gravity as the intersection theory on the moduli space of complex curves. Each of the models can be considered as a generating function in infinite number of indeterminates. Witten [39] conjectured the equivalence of these two models (i.e., the coincidence of the two generating functions), and one of the most exciting achievements of the last few years was the Kontsevich's proof of this conjecture. The proof is based on constructing an intermediate model (the Kontsevich model) and proving the equivalence of this intermediate model to both models in question.

Since even a precise statement of Witten's conjecture would require a separate chapter, we restrict ourselves with describing the Kontsevich model and its connection with embedded graphs.

3.2 Model problem: one face maps

Consider a square. Gluing its edges pairwise we obtain a two-dimensional (oriented) surface. Recall that in order to obtain an oriented surface we glue edges with opposite directions (with respect to the orientation of the polygon). All three ways to glue the edges are presented on Figure 3.2.

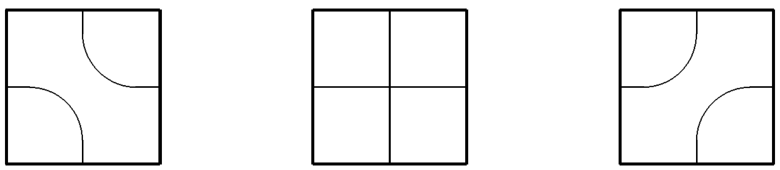


Figure 3.2: Three possible gluings of the square

Here we mark the edges to be glued together by an arc connecting these edges. Note that *we distinguish between a gluing and another gluing that can be obtained from the first one by a rotation of the square.*

The first and the third ways of gluing give the sphere as the result, see Figure 3.3. The second way gives the torus.

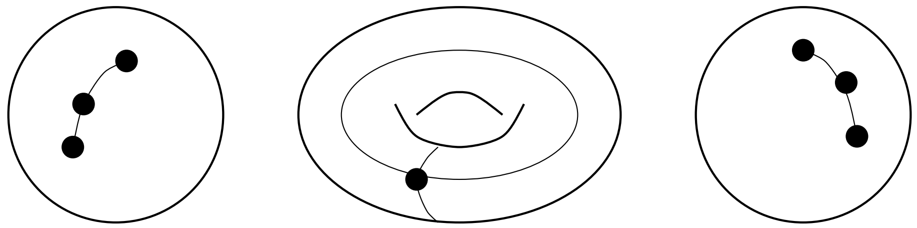


Figure 3.3: Surfaces glued from the square

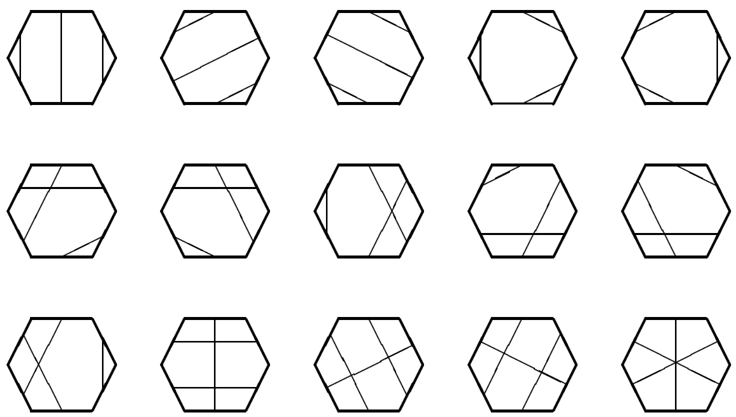


Figure 3.4: All possible ways to glue the edges of the hexagon

Similarly, there exist fifteen ways of gluing the hexagon, see Figure 3.4. The first five of them result in the sphere, while the last ten ways give the torus. This statement can be easily verified in the following way. Consider the boundary of the original polygon. The image of this boundary is an embedded graph on the resulting surface. We know the number of edges in this graph: for a polygon with $2n$ sides this number equals n since two sides of the polygon are glued together producing an edge of the graph.

The number of faces equals one. Therefore the genus of the surface is uniquely determined by the number of vertices of the graph.

This number can be easily computed looking on the picture of gluing. We can, for example, label all the vertices of the polygon that are glued together by the same label. Then the number of different labels will be the number of resulting vertices, see Figure 3.5.

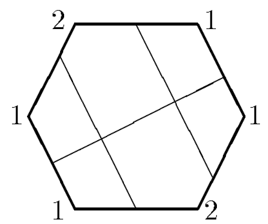


Figure 3.5: A labeling of glued vertices

For the gluing on the picture we conclude that the Euler characteristic is $2 - 2g = V - E + F = 2 - 3 + 1 = 0$ and the surface is the torus.

Exercise 3.2.1 Prove the equivalence of the following statements:

1. the resulting surface is the sphere;

2. the resulting graph is a tree;

3. chords, connecting pairs of gluing sides. do not intersect each other.

In Table 3.1 one can see the number of ways to obtain a surface of genus g from a regular $2n$-gon for small values of n. The table is borrowed from [20].

The sum of all numbers in a row is $(2n-1)!!$ since there exist precisely $(2n-1)!!$ ways to split $2n$ edges of the polygon into disjoint pairs.

The problem we consider now is

$n\backslash g$	0	1	2	3	4	5
1	1	0	0	0	0	0
2	2	1	0	0	0	0
3	5	10	0	0	0	0
4	14	70	21	0	0	0
5	42	420	483	0	0	0
6	132	2310	6468	1485	0	0
7	429	12012	66066	56628	0	0
8	1430	60060	570570	1169740	225225	0
9	4862	291720	4390386	17454580	12317877	0
10	16796	1385670	31039008	211083730	351683046	59520825

Table 3.1: The number of g-genus gluings of $2n$-gon

Problem 3.2.2 Enumerate the ways to obtain a surface of genus g from the $2n$-gon.

Note that in dual setting gluing a surface from a polygon with $2n$ sides is substituted with gluing a surface from a star with $2n$ needles. In this case the number of vertices is always 1, the number of edges equals n, and the number of faces depends on a gluing, thus determining the genus.

A reader familiar with enumerative combinatorics may note that the first column of the table consists of the famous Catalan numbers $\mathrm{Cat}_n = \frac{1}{n+1}\binom{2n}{n}$. The 0-th Catalan number is usually set to $\mathrm{Cat}_0 = 1$.

Exercise 3.2.3 Prove that the number of ways to glue the sphere from a polygon with $2n$ sides equals Cat_n.

Hint. Use the fact that there is a one-to-one correspondence between spherical gluings and rooted plane trees [21].

We will give an answer to the problem in the following form. Introduce the sequence of polynomials

$$T_1(N) = N^2$$

$$\begin{aligned} T_2(N) &= 2N^3 + N \\ T_3(N) &= 5N^4 + 10N^2 \\ T_4(N) &= 14N^5 + 70N^3 + 21N \\ T_5(N) &= 42N^6 + 420N^4 + 483N^2 \end{aligned}$$

and so on. The polynomials T_n are generating polynomials for the rows of the table: $T_n(N) = \sum_{\sigma} N^{V(\sigma)}$, where we sum over all possible ways of gluing σ and $V(\sigma)$ is the number of vertices in the resulting map.

Exercise 3.2.4 Prove that the polynomial T_n is odd for n even, and it is even for n odd.

Consider the (exponential) generating function for the sequence of polynomials $T_n(N)$:

$$\begin{aligned} T(N,s) &= 1 + 2Ns + 2s\sum_{n=1}^{\infty}\frac{T_n(N)}{(2n-1)!!}s^n \\ &= 1 + 2Ns + 2N^2s^2 + \frac{2}{3}(2N^3+N)s^3 + \frac{2}{15}(5N^4+10N^2)s^4 + \dots \end{aligned} \tag{3.1}$$

Theorem 3.2.5 $T(N,s) = \left(\frac{1+s}{1-s}\right)^N$.

Denote by $\varepsilon_g(n)$ the number of gluings of the $2n$-gon producing a surface of genus g, i. e., $T_n(s) = \varepsilon_0(n)N^{n+1} + \varepsilon_1(n)N^{n-1} + \dots$. Theorem 3.2.5 implies a convenient recursion formula for computing the values of $\varepsilon_g(n)$.

Corollary 3.2.6

$$\varepsilon_g(n) = \frac{4n-2}{n+1}\varepsilon_g(n-1) + \frac{(n-1)(2n-1)(2n-3)}{n+1}\varepsilon_{g-1}(n-2),$$

and

$$\varepsilon_g(0) = \begin{cases} 1, & \text{if } g = 0 \\ 0, & \text{otherwise} \end{cases}$$

There exist different ways to prove this theorem. We present below in Section 3.3.8 the original proof from [20]. This proof is based on an expression of the polynomials T_n through matrix Gaussian integrals.

3.3 Gaussian integrals

Important integrals coming from various applications often have nontrivial combinatorial meaning. In this section we study the combinatorial meaning of integrals over Gaussian measure. Even simple integrals of this kind on the real line have combinatorial interpretation. And integration over the Gaussian measure on the space of Hermitian matrices happen to be closely connected to enumeration of embedded graphs with respect to their genera. Although we concentrate mainly on the space of Hermitian matrices, corresponding to the orientable case, the reader must have in mind that there exist other important applications of Gaussian integrals. For example, integration over the space of real symmetric matrices enumerates graphs embedded in both orientable and non-orientable surfaces.

The standard reference on Gaussian measures is [16].

3.3.1 The Gaussian measure on the line

This is the measure μ on the real line $\mathbf{R}$ with the density

$$d\mu(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad x \in \mathbf{R}.$$

You may see the graph of the density in Figure 3.6.

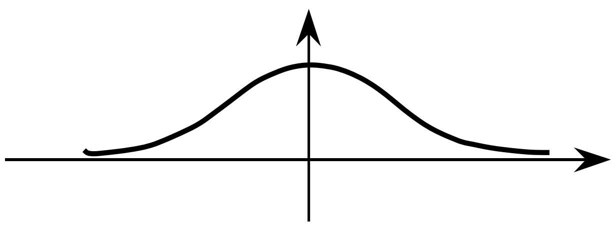


Figure 3.6: Standard Gaussian density on the line

This measure has the following properties:

1. It is *normalized*, that is, the integral of the density over the line is equal to 1:

$$\int_{\mathbf{R}} d\mu(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1.$$

2. The *mean* is equal to 0:

$$\int_{\mathbf{R}} x d\mu(x) = 0.$$

3. The *variance* is equal to 1:

$$\int_{\mathbf{R}} x^2 d\mu(x) = 1.$$

4. The *characteristic function* (or Fourier transform) is

$$\varphi(t) := \int_{\mathbf{R}} e^{itx} d\mu(x) = e^{-\frac{t^2}{2}}.$$

5. The integral

$$\int_{\mathbf{R}} p(x) e^{-\frac{x^2}{2}} dx$$

converges for any polynomial p.

Exercise 3.3.1 Prove these properties.

As an example we present a proof of the first property.

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx\right)^2 &= \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx dy \\
&= \int_{0}^{\infty} \int_{0}^{2\pi} r e^{-\frac{1}{2}r^2} d\varphi dr \\
&= 2\pi \int_{0}^{\infty} r e^{-\frac{1}{2}r^2} dr \\
&= 2\pi
\end{aligned}$$

Notation 3.3.2 We will regularly use the following notation borrowed from physics: for any measure μ on X, and for any function $f : X \to \mathbf{R}$ or

$f : X \to \mathbf{C}$ we will denote by $<f>$ the *mean*, or *average*, of f with respect to the measure μ:

$$<f> := \int_X f(x) d\mu(x).$$

The measure μ and the space X not explicitly mentioned in this notation will usually be clear from the context. For example, the above formulas may be rewritten as

$$<1> = 1, \quad <x> = 0, \quad <x^2> = 1, \quad \varphi(t) := <e^{itx}> = e^{-\frac{t^2}{2}}.$$

Exercise 3.3.3 Prove that

$$\int_{-\infty}^{\infty} x^{2n} d\mu(x) = (2n-1)!! \tag{3.2}$$

Hint. Taking the integral by parts twice prove the recursion

$$\int_{-\infty}^{\infty} x^{2n} d\mu(x) = (2n-1) \int_{-\infty}^{\infty} x^{2n-2} d\mu(x).$$

Note that $(2n-1)!!$ is precisely the number of ways to glue an oriented surface from the regular $2n$-gon. As we are going to see this coincidence is not accidental.

The *general Gaussian measure on the line* has the density

$$d\mu(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx, \quad x, m, \sigma \in \mathbf{R}, \quad \sigma > 0. \tag{3.3}$$

This measure is also normalized, and the mean, the variance, and the characteristic function are as follows:

$$<x> = m, \quad <(x-m)^2> = \sigma^2, \quad \varphi(t) := <e^{itx}> = e^{imt - \frac{\sigma^2 t^2}{2}}. \tag{3.4}$$

Remark 3.3.4 As is well known, a probabilistic distribution is uniquely determined by its characteristic function. Hence, we could give the last formula for $\varphi(t)$ as a *definition* of a Gaussian measure on the line.

3.3.2 Gaussian measure in $\mathbf{R}^k$

This measure will be the main object of our study. To introduce it we must start with its characteristic function.

Let $x = (x_1, x_2, \dots, x_k) \in \mathbf{R}^n$. By (x, y) we denote the ordinary scalar product in $\mathbf{R}^n$, that is, $(x, y) := x_1 y_1 + \dots + x_k y_k$.

Definition 3.3.5 A measure μ on $\mathbf{R}^k$ is called *Gaussian*, if its characteristic function has the form

$$\varphi(t) := \int_{\mathbf{R}^k} e^{i(t,x)} d\mu(x) = \exp\left\{ i(m, t) - \frac{1}{2}(Ct, t) \right\},$$

where $t, m \in \mathbf{R}^k$ are vectors, and $C = (c_{ij})$ is a non-negatively defined $(k \times k)$-matrix of a quadratic form. The vector m is the *mean vector* of the measure μ, that is, $< x_i > = m_i$, and the matrix C is its *covariance matrix*, that is, $<(x_i - m_i)(x_j - m_j)> = c_{ij}$.

Remark 3.3.6 In what follows, for the sake of simplicity we consider only the case of zero mean $m = 0$, thus

$$\varphi(t) = \exp\{-\frac{1}{2}(Ct, t)\}, \tag{3.5}$$

and

$$<x_i> = 0, \quad <x_i x_j> = c_{ij}, \quad i, j = 1, 2, \dots, k.$$

The covariance matrix C may well be degenerate. Then the measure has no density and is concentrated on a vector subspace of $\mathbf{R}^k$. When C is *non* degenerate, the measure has the *density*

$$d\mu(x) = \text{Const} \times \exp\left\{ -\frac{1}{2}(Bx, x) \right\} dv(x), \tag{3.6}$$

where the matrix $B = C^{-1}$ and $dv(x) = dx_1 \dots dx_k$ is the standard volume form. For this measure to be normalized, i.e., the measure of the whole space $\mathbf{R}^k$ to be equal to 1, we must take the constant factor in the last formula equal to

$$\text{Const} = (2\pi)^{-k/2} (\det B)^{1/2}. \tag{3.7}$$

Indeed, considering a linear orientation preserving orthogonal transformation $x = Oy$ in $\mathbf{R}^k$ we obtain

$$\begin{aligned}\int_{\mathbf{R}^k} e^{-(Bx,x)/2} dv(x) &= \int_{\mathbf{R}^k} e^{-(BOy,Oy)/2} dv(Oy) \\ &= \int_{\mathbf{R}^k} e^{-(O^{-1}BOy,y)/2} dv(y)\end{aligned}$$

since an orthogonal transformation does not change the volume form. Choosing O diagonalizing the matrix B we obtain a diagonal matrix $B_1 = O^{-1}BO$ with positive elements $b_1, \ldots, b_k$, the eigenvalues of B, on the diagonal. Now the integration can be done separately in each of the variables $y_1, \ldots, y_k$, and the result will be

$$\int_{\mathbf{R}^k} e^{-(Bx,x)/2} dv(x) = \prod_{i=1}^{k} \left(\frac{2\pi}{b_i}\right)^{1/2} = (2\pi)^{k/2} \left(\prod_{i=1}^{k} b_i\right)^{-1/2} = (2\pi)^{k/2} (\det B)^{-1/2},$$

i. e., precisely what we have expected.

All the above formulas (3.5), (3.6), (3.7) are consistent with the one-dimensional case (3.3), (3.4), where $C = (\sigma^2)$, $B = (\sigma^{-2})$, and Const $= (2\pi)^{-1/2}\sigma^{-1}$.

A Gaussian measure in $\mathbf{R}^n$ is called *standard* if both B and C are identity matrices.

3.3.3 Integrals of polynomials and Wick formula

A great part of the machinery of quantum physics consists of methods of computing the integrals with respect to a Gaussian measure (not necessarily in the finite dimensional case); see, for example, [35]. The goal of this section is to develop a technique for integration of polynomials.

Knowing that $< x_i > = 0$, $i = 1, \ldots, n$ and $< x_i x_j > = c_{ij}$, $i, j = 1, \ldots, n$, one can easily compute the integral of any polynomial in $x_1, \ldots, x_n$ of degree 2. What about higher degrees?

Lemma 3.3.7 *If $f(x)$ is a monomial of odd degree, then $< f > = 0$.*

Indeed, the integral of an odd form vanishes.

The following theorem reduces the integration of any polynomial to that of degree 2.

Theorem 3.3.8 (Wick formula) *Let $f_1, f_2, \ldots, f_{2n}$ be a set of linear functions* (not necessarily different) *of $x_1, \ldots, x_k$. Then*

$$< f_1 f_2 \ldots f_{2n} >= \sum < f_{p_1} f_{q_1} > < f_{p_2} f_{q_2} > \ \ldots \ < f_{p_n} f_{q_n} >,$$

where the sum is taken over all permutations $p_1 q_1 p_2 \ldots q_n$ of the set of indices $1, 2, \ldots, 2n$ such that $p_1 < p_2 < \ldots < p_n$, $p_1 < q_1, \ldots, p_n < q_n$.

The number of summands in the right hand side is equal to $(2n-1)!!$ A partition of the set $1, 2, \ldots, 2n$ into couples (p_i, q_i) satisfying the conditions of the Wick formula is called a *Wick coupling.*

Example 3.3.9 As an application of the Wick formula let us compute the one-dimensional integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} dx = < x^4 > .$$

We have $x^4 = f_1 f_2 f_3 f_4$, where $f_1 = f_2 = f_3 = f_4 = x$. Hence,

$$< f_1 f_2 f_3 f_4 > = < f_1 f_2 > < f_3 f_4 > \ + \ < f_1 f_3 > < f_2 f_4 > \ + \ < f_1 f_4 > < f_2 f_3 > .$$

Now each $< f_i f_j > = < x^2 > = 1$; therefore the result is $1^2 + 1^2 + 1^2 = 3$ in consistency with our earlier calculations.

In the same way we may find $< x^{2n} > = (2n-1)!!$ (sf. Exercise 3.3.3).

3.3.4 The Wick formula: a sketch of the proof

We outline in this section a proof of the Wick formula. Another proof is given in Section 3.8. It is based on a formula for the logarithm of a power series, and it has various combinatorial applications.

After an orthogonal change of variables the covariance matrix C can be made diagonal. Note now that both sides in the Wick formula are linear in f_i. It is therefore sufficient to prove it for products of coordinate functions, i. e., for monomials. For a diagonal matrix C we have $\langle x_i^2 \rangle = c_{ii}$, while $\langle x_i x_j \rangle = 0$ for $i \neq j$. Then the Wick formula becomes obvious:

$$\langle x_1^{2n_1} \ldots x_k^{2n_k} \rangle = \langle x_1^{2n_1} \rangle \ldots \langle x_k^{2n_k} \rangle = \left(\sum \langle x_1^2 \rangle \right) \ldots \left(\sum \langle x_k^2 \rangle \right),$$

where the ith sum is taking over all $(2n_i - 1)!!$ possible couplings of $2n_i$ copies of the monomial x_i. All the other couplings are equal to zero.

3.3.5 Gaussian measure on the space of Hermitian matrices

Let $H = (h_{ij})$ denote an Hermitian $(N \times N)$-matrix; that is, its entries are complex numbers: $h_{ij} \in \mathbf{C}$, and $h_{ij} = \overline{h}_{ji}$, where the bar denotes the complex conjugation. We denote by $\mathcal{H}_N$ the space of all such matrices. This space is an N^2-dimensional subspace in the $2N^2$-dimensional space of complex $N \times N$-matrices.

Every Hermitian matrix can be described by N^2 real parameters: $x_{ii} = h_{ii} \in \mathbf{R}$, $i = 1, \dots, N$ for diagonal entries, and $x_{ij} = \mathrm{Re}(h_{ij})$, $y_{ij} = \mathrm{Im}(h_{ij})$, $1 \leq i < j \leq N$ for over-diagonal entries (the under-diagonal entries being determined by the over-diagonal ones via complex conjugation). Thus, the space $\mathcal{H}_N$ is isomorphic to the vector space $\mathbf{R}^{N^2}$. We introduce the ordinary volume form in $\mathcal{H}_N$:

$$dv(H) := \prod_{i=1}^{N} dx_{ii} \prod_{i<j} dx_{ij} dy_{ij}.$$

In principle the space $\mathcal{H}_N \cong \mathbf{R}^{N^2}$ does not differ from any other vector space of the same dimension. But it is convenient to express some characteristics of the Gaussian measure on it in terms of matrix operations. In order to introduce a Gaussian density in the form (3.6), we need a nondegenerate quadratic form on $\mathcal{H}_N$. We take for such a form the following one: $\mathrm{tr}(H^2)$. This quadratic form is simply the restriction of the ordinary scalar product on the space $\mathbf{R}^{2N^2}$ of complex $N \times N$-matrices to the subspace of Hermitian matrices. For two complex matrices X and Y this scalar product looks like $(X, Y) = \mathrm{tr}(X\overline{Y}^t)$.

Let us see what does this form look like in the coordinates x_{ij}, y_{ij}. A term $(\cdot)_{ik}$ of H^2 is equal to $\sum_{j=1}^{N} h_{ij}h_{jk}$; hence a diagonal term $(\cdot)_{ii}$ is equal to $\sum_{j=1}^{N} h_{ij}h_{ji}$, and the trace of H^2 is equal to

$$\mathrm{tr}(H^2) = \sum_{i,j=1}^{N} h_{ij}h_{ji} = \sum_{i,j=1}^{N} h_{ij}\overline{h}_{ij} = \sum_{i=1}^{N} x_{ii}^2 + \sum_{i \neq j} (x_{ij}^2 + y_{ij}^2) = \sum_{i=1}^{N} x_{ii}^2 + 2\sum_{i<j} (x_{ij}^2 + y_{ij}^2).$$

Let us specify the matrix B of this quadratic form (see (3.6)). It must be of size $N^2 \times N^2$. But the above expression shows that this matrix is very simple; indeed, it is diagonal! We have N diagonal terms equal to 1 (they correspond to the coordinates x_{ii}), and $N^2 - N$ terms equal to 2 (they correspond to the coordinates x_{ij}, y_{ij} for $i < j$):

$$B = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 2 & & & \\ & & & & 2 & & \\ & & & & & \ddots & \\ & & & & & & 2 \end{pmatrix}.$$

Note that there is nothing unexpected in appearing of "twos" in the quadratic form matrix. Indeed, the subspace of off-diagonal Hermitian matrices in the space of off-diagonal complex matrices resembles the diagonal $x = y$ in the two-dimensional real plane $\mathbf{R}^2$. Restricting the ordinary scalar product in $\mathbf{R}^2$ to the diagonal and taking x as the coordinate on the diagonal we otain $(e, e) = 2$ for the unit diagonal vector.

For this matrix it is very easy to find all the necessary ingredients for the subsequent work. For example,

$$\det B = 2^{N^2 - N},$$

so the density (3.6), (3.7) takes the form

$$d\mu(H) = (2\pi)^{-N^2/2} 2^{(N^2-N)/2} \exp\{-\frac{1}{2}\mathrm{tr}(H^2)\} dv(H). \tag{3.8}$$

Exercise 3.3.10 Compute the covariance matrix $C = B^{-1}$. Infer from it that $<x_{ii}^2> = 1$, and $<x_{ij}^2> = <y_{ij}^2> = 1/2$ for $i < j$.

We will also need the following preparatory lemma:

Lemma 3.3.11 *Taking for linear functions of the coordinates x_{ij}, y_{ij} the matrix elements h_{ij}, we have*

$$<h_{ij} h_{ji}> = 1,$$

while all the other second moments are equal to zero:

$$<h_{ij} h_{kl}> = 0, \quad \text{whenever} \quad (i, j) \neq (l, k).$$

Proof. If $i = j$, then
$$< h_{ii}^2 > = < x_{ii}^2 > = 1.$$
If $i \neq j$, then
$$< h_{ij} h_{ji} > = < h_{ij} \overline{h}_{ij} > = < x_{ij}^2 + y_{ij}^2 > = \frac{1}{2} + \frac{1}{2} = 1.$$
All the other second moments of the type $< h_{ij} h_{kl} >$, $(i,j) \neq (l,k)$ involve only off-diagonal terms of the covariance matrix C.

Exercise 3.3.12 This exercise starts a series of exercises, which allow the reader to develop the theory of integration over the space of symmetric matrices parallel to that of Hermitian matrices by his/her own. Let $\mathcal{S}_N$ denote the space of real symmetric $N \times N$-matrices, i. e., for an element $S = (s_{ij}) \in \mathcal{S}_N,\ s_{ij} = s_{ji}$ for all $i, j = 1, \ldots, N$.

1. Verify that $\mathcal{S}_N$ is a subspace in the space of all real $N \times N$-matrices. Find the dimension of this subspace.
2. Verify that the trace form $\mathrm{tr}(S^2)$ determines a quadratic form on the space of symmetric matrices coinciding with the restriction of the ordinary scalar product $(X, Y) = \mathrm{tr}(XY^t)$ in the space of all real matrices.
3. Prove that $\langle s_{ij} s_{kl} \rangle \neq 0$ iff either $i = k, j = l$ or $i = l, j = k$.
4. Find the matrix of the trace form $\mathrm{tr}(S^2)$ and the scalar products $\langle s_{ij}^2 \rangle = \langle s_{ij} s_{ji} \rangle$.

   In what follows we denote by $d\mu(S) = c \exp\{-\frac{1}{2}\mathrm{tr}(S^2)\} dv(S)$ the Gaussian measure on the space of symmetric matrices. Here $dv(S) = \prod\limits_i ds_{ii} \prod\limits_{i<j} ds_{ij}$ and c is the normalizing constant chosen in such a way that $\int_{\mathcal{S}_N} d\mu(S) = 1$.
5. Find the value of the constant c.

3.3.6 Feynman diagrams

Let us compute the integral $\int\limits_{\mathcal{H}_N} \mathrm{tr}(H^4) d\mu(H)$. The integrand here is a sum of monomials
$$\mathrm{tr}(H^4) = \sum_{i,j,k,l=1}^{N} h_{ij} h_{jk} h_{kl} h_{li}.$$

The entries h_{kl} are linear functions in x_{kl}, y_{kl} and we can apply Wick's formula:

$$\begin{aligned}
\langle h_{ij}h_{jk}h_{kl}h_{li}\rangle &= \langle h_{ij}h_{jk}\rangle\langle h_{kl}h_{li}\rangle \\
&+ \langle h_{ij}h_{kl}\rangle\langle h_{jk}h_{li}\rangle \\
&+ \langle h_{ij}h_{li}\rangle\langle h_{jk}h_{li}\rangle. \qquad (3.9)
\end{aligned}$$

Would the integrand be more complicated, both expressions on the left and on the right hand sides would become a heap of sums, products, letters and indices. The Feynman diagram techniques has been invented specially for avoiding these cumbersome presentations. It is a method for expressing the integral in a compact geometric image.

Consider a 4-star. There are three ways to glue the ends of the four needles into pairs, Figure qu-ga-ff4. We establish a natural one-to-one correspondence between these three ways and three terms in the right hand side of Equation (3.9). The entries of the monomial $h_{ij}h_{jk}h_{kl}h_{li}$ may be considered as cyclically ordered and the right hand side presents all possible couplings of these entries.

In order to make the correspondence more clear t'Hooft suggested to double the needles of the star and to mark each needle with the corresponding pair of indices. Note that each (broken) segment is then marked twice with the same index. A gluing of the needles provides then a gluing of pairs of segments, Figure 3.7.

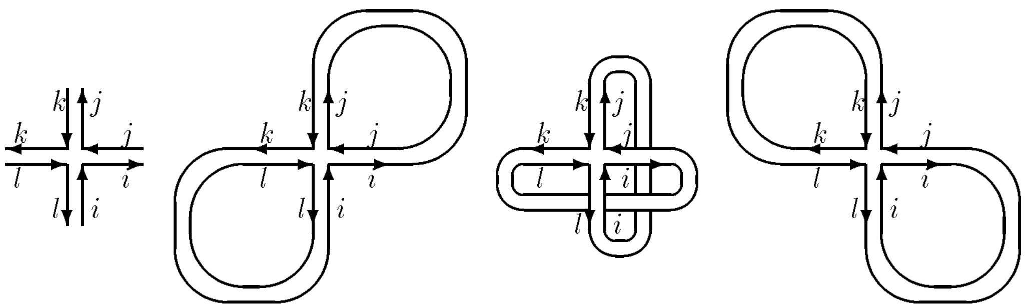


Figure 3.7: Feynman diagrams for the 4-star

Consider the first term in the right hand side of (3.9). The first factor $\langle h_{ij}h_{jk}\rangle$ differs from zero iff $i = k$. In this case it equals 1 (see Lemma 3.3.11).

The same statement is valid for the second factor $\langle h_{kl}h_{li}\rangle$. Since indices i, j, k, l vary from 1 to N the contribution of the first term summed over all indices equals N^3: the values of j, l and $i = k$ are independent. Note that the corresponding gluing has three faces marked with indices j, l and $i = k$ respectively.

The second term differs from zero iff $i = j = k = l$. Then the contribution of this term is N. The corresponding gluing has one face marked with the index $i = j = k = l$.

The third term is similar to the first one and its contribution is therefore N^3.

We conclude that

$$\int\limits_{\mathcal{H}_N} \operatorname{tr}(H^4)d\mu(H) = 2N^3 + N = T_2(N).$$

In the language of physics one would say that each of three gluings serves as a *Feynman diagram* for our integral.

On the other hand, gluing needles of the 4-star is dual to gluing a surface from the square. The last equation may be then written down as

$$\int\limits_{\mathcal{H}_N} \operatorname{tr}(H^4)d\mu(H) = \sum_{\tilde{\sigma}} N^{F(\tilde{\sigma})},$$

where the sum is taken over all gluings $\tilde{\sigma}$ of the 4-star and $F(\tilde{\sigma})$ is the number of faces of the gluing $\tilde{\sigma}$.

Precisely the same considerations for arbitrary star lead to the following statement.

Proposition 3.3.13

$$\int\limits_{\mathcal{H}_N} \operatorname{tr}(H^{2m})d\mu(H) = T_m(N) \textit{ for } m = 1, 2, \ldots. \tag{3.10}$$

Note that substituting $N = 1$ in this formula we obtain the formula $\int_{\mathbf{R}} x^{2m} d\mu(x) = (2m-1)!!$ from Section 3.2. We conclude thus that

integrating over the space $\mathcal{H}_N$ of Hermitian matrices allows to separate the ways of gluing the regular $2m$-gon according to the genus of the resulting surface.

The number of ways to obtain a surface of given genus is a coefficient of the resulting polynomial in N.

In order to make Proposition 3.3.13 more clear, let us consider one more example.

Example 3.3.14 Let $m = 4$, so we deal with the sum of the N^8 products of the form

$$h_{i_1 i_2} h_{i_2 i_3} h_{i_3 i_4} h_{i_4 i_5} h_{i_5 i_6} h_{i_6 i_7} h_{i_7 i_8} h_{i_8 i_1}.$$

Let us choose an arbitrary Wick coupling (out of $7!! = 105$ ones). For example, let us couple $h_{i_1 i_2}$ with $h_{i_4 i_5}$; $h_{i_2 i_3}$ with $h_{i_5 i_6}$; $h_{i_3 i_4}$ with $h_{i_8 i_1}$; and $h_{i_6 i_7}$ with $h_{i_7 i_8}$; in other words, let us consider the product

$$< h_{i_1 i_2} h_{i_4 i_5} > < h_{i_2 i_3} h_{i_5 i_6} > < h_{i_3 i_4} h_{i_8 i_1} > < h_{i_6 i_7} h_{i_7 i_8} > . \qquad (3.11)$$

Recalling Lemma 3.3.11 we see that each factor of this product has quite good chances to be equal to 0 (and then the whole product will become 0). In order for the product to be non zero we need *all* of its factors to be equal to 1 (then the product itself is also equal to 1). And for that we must impose rather restrictive conditions on indices (see once more Lemma 3.3.11):

$$\begin{array}{lcccl} < h_{i_1 i_2} h_{i_4 i_5} > & = & 1 & \Longleftrightarrow & i_1 = i_5, \ \ i_2 = i_4, \\ < h_{i_2 i_3} h_{i_5 i_6} > & = & 1 & \Longleftrightarrow & i_2 = i_6, \ \ i_3 = i_5, \\ < h_{i_3 i_4} h_{i_8 i_1} > & = & 1 & \Longleftrightarrow & i_3 = i_1, \ \ i_4 = i_8, \\ < h_{i_6 i_7} h_{i_7 i_8} > & = & 1 & \Longleftrightarrow & i_6 = i_8, \ \ i_7 = i_7. \end{array}$$

Finally we have the following "chains of equalities":

$$\begin{array}{lllllllll} i_1 & = & i_5 & = & i_3 & = & i_1 & & \\ i_2 & = & i_4 & = & i_8 & = & i_6 & = & i_2 \\ i_7 & = & i_7 & & & & & & \end{array}$$

which gives us N^3 possible combinations of indices (the indices i_1, i_2 and i_7 may be chosen arbitrarily, while the values of all the other indices are determined by this choice). This fact is usually expressed by saying that the *contribution* of the coupling (3.11) is equal to N^3.

Consider now a polygon with 8 sides, and glue its sides in pairs as is shown in Figure 3.8.

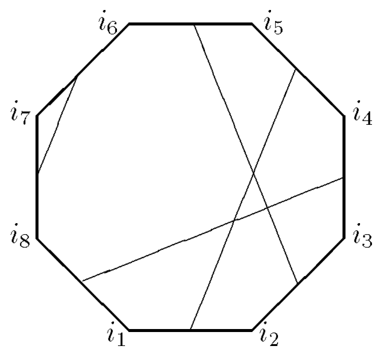


Figure 3.8: A polygon whose sides are glued together in pairs

The fact that the side labeled i_1i_2 is glued to the one labeled i_4i_5 means that the polygon vertex i_1 is identified with i_5, as well as i_2 is identified with i_4. We express this by writing

$$i_1 = i_5, \quad i_2 = i_4.$$

In the same vein, the side i_2i_3 is glued to i_5i_6, hence

$$i_2 = i_6, \quad i_3 = i_5,$$

and the two remaining identifications give us

$$i_3 = i_1, \quad i_4 = i_8, \quad i_6 = i_8, \quad i_7 = i_7.$$

Finally we see that the 8 original vertices of the polygon are identified in the following way (thus producing 3 vertices of the corresponding map):

$$\begin{array}{lllllllll} i_1 & = & i_5 & = & i_3 & = & i_1 & & \\ i_2 & = & i_4 & = & i_8 & = & i_6 & = & i_2 \\ i_7 & = & i_7 & & & & & & \end{array}$$

Exercise 3.3.15 (Continuation of Exercise 3.3.12.)

1. Compute the integrals

$$\int_{\mathcal{S}_N} \mathrm{tr}(S^2) d\mu(S), \quad \int_{\mathcal{S}_N} \mathrm{tr}(S^2) d\mu(S).$$

2. Define a Feynman diagram for gluing an arbitrary (either orientable or non-orientable surface) from a $2n$-star. Draw all gluings of the 2-star and of the 4-star.

3. Find the number of arbitrary (either orientable or non-orientable) gluings of the $2n$-star.

4. Define the *number of faces* $F(\sigma)$ for a non-orientable gluing σ.

5. Prove that
$$\sum_{\sigma} N^{F(\sigma)} = 2^n \int_{\mathcal{S}_N} \operatorname{tr}(S^{2n}) d\mu(S)$$
where the sum in the left hand side is taken over all possible (either orientable or non-orientable) gluings of the $2n$-star.

3.3.7 Computation of Gaussian integrals. The unitary invariance

The main observation allowing to compute the integral is that the integrand $\operatorname{tr}(H^{2m})$ is invariant under the action of the unitary group $\mathcal{U}_N$. In practice, functions coming from physics usually satisfy this assumption. The integral can then be simplified which allows us in some situations to find its asymptotics or to compute it explicitly.

First let us define the action of the unitary group on the space of Hermitian matrices.

The *unitary group* $\mathcal{U}_N$ consists of *unitary matrices*, i. e., of complex $N\times N$-matrices U such that $|U^t\overline{U} = 1$. Here U^t is the transposed matrix, and $\overline{U}$ is the complex conjugate matrix to U.

Exercise 3.3.16 Prove that the set of unitary matrices forms a group with respect to matrix multiplication. Prove that this group is compact.

The group $\mathcal{U}_N$ acts on the space $\mathcal{H}_N$ of Hermitian matrices of order N by $H \mapsto U^{-1}HU$. This action is called the *adjoned action*. The set of eigenvalues of an Hermitian matrix U is invariant with respect to this action.

Exercise 3.3.17 Verify that this mapping determines an action of $\mathcal{U}_N$ on $\mathcal{H}_N$. (The main point in the verification is checking that the matrix $U^{-1}HU$ is an Hermitian matrix.)

The construction below resembles "polar coordinates" and integration of "rotation-invariant" functions. Diagonal Hermitian matrices correspond to the radial coordinate while unitary matrices correspond to the angle coordinate.

Let F be a polynomial on $\mathcal{H}_N$ invariant with respect to this action, i.e. $F(H) = F(U^{-1}HU)$ for any Hermitian matrix H and any unitary matrix U. There exists a natural one-to-one correspondence between invariant polynomials and symmetric polynomials in eigenvalues of unitary matrices, or polynomials in traces of H^k: any invariant polynomial restricted to the subspace of diagonal matrices is a symmetric polynomial in the diagonal elements.

Proposition 3.3.18

$$\int_{\mathcal{H}_N} F(H)d\mu(H) = c_N \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} F(\Lambda) \prod_{1\leq i<j\leq N} (\lambda_i - \lambda_j)^2 d\lambda_1 \ldots d\lambda_N \qquad (3.12)$$

for some constant c_N depending only on the dimension N of matrices. Here the matrix Λ is the diagonal $N \times N$-matrix with $\lambda_1, \ldots, \lambda_N$ on the diagonal.

In this form the proposition belongs to H. Weyl [37], see also [38]. The precise value of the constant c_N is given in [17]; we are not sufficiently sure, however, that there are no predecessors. We present the computation of c_N below in Sec. 3.5.

Let us remark also that there is nothing unexpected in the presentation (3.12). Indeed, the λ-Jacobian $\prod_{1\leq i<j\leq N}(\lambda_i - \lambda_j)^2$ is the symmetric polynomial of the smallest degree, which vanishes whenever two eigenvalues coincide. And our "change of variables" fails to be this precisely on the subset of matrices with coinciding eigenvalues. It is convinient sometimes to present this Jacobian as the square of the Vandermonde determinant:

$$\prod_{1\leq i<j\leq N} (\lambda_i - \lambda_j)^2 = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_N \\ \lambda_1^2 & \lambda_2^2 & \ldots & \lambda_N^2 \\ \ldots & \ldots & \ddots & \ldots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \ldots & \lambda_N^{N-1} \end{vmatrix}^2$$

Proof. Each Hermitian matrix H is unitary equivalent to a diagonal matrix Λ having its eigenvalues on the diagonal. The presentation $H = U^{-1}\Lambda U$ is unique up to

1. The matrix Λ is generally determined up to $N!$ permutations of the eigenvalues (each permutation causes the corresponding permutation of columns of the matrix U). The word "generally" means: except the set of matrices with at least two coinciding eigenvalues.

2. The matrix U is determined up to multiplication by a diagonal unitary matrix with numbers $e^{i\theta_k}$, $\theta_k \in \mathbf{R}$ for $k = 1, \ldots, N$ on the diagonal. (The set of all such diagonal matrices is an n-dimensional torus $\mathbf{T}^n$.)

Note that the subset of Hermitian matrices with coinciding eigenvalues has measure zero and is therefore negligible from the integration point of view.

For any matrix H without coinciding eigenvalues the preimage of the map

$$(\Lambda, U) \mapsto H = U^{-1}\Lambda U$$

consists of $N!$ copies of the N-dimensional torus.

Recall that we use matrix notation as a convinient tool for presenting a large number of independent coordinates. For a matrix $A = (a_{ij})$, denote by dA the matrix $dA = (da_{ij})$.

Exercise 3.3.19 1. Verify the Leibnits identity for this differential: for any two $N \times N$ matrices A and B

$$d(AB) = (dA)B + A(dB);$$

2. Verify that the analog to the formula $d(1/x) = -(1/x^2)dx$ for functions is the formula $d(A^{-1}) = -A^{-1}dAA^{-1}$.

Differentiating the formula $H = U^{-1}\Lambda U$ we obtain

$$\begin{aligned}
dH &= d(U^{-1}) \cdot \Lambda U + U^{-1} \cdot d\Lambda \cdot U + U^{-1}\Lambda \cdot dU \\
&= U^{-1} \cdot U \cdot d(U^{-1}) \cdot \Lambda U + U^{-1} \cdot d\Lambda \cdot U + U^{-1}\Lambda \cdot dU \cdot U^{-1} \cdot U \\
&= U^{-1}(-\Omega\Lambda + d\Lambda + \Lambda\Omega)U \\
&= U^{-1}AU;
\end{aligned}$$

here $\Omega = dU \cdot U^{-1}$, $A = -\Omega\Lambda + d\Lambda + \Lambda\Omega$ and we use the relation $U \cdot d(U^{-1}) = -\Omega$ following from $0 = d(UU^{-1}) = d(U) \cdot U^{-1} + U \cdot d(U^{-1})$.

Denote entries of the matrix Ω by ω_{ij}, i. e., $\Omega = (\omega_{ij})$. The matrix $A = d\Lambda + \Lambda\Omega - \Omega\Lambda$ has diagonal entries $d\lambda_i$ and off-diagonal entries $(\lambda_i - \lambda_j)\omega_{ij}$. The volume form is the product of all the entries of A and it is thus

$$\prod_{1 \le i < j \le N} (\lambda_i - \lambda_j)^2 d\lambda_1 \dots d\lambda_N \times (\text{some differential form independent of } \lambda). \tag{3.13}$$

The adjoint action of a unitary matrix U on a matrix A does not change the volume form for the matrix A. Indeed, suppose the converse. Then $dv(U^{-1}AU) = c_U dv(A)$, where c_U is a function depending on coordinates U, Λ. Applying this action k times we obtain $dv(U^{-k}AU^k) = c_U^k dv(A)$. On the other hand, a number k can be chosen such that U^k is arbitrarily close to the unit matrix since the unitary group is compact. This contradiction proves the statement.

Therefore, after integrating over a fiber we obtain the form

$$c_N \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_i d\lambda_i$$

on the space $\mathbf{R}^N$ of diagonal matrices.

The Proposition is proved.

Let us understand now the geometric meaning of the constant c_N. The restriction of the quadratic form $\mathrm{tr}(A\overline{a}^t)$ to the unitary group $\mathcal{U}_N$ detrmines a metric on this group. Associate with this metric a volume form on $\mathcal{U}_N$ and denote by $\mathrm{vol}\,\mathcal{U}_N$ the total volume of the unitary group with respect to this metric. Then c_N coincides with $\mathrm{vol}(\mathcal{U}_N)$ up to a natural factor.

Lemma 3.3.20

$$c_N = N!(2\pi)^N \,\mathrm{vol}(\mathcal{U}_N).$$

The coefficient in the expression for c_N reflects the ambiguity in matrix presentation $H = U^{-1}\Lambda U$: the matrix Λ is uniquely determined up to $N!$ permutations of diagonal elements, and whenever Λ is specified, the matrix U is unique up to the $\mathbf{T}^n$ goup action. The coefficient $(2\pi)^N$ is precisely the volume of this group with respect to the above metric.

Proof. Equation 3.13 shows that the volume form in the space of Hermitian matrices splits into a product of two volume forms: one in variables λ, and another one in coordinates along an orbit of the unitary group action. The constant c_N is simply the volume of such an orbit (volumes of all orbits coincide).

Exercise 3.3.21 (Continuation of Exercise 3.3.15.)

1. Verify that the group SO_N of orientation preserving orthogonal $N \times N$-matrices acts on the space of $N \times N$ symmetric matrices by $S \mapsto O^{-1}SO$ for $O \in \mathrm{SO}_N,\ S \in \mathcal{S}_N$.

2. Prove that for a function F on the space of symmetric matrices invariant with respect to the action of the orthogonal group

$$\int_{\mathcal{S}_N} F(S)d\mu(S) = \tilde{c}_N \int_{\mathbf{R}^N} F(\Lambda) \prod_{i<j} (\lambda_i - \lambda_j)^2 d\mu(\lambda_1) \ldots d\mu(\lambda_N),$$

where $\tilde{c}_N$ is a constant depending only on N. Its precise value will be a result of a further exercise.

3.3.8 Computation of the integral for one face gluings

This section is devoted to the proof of Theorem 3.2.5.

The computation of the integral (3.10) splits into two stages. First of all we prove that for a fixed value of N the function $t(N,n) = \frac{T_n(N)}{(2n-1)!!}$ is a polynomial in n. The degree of this polynomial is $N-1$. In particular, as we have seen before, $t(1,n) \equiv 1$. Strangely, this proof is the only point where the matrix integral representation of the generating function is essential.

The second step is a purely combinatorial computation of this polynomial.

Lemma 3.3.22 *The function $t(N,n) = \frac{T_n(N)}{(2n-1)!!}$ is polynomial in n.*

Proof. Let us make use of Equation 3.12. It yields

$$\int_{\mathcal{H}_N} \operatorname{tr} H^{2n} d\mu(H) = c_N \int_{\mathbf{R}^N} \operatorname{tr} \Lambda^{2n} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{N} d\mu(\lambda_i).$$

The integrand $\operatorname{tr} \Lambda^{2n} = \lambda_1^{2n} + \ldots + \lambda_N^{2n}$ may be substituted with $N\lambda_1^{2n}$ since the integral values for all the summands are the same because of the symmetry. Expanding the square of the Vandermonde determinant as a polynomial in λ_1 and integrating over all other variables $\lambda_2, \ldots, \lambda_N$ we obtain an integral over one variable λ_1. The integrand is a polynomial of degree $2n + 2N - 2$.

The coefficients of the polynomial are constant for fixed N. The leading coefficient is Nc_N.

Integrating a monomial λ_1^{2n+2k} and dividing the result by $(2n-1)!!$ we obtain $\frac{(2n+2k-1)!!}{(2n-1)!!}$, which is a polynomial of degree $k-1$ in n. The Lemma is proved.

Problem 3.3.23 Find a purely combinatorial proof of the previous Lemma.

Now suppose that vertices of our polygon are colored in some colors. We say that a gluing *agree with the coloring* if only vertices of the same color are glued together.

Lemma 3.3.24 *The number $T_n(N)$ is precisely the number of gluings of the $2n$-gon that agree with possible colorings of the vertices in (not more than) N colors.*

Indeed, let the boundary of the polygon become after a gluing an embedded graph with V vertices. Color each of the vertices in one of N colors. Any such coloring determines a coloring of the vertices of the polygon itself, and the gluing agree with this coloring. And there is precisely N^V ways to color the vertices of the graph in N colors. (Note that we do not require that neighboring vertices have different colors.)

Introduce now the function $\tilde{T}_n(N)$, the number of gluings of the $2n$-gon that agree with colorings of its vertices in precisely N colors. Then

$$T_n(N) = \sum_{L=1}^{N} \binom{N}{L} \tilde{T}_n(L).$$

Indeed, there are $\binom{N}{L}$ different ways to choose L colors from N given colors. Obviously, $\tilde{T}_0(N) = \tilde{T}_1(N) = \ldots = \tilde{T}_{N-2}(N) = 0$. Indeed, the resulting graph on the surface has at most $n+1$ vertices (in this case it is a tree). Therefore there is no coloring of the vertices of a $2n$-gon in more than $n+1$ different colors that may agree with any gluing.

The function $\tilde{t}(N,n) = \frac{\tilde{T}_n(N)}{(2n-1)!!}$ is a polynomial in n of degree $N-1$. We know $N-1$ roots $0, 1, 2, \ldots, N-2$ of this polynomial. Therefore there exists a constant A_N such that

$$\tilde{t}(n,N) = A_N n(n-1)(n-2)\ldots(n-N+2) = A_N \binom{n}{N-1}(N-1)!.$$

Substituting this formula in the expression for $T_n(N)$ yields

$$T_n(N) = (2n-1)!! \sum_{L=1}^{N} A_L \binom{n}{L-1} \binom{N}{L} (L-1)!$$

In particular, the leading coefficient in N of the polynomial $T_n(N)$ equals

$$(2n-1)!! \frac{A_{n+1}}{(n+1)!}.$$

On the other hand, we know that this coefficient is the n-th Catalan number:

$$(2n-1)!! \frac{A_{n+1}}{(n+1)!} = c_n = \frac{(2n)!}{n!(n+1)!}.$$

Then it follows that $A_{n+1} = \frac{2^n}{n!}$. Thus,

$$T_n(N) = (2n-1)!! \sum_{L=1}^{N} 2^{L-1} \binom{N}{L} \binom{n}{L-1}.$$

This is precisely the formula for the expansion of the function $\left(\frac{1+s}{1-s}\right)^N$ in powers of s.

The Harrer-Zagier theorem is proved.

3.4 Universal one-matrix model

3.4.1 The matrix integral for an arbitrary gluing

Suppose now we investigate gluing a set of stars. Then a matrix integral allows to enumerate embedded graphs thus obtained according to their genera. It may happen however that a resulting graph is disconnected, so the notion of "genus" must be made more precise.

Suppose we consider a set of α_1 1-stars, α_2 2-stars,..., α_k k-stars ($\alpha_1, \dots, \alpha_k$ are nonnegative integers). Consider a gluing σ of the ends of the needles. Such a gluing determines a set of embedded graphs $\Gamma_1, \dots, \Gamma_m$: each graph in the set is a connected component of the glued graph. Define the number of faces $F(\sigma)$ as the sum of numbers of faces for all connected components,

$$F(\sigma) = \sum_{i=1}^{m} F(\Gamma_i).$$

We are now in position to state the enumerative formula.

Proposition 3.4.1

$$\sum_{\sigma} N^{F(\sigma)} = \int\limits_{\mathcal{H}_N} (\operatorname{tr} H)^{\alpha_1} (\operatorname{tr} H^2)^{\alpha_2} \dots (\operatorname{tr} H^k)^{\alpha_k} d\mu(H)$$

where the sum is taken over all possible couplings of the needles.

The proof of this proposition is absolutely the same as the proof of Proposition 3.3.13. The integrand in the right hand side may be presented as the sum of monomials of the type

$$h_{i_1 i_1} \dots h_{i_{\alpha_1} i_{\alpha_1}} h_{j_1 k_1} h_{k_1 j_1} \dots h_{j_{\alpha_2} k_{\alpha_2}} h_{k_{\alpha_2} j_{\alpha_2}} \dots$$

The further computation looks as follows.

1. Label each needle of a star with a pair of indices.
2. Express the integral of each monomial through the Wick formula.
3. Each monomial with non zero contribution determines a Wick coupling σ of the needles.
4. Compute the number of possible values of indices producing non zero contribution to the value of the integral. This number equals precisely $N^{F(\sigma)}$.
5. Sum over all possible couplings.

The Proposition is proved.

Exercise 3.4.2 (Continuation of Exercise 3.3.21.)
Prove that

$$\sum_{\sigma} N^{F(\sigma)} = 2^{\frac{1}{2}(1\cdot\alpha_1 + 2\cdot\alpha_2 + \dots + k\cdot\alpha_k)} \int\limits_{\mathcal{S}_N} (\operatorname{tr} S)^{\alpha_1} (\operatorname{tr} S^2)^{\alpha_2} \dots (\operatorname{tr} S^k)^{\alpha_k} d\mu(S)$$

where the sum is taken over all (either orientable or not) gluings of the needles.

3.4.2 Getting rid of disconnected graphs

The formula in Proposition 3.4.1 enumerates both connected and disconnected graphs. It is not convenient since we are regularly interested in enumerating "elementary objects" — connected graphs. Unfortunately there is no regular way to get rid of disconnected graphs in an isolated formula like this. However, disconnected graphs can be excluded if we consider the exponential generating function for a family of finite sets of glued stars, not for a single set of stars.

As an example consider a problem of gluing n 4-stars for $n = 0, 1, 2, \ldots$, see Figure 3.9.

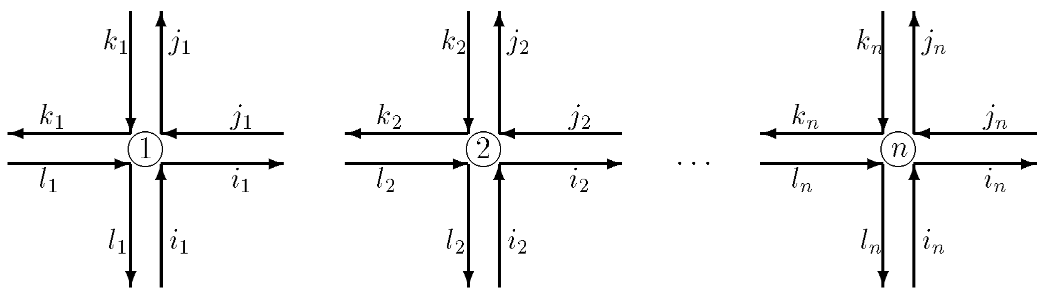


Figure 3.9: An image for $(\mathrm{tr}(H^4))^n$

As we know, the function

$$F_n = \int_{\mathcal{H}_N} (\mathrm{tr}\, H^4)^n d\mu(H)$$

can be represented as a polynomial in N:

$$F_n(N) = \sum_\sigma N^{F(\sigma)}.$$

As we have seen, $F_1(N) = 2N^3 + N$.

Exercise 3.4.3 Compute $F_2(N), F_3(N)$.

Now denote by $C_n(N),\ n = 1, 2, \ldots$ the sum

$$C_n(N) = \sum_{\sigma_c} N^{F(\sigma_c)}$$

taken over all gluings σ_c that provide connected graphs.

Consider two exponential generating functions

$$F(s, N) = \sum_{n=0}^{\infty} \frac{1}{n!} F_n(N) s^N$$

$$C(s, N) = \sum_{n=1}^{\infty} \frac{1}{n!} C_n(N) s^N,$$

where s is a formal parameter. (Note that the second sum starts with the term $C_1(N)s$.)

Proposition 3.4.4 *The generating functions $F(s, N)$ and $C(s, N)$ are related by*

$$F(s, N) = \exp(C(s, N)).$$

The equality in Proposition has in fact nothing special to do with gluings of 4-stars. The same statement is valid in much more general situation. The operation of taking the logarithm is very well known in combinatorics; see, for example, [21]. If one has an *exponential* generating function for a class of *labeled* objects, then its logarithm is the generating function for the *connected* objects of the same type.

Recall that the word "exponential" means "with $n!$ in the denominator". The word "labeled" must be explained at more length. But we will limit ourselves to a single remark. One may consider objects (such as graphs) with additional structures (such as cyclic orders of edges around vertices) and with some properties imposed (such as prescribed vertex degrees). But one must be careful: a property has a legal right to be imposed when *it is valid for an object if and only if it is valid for all its connected components.* For example, the property

Every vertex of a graph is colored in one of k colors

is acceptable; on the contrary, the property

> Every vertex of a graph is colored in one of k colors, and *all k colors are used*

is not good, because it may be valid for a graph as a whole but not valid for some of its components.

Now let us give the precise formal statement. Consider a graded set of objects $\mathcal{O} = \mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{O}_3 \dots$ satisfying the following properties:

1. Each of the object sets $\mathcal{O}_k$ is finite;

2. An operation of union of the objects $\mathcal{O} \times \mathcal{O} \to \mathcal{O}$ is defined, which takes a pair (o_1, o_2) with $o_1 \in \mathcal{O}_k, o_2 \in \mathcal{O}_m$ into an object $o_1 + o_2 \in \mathcal{O}_{k+m}$;

3. Any object $o \in \mathcal{O}$ admits the unique (up to the order) decomposition into the union of connected objects $o = o_1 + o_2 + \dots + o_k$. (A *connected object* is an object o that admits no decomposition $o = o_1 + o_2$.)

A *labeled* object is an object $o \in O_n$ together with the decomposition $o = o_1 + o_2 + \dots$ into the union of connected objects and a *labeling*: the set of indices $\{1, \dots, n\}$ is distributed among the connected objects $o_1, o_2, \dots$ in such a way that each of the objects obtains the number of indices equal to its grading.

Denote the subset of connected objects by $C \subset \mathcal{O}$, $C = C_1 \cup C_2 \cup \dots$ and denote the set of labeled objects by $L = L_1 \cup L_2 \cup \dots$.

Let now $F : \mathcal{O} \to K$ be a multiplicative mapping of the set $\mathcal{O}$ into a ring K, i.e. $F(o_1 + o_2) = F(o_1)F(o_2)$. We can set, for example, $F(o) = t^n \in \mathbf{Z}[t]$ for $o \in \mathcal{O}_n$.

Consider two exponential generating functions

$$F_L(s) = 1 + \frac{1}{1!}s \sum_{o \in L_1} F(o) + \frac{1}{2!}s^2 \sum_{o \in L_2} F(o) + \dots$$

$$F_C(s) = \frac{1}{1!}s \sum_{o \in \mathbf{C}_1} F(o) + \frac{1}{2!}s^2 \sum_{o \in \mathbf{C}_2} F(o) + \dots$$

for the set of labeled objects and for the set of connected objects respectively.

Proposition 3.4.5

$$F_L(s) = \exp(F_C(s)).$$

Proof. Let us look at the coefficient at s^n in the exponent in the right hand side. A contribution to this coefficient is given by a set of connected objects $o_1, \ldots, o_k$ such that $o = o_1 + \ldots + o_k \in O_n$. Suppose $o_1 \in O_{i_1}, \ldots, o_k \in O_{i_k}$, where $i_1 + \ldots + i_k = n$. Then the contribution is equal to $\frac{1}{i_1!} \ldots \frac{1}{i_k!}$. On the other hand, the disconnected object o admits $\frac{n!}{i_1! \ldots i_k!}$ labelings. Therefore its contribution to the coefficient at s^n in the left hand side generating function is $\frac{1}{i_1! \ldots i_k!}$.

The Proposition is proved.

3.4.3 The integral of the universal model

Let us set

$$\mathcal{F}(N; t_1, t_2, \ldots) = \log \int_{\mathcal{H}_N} e^{-\operatorname{tr}(t_1 H + \frac{1}{2} t_2 H^2 + \frac{1}{3} t_3 H^3 + \frac{1}{4} t_4 H^4 + \ldots)} d\mu(H). \qquad (3.14)$$

As we have seen before this integral describes all (connected) gluings of all possible sets of stars. The function $\mathcal{F}$ is a formal power series in the infinite set of variables $t_1, t_2, t_3, \ldots$. The coefficient at a monomial $t_1^{\alpha_1} t_2^{\alpha_2} t_3^{\alpha_3} \ldots$ is a polynomial in N. The coefficient at N^k of this polynomial equals the number of connected gluings of α_1 1-stars, α_2 2-stars, ... with k faces.

The coefficient $\frac{1}{n}$ at $t_n H^n$ reflects the fact that an n-star admits the group $\mathbf{Z}_n$ as the group of automorphisms. The group $\mathbf{Z}_n$ acts on an n-star by rotations.

The function $\mathcal{F}$ is called the *integral of the universal one–matrix model.*

Exercise 3.4.6 Gluing stars compute first few coefficients of the formal series (3.14).

One can associate with the function 3.14 a formal power series F in variables $t_1, t_2, t_3, \ldots$ whose coefficients are numbers, not polynomials in N. The construction proceeds as follows.

Let us fix a finite set of stars including α_k of k-stars for $k = 1, \ldots, m$, and let n be greater than m.

Proposition 3.4.7 *For α_n large enough the coefficient at $N^{2-2g} t_1^{\alpha_1} \ldots t_m^{\alpha_m} t_n^{\alpha_n}$ in the series* (3.14) *grows as*

$$e^{c_1 \alpha_n} \alpha_n^{c_2 (2-2g)-1} f_g(\alpha_1, \ldots, \alpha_m).$$

Here c_1, c_2 are some constants, and the factor $f_g(\alpha_1, \ldots, \alpha_m)$ does not depend on the choice of n.

3.5 Computation of matrix integrals

3.5.1 Example: computation of the unitary group volume

In order to compute the unitary group volume we need to compute the integral

$$\int_{\mathbf{R}^N} \prod_{i<j} (\lambda_i - \lambda_j)^2 d\mu(\lambda_1) \dots d\mu(\lambda_N). \tag{3.15}$$

This is the most simple integral of the kind we are interested in. Indeed, it corresponds to integrating over the space of Hermitian matrices of the function identically equal to one.

Hermite polynomials provide a very convenient tool to compute this integral.

The *n-th Hermite polynomial* is the polynomial

$$H_n(\lambda) = (-1)^n e^{\frac{x^2}{2}} \left(e^{-\frac{x^2}{2}} \right)^{(n)},$$

where the symbol (n) denotes the n-th derivative.

Here are the first few Hermite polynomials:

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \\ H_4(x) &= x^4 - 6x^2 + 3 \end{aligned}$$

The usefulness of Hermite polynomials is due to the properties they satisfy:

1. H_n is a polynomial in x of degree n;

2. the leading coefficient of H_n equals 1;

3. polynomials H_n form an orthogonal set of polynomials with respect to the standard Gaussian measure on the line: $(H_n, H_m) = 0$ whenever

$n \neq m$; here (P, Q) denotes the scalar product for two polynomials P, Q:

$$(P, Q) = \int_{\mathbf{R}} PQ\, d\mu(\lambda).$$

4. H_n is the unique sequence of polynomials satisfying the previous properties.

Exercise 3.5.1 Prove the above properties of Hermite polynomials.

It is convenient to substitute Hermite polynomials in variables λ_i for the monomials in the Vandermonde determinant:

$$\begin{aligned}\prod_{i<j}(\lambda_i - \lambda_j) &= \begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_N \\ \lambda_1^2 & \lambda_2^2 & \dots & \lambda_N^2 \\ \dots & \dots & \ddots & \dots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \dots & \lambda_N^{N-1} \end{vmatrix} \\ &= \begin{vmatrix} H_0(\lambda_1) & H_0(\lambda_2) & \dots & H_0(\lambda_N) \\ H_1(\lambda_1) & H_1(\lambda_2) & \dots & H_1(\lambda_N) \\ H_2(\lambda_1) & H_2(\lambda_2) & \dots & H_2(\lambda_N) \\ \dots & \dots & \ddots & \dots \\ H_{N-1}(\lambda_1) & H_{N-1}(\lambda_2) & \dots & H_{N-1}(\lambda_N) \end{vmatrix}\end{aligned}$$

The square of the last determinant is a polynomial in $H_i(\lambda_j)$. Each term $H_i(\lambda_j)$ has degree not more than 2 in each of the monomials of the determinant expansion. Let us take the integral separately in each of the variables $\lambda_1, \dots, \lambda_n$. Then the orthogonality property of the Hermite polynomials implies that only monomials of the form $H_0^2(\lambda_{\sigma(1)})H_1^2(\lambda_{\sigma(2)})\dots H_{N-1}^2(\lambda_{\sigma(N)})$ contribute to the integral. Here σ is a permutation on the set of indices $1, \dots, N$. The number of permutations is $N!$ and the contribution of all these monomials is the same. Whence,

$$\int_{\mathbf{R}^N} \prod_{i<j}(\lambda_i - \lambda_j)^2 d\mu(\lambda_1)\dots d\mu(\lambda_N) = N!(H_0, H_0)(H_1, H_1)\dots(H_{N-1}, H_{N-1}).$$

The only thing which remains to be computed is the scalar squares (H_i, H_i). In order to compute them let us find a recurrency formula for these values. The definition of Hermite polynomials obviously imply

$$\begin{aligned}
H_n &= (-1)^n e^{\frac{x^2}{2}} \left(e^{-\frac{x^2}{2}}\right)^{(n)} \\
&= -e^{\frac{x^2}{2}} \left(\left((-1)^{n-1} e^{-\frac{x^2}{2}}\right)^{(n-1)}\right)' \\
&= -e^{\frac{x^2}{2}} \left(H_{n-1} e^{-\frac{x^2}{2}}\right)' \\
&= -e^{\frac{x^2}{2}} \left(H_{n-1}' e^{-\frac{x^2}{2}} - x H_{n-1} e^{-\frac{x^2}{2}}\right) \\
&= x H_{n-1} - H_{n-1}'
\end{aligned}$$

On the other hand, $H_{n-1}' = (n-1)H_{n-2}$, and we obtain

$$\begin{aligned}
(H_n, H_n) &= (H_n, xH_{n-1} - (n-1)H_{n-2}) \\
&= (H_n, xH_{n-1}) \\
&= (xH_n, H_{n-1}) \\
&= (H_{n+1} + nH_{n-1}, H_{n-1}) \\
&= n(H_{n-1}, H_{n-1})
\end{aligned}$$

Whence

$$(H_n, H_n) = n!$$

and, finally

$$\int_{\mathbf{R}^N} \prod_{i<j} (\lambda_i - \lambda_j)^2 d\mu(\lambda_1) \ldots d\mu(\lambda_N) = N!(N-1)! \ldots 2!1!.$$

3.5.2 Generalized Hermite polynomials

In order to compute the integral

$$\int_{\mathcal{H}_N} e^{t \operatorname{tr} H^4} d\mu(H) = c_N \int_{\mathbf{R}^N} e^{-t(\lambda_1^4 + \ldots + \lambda_N^4)} \prod (\lambda_i - \lambda_j)^2 d\mu(\lambda)$$

we introduce generalized Hermite polynomials.

Consider the measure

$$d\nu_V(\lambda) = e^{-V(\lambda)} d\mu(\lambda).$$

For the time being we approve that $V = V(\lambda, t) = t\lambda^4$. For two functions P, Q denote by $(P, Q)_V$ their scalar product with respect to this measure, $(P, Q)_V = \int PQ d\nu_V$.

The *generalized Hermite polynomials* with respect to this measure are the polynomials (in λ) $P_{V,0}(\lambda, t), P_{V,1}(\lambda, t), \dots$ such that

1. The polynomial $P_{V,k}$ is a polynomial of degree k with the leading term λ^k.

2. The polynomials $P_{V,k}$ are orthogonal with respect to the measure $d\nu_V$, $(P_{V,k}, P_{V,m})_V = 0$ for $m \neq k$.

In the case $V \equiv 0$ these polynomials are classic Hermite polynomials.

Exercise 3.5.2 Prove that the polynomials $P_{V,k}$ are even for k even and odd for k odd.

Denote the scalar square of the polynomial $P_{V,k}$ with respect to the measure ν_V by $h_{V,k} = h_{V,k}(t)$:

$$h_{V,k} = (P_{V,k}, P_{V,k})_V$$

The reason why the generalized Hermite polynomials are introduced is that they allow one to get rid of the square of the Vandermonde determinant factor in the integrand. Indeed, substituting linear combinations of the rows of the Vandermonde determinant instead of the original rows, we can change the polynomials in the determinant with the generalized Hermitian polynomials:

$$\begin{aligned}
\Delta(\lambda) &= \prod(\lambda_i - \lambda_j) \\
&= \begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_N \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \dots & \lambda_N^{N-1} \end{pmatrix} \\
&= \begin{pmatrix} P_0(\lambda_1) & P_0(\lambda_2) & \dots & P_0(\lambda_N) \\ P_1(\lambda_1) & P_1(\lambda_2) & \dots & P_1(\lambda_N) \\ \vdots & \vdots & \ddots & \vdots \\ P_{N-1}(\lambda_1) & P_{N-1}(\lambda_2) & \dots & P_{N-1}(\lambda_N) \end{pmatrix}
\end{aligned}$$

The expansion of the determinant and separate integration over each of N variables λ_k give

$$\int_{\mathcal{H}_N} e^{-t\,\mathrm{tr}\,H^4} d\mu(H) = c_N N! h_{N-1} h_{N-2} \dots h_0.$$

Note that this representation of the integral as a product is very convenient for our purposes since we need to take a logarithm of the integral.

Our goal now is to calculate the values of the constants h_k. For small values of k these constants may be explicitly calculated as some integrals but these formulae carry small information on their behavior with k increasing. So we need some more effective tool. Such a tool is a recurrency formula.

Consider the operator of multiplication by λ on the space of polynomials in λ and find the matrix of this presentation in the basis $P_{V,k}$.

Lemma 3.5.3 $\lambda P_{V,k} = P_{V,k+1} + R_k P_{V,k-1}$ *for some constant* $R_k = R_k(t)$.

Proof. Indeed, the polynomial $\lambda P_{V,k}$ has the degree $k+1$ so it is a linear combination of the polynomials $P_{V,0}, \dots, P_{V,k+1}$. The coefficient at $P_{V,k+1}$ is 1 because the leading coefficients coincide. The coefficient at $P_{V,k}$ is zero because of the parity. The coefficient at $P_{V,m}$ for $m < k-1$ vanishes because of $(\lambda P_{V,k}, P_{V,m})_V = (P_{V,k}, \lambda P_{V,m})_V = 0$ since $\lambda P_{V,m}$ is a polynomial of degree $< k$.

The Lemma is proved.

The constants $h_{V,k}$ are related to the constants $R_{V,k}$ by the formula

$$\begin{aligned} h_{V,k} &= (P_{V,k}, \lambda P_{V,k-1})_V \\ &= (\lambda P_{V,k}, P_{v,k-1})_V \\ &= (P_{V,k+1} + R_{V,k} P_{V,k-1}, P_{V,k-1})_V \\ &= R_{V,k} h_{V,k-1} \end{aligned}$$

and the formula for the integral becomes

$$\int_{\mathcal{H}_N} e^{-t\,\mathrm{tr}\,H^4} d\mu(H) = c_N N! R_1^{N-1} R_2^{N-2} \dots R_{N-1} h_0^N. \tag{3.16}$$

For two functions P, Q integration by parts gives

$$(P, Q')_V = \frac{1}{\sqrt{2\pi}} \int P Q' e^{-\frac{1}{2}\lambda^2 - V(\lambda)} = -(P', Q)_V + (P, Q(\lambda) + V'(\lambda))_V.$$

In particular,

$$\begin{aligned}
kh_{V,k-1} &= (P_{V,k-1}, P'_{V,k})_V \\
&= -(P'_{V,k-1}, P_{V,k})_V + (P_{V,k-1}, P_{V,k}(\lambda + 4t\lambda^3))_V \\
&= (\lambda P_{V,k-1}, P_{V,k})_V + 4t(\lambda P_{V,k-1}, \lambda^2 P_{V,k})_V \\
&= h_{V,k} + 4t(P_{V,k} + R_{V,k-1}P_{V,k-2}, \lambda^2 P_{V,k})_V \\
&= h_{V,k} + 4t[(\lambda P_{V,k}, \lambda P_{V,k})_V + R_{V,k-1}(\lambda^2 P_{V,k-2}, P_{V,k})_V] \\
&= h_{V,k} + 4t(h_{V,k+1} + R^2_{V,k}h_{V,k-1} + R_{V,k-1}h_{V,k}) \\
&= h_{V,k-1}(R_{V,k} + 4tR_{V,k}(R_{V,k+1} + R_{V,k} + R_{V,k-1}))
\end{aligned}$$

and finally we obtain

$$k = R_{V,k} + 4tR_{V_k}(R_{V,k+1} + R_{V,k} + R_{V,k-1}). \tag{3.17}$$

3.5.3 Planar approximations

The coefficients of the expansion in t of $\log \int_{\mathcal{H}_N} e^{-t\,\mathrm{tr}\,H^4} d\mu(H)$ are polynomials in N. We know that the leading coefficients describe planar connected gluings. These leading coefficients correspond to the power N^{n+2}. Indeed, a connected embedded graph with n vertices and $2n$ edges is planar iff the number of its faces equals $n+2$.

Planar gluings of 4-stars are thus enumerated by the generating function

$$e_0(t) = \lim_{N\to\infty} \frac{1}{N^2} \log \int_{H_N} e^{-\frac{t}{N}\,\mathrm{tr}\,H^4} d\mu(H).$$

Divide both parts of the equation (3.17) by N:

$$\frac{k}{N} = \frac{R_{V,k}}{N}\left(1 + 4t\left(\frac{R_{V,k+1}}{N} + \frac{R_{V,k}}{N} + \frac{R_{V,k-1}}{N}\right)\right).$$

As N tends to infinity for $\frac{k}{N} = x \in [0,1]$ fixed we obtain

$$x = r_V(1 + 12tr_V),$$

where $r_V = r_V(x,t)$ is the limit function $r_V(x,t) = \lim_{N\to\infty} \frac{R_{xN}(t)}{N}$.

On the other hand, taking logarithm of both parts of Equation (3.16) yields

$$-\frac{1}{N}\sum_{k=1}^{N}\left(1-\frac{k}{N}\right)\ln\frac{R_k(t)}{R_k(0)}-\ln\frac{h_0(t)}{h_0(0)}$$

and the limit as $N\to\infty$ is

$$e_0(t)=-\int_0^1 dx(1-x)\ln\left(\frac{r_V(x)}{x}\right).$$

The integration gives

$$\begin{aligned}
e_0(t) &= -\int_0^1 dx(1-x)\ln\left(\frac{r_V(x)}{x}\right)\\
&= \int_0^{\alpha}(1-r_V-12tr_V)(1+24tr_V)\ln(1+12tr_V)dr_V\\
&= -\frac{1}{2}(-12t\alpha^3+\alpha^2-2\alpha)(1+12t\alpha)\ln(1+12t\alpha)+(18t^2\alpha^4-2t\alpha^3+\frac{3}{2}\alpha)\\
&= -\frac{1}{2}\ln\alpha+\frac{1}{8}(1-\alpha)(\alpha-9)
\end{aligned}$$

Substituting the value of α in the last formula yields

$$e_0(t)=\sum_{k=0}^{\infty}(12)^k\frac{(2k-1)!}{(k+2)!}\frac{t^k}{k!}.$$

3.6 Enumeration of colored graphs

3.6.1 Gauss problem

Consider an immersion $u: S^1\to\mathbf{R}^2$ of the circle into the plane, $u'(x)\neq 0\quad\forall x\in S^1$. Suppose also that the image $u(S^1)$ has only double points of intersection and the intersection in each double point is transversal (i.e., the tangent vectors to branches of the image are not collinear), see Figure 3.10.

Following V. I. Arnold we call the image of such an immersion a *unicursal curve*.

Let n be the number of double points in a unicursal curve.

Table 3.2 borrowed from [1] contains the numbers of plane and spherical unicursal curves with given number n of the self-intersection points. Note that these numbers are obtained by direct drawing of all the curves.

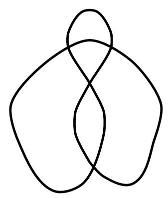


Figure 3.10: A unicursal curve

n	0	1	2	3	4	5
$\mathbf{R}^2$	1	2	5	21	102	639
S^2	1	1	2	6	21	99

Table 3.2: The number of unicursal plane and spherical curves

Considered up to an isotopy of the plane, there are 2 curves with one double point, 5 curves with two double points and 21 curves with three double points.

Note, that considered as curves on the sphere there is 1 curve with one double point, 2 curves with two double points and 6 curves with three double points.

The Gauss problem is:

How many are there unicursal curves (up to an isotopy) with given number of double points.

This problem is unresolved yet. We are going to write out a Gaussian integral describing the number of isotopy classes of unicursal curves on the sphere.

Consider n 4-stars on the plane. We know already how to select planar gluings of the needles. What we need is to select unicursal gluings. In order to do this color each pair of opposite needles in one of c colors. We consider further only *admissible* gluings, i.e., those that glue only needles of one and the same color.

Each gluing provides a curve with some number k of "components". Uni-

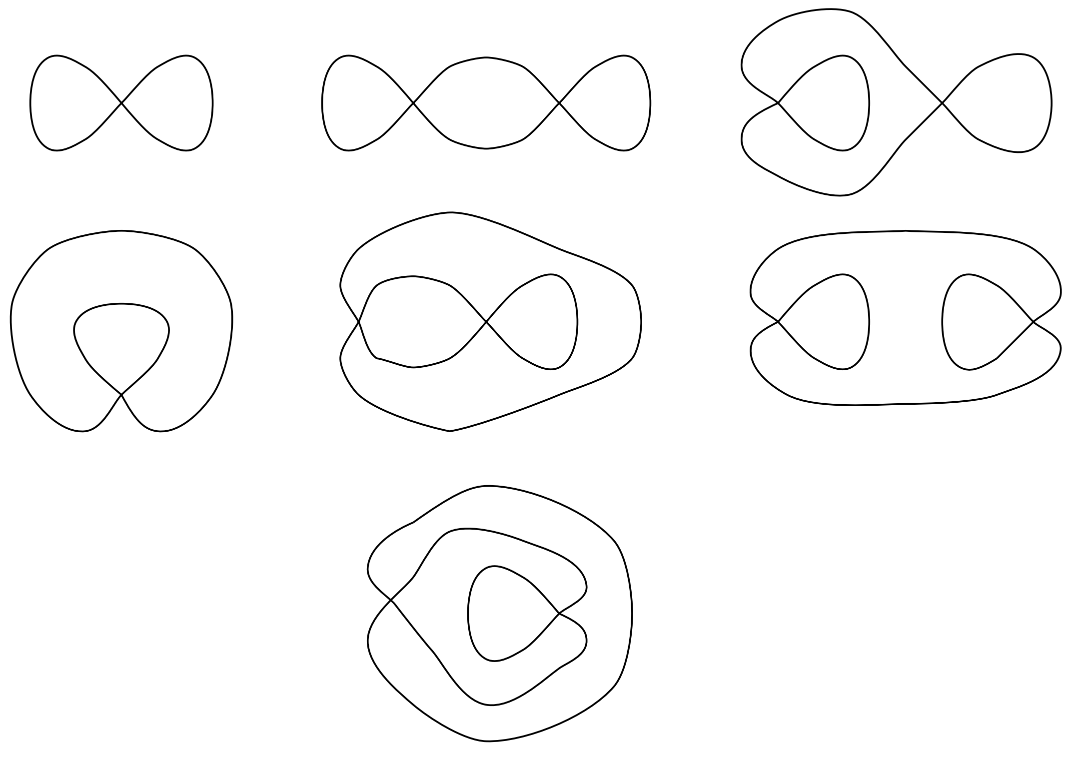


Figure 3.11: Plane unicursal curves with one and two points of selfintersection

cursal curves are precisely those that consist of one "component" only. A k-"component" curve admits c^k admissible colorings.

In order to involve colorings in the integral, consider the integrand in the form

$$\sum_{i,j=1}^{c} \mathrm{tr}(H_i H_j H_i H_j)^n.$$

Integrating this expression over all spaces $\mathcal{H}_{N,i}, \quad i = 1, \ldots, c$ we obtain a function, which will be a polynomial both in N and in c. The coefficient at the monomial linear in c and having the leading power of N is precisely the number of all planar (i.e., spherical) unicursal gluings. Note that a unicursal curve is automatically connected.

Some of these gluings provide one and the same unicursal planar curve. Each unicursal curve is represented by precisely $n!4^n$ gluings. Indeed, in each gluing we are free in exchanging 4-stars and in rotating each star.

We have proved thus the following theorem.

Theorem 3.6.1 *The number of spherical unicursal curves with n double points is equal to the coefficient at the monomial $(-1)^n cN^{n+1}s^n$ in the expansion of*

$$\int_{\mathcal{H}_N} \dots \int_{\mathcal{H}_N} e^{-\frac{s}{4}\operatorname{tr}\sum_{i,j=1}^{c} H_i H_j H_i H_j} d\mu(H_1) \dots d\mu(H_c).$$

3.6.2 Meanders

Let us fix a straight line on the plane and $2n$ points on it. Consider a simple connected not self–intersecting closed curve intersecting the line in exactly those points. The equivalence class of such curves with respect to isotopies of the plane leaving the line fixed is called a *meander* of order n.

In Figure 3.12 all the meanders of order $1, 2, 3$ are shown.

Considerations similar to that in the previous subsection immediately give

Theorem 3.6.2 *The number of meanders of order n is equal to the coefficient at the monomial $4nc^2 N^{2(n+1)} s^{2n}$ in the expansion of*

$$\int_{\mathcal{H}_N} \dots \int_{\mathcal{H}_N} \int_{\mathcal{G}_N} \dots \int_{\mathcal{G}_N} e^{-\frac{s}{2}\operatorname{tr}\sum_{i,j=1}^{c} H_i G_j H_i G_j} d\mu(H_1) \dots d\mu(H_c) d\mu(G_1) \dots d\mu(G_c).$$

3.7 The Kontsevich model

3.7.1 The integral of the model

Let Λ be a diagonal $N \times N$-matrix with positive entries $\Lambda_1, \dots, \Lambda_N$ on the diagonal. The *Kontsevich model* associates to the matrix Λ a new measure on the space of Hermitian matrices. This measure has the form

$$d\mu_\Lambda(H) = c_{\Lambda,N} e^{-\frac{1}{2}\operatorname{tr} H^2 \Lambda} dv(H). \tag{3.18}$$

In fact, this measure is a set of measures parametrized with the diagonal entries of Λ.

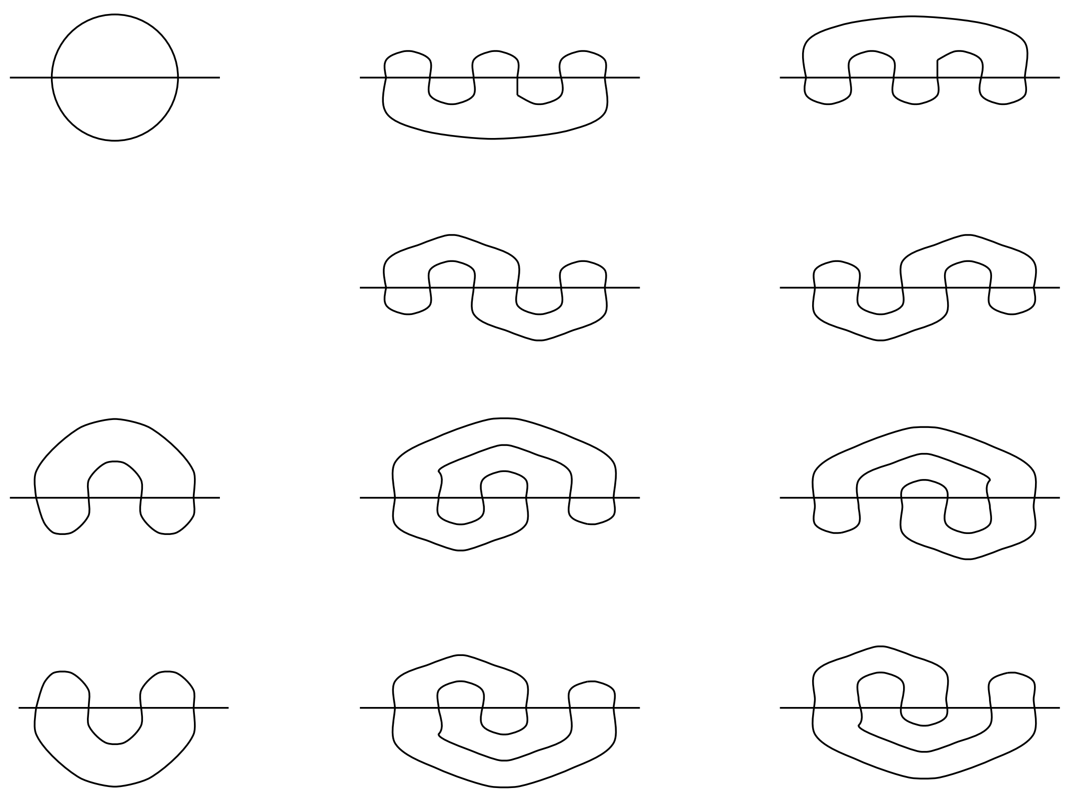


Figure 3.12: Meanders of order 1,2,3

The *integral of the Kontsevich model* is the following integral over the space of Hermitian matrices:

$$\log \int_{\mathcal{H}_N} e^{\frac{\sqrt{-1}}{6}\operatorname{tr} H^3} d\mu_\Lambda(H). \tag{3.19}$$

This integral is a function of $\Lambda_1, \ldots, \Lambda_N$, but it will be more convenient for us to understand this integral as a function in a different set of variables.

We are already familiar with essential parts of integral (3.19), at least for the identity matrix Λ. The expansion of the exponent in the integrand over powers of tr H^2 looks like

$$1 + \frac{1}{1!}\frac{\sqrt{-1}}{6}\operatorname{tr} H^3 - \frac{1}{2!}\frac{1}{6^2}(\operatorname{tr} H^3)^2 - \frac{1}{3!}\frac{\sqrt{-1}}{6^3}(\operatorname{tr} H^3)^3 \ldots,$$

and we conclude that this integral enumerates gluings of 3-stars (up to a nonessential numerical normalizing factor). Note that the number of ways to glue an odd number of 3-stars is zero, hence the imaginary part of the expansion of (3.19) vanishes. Taking logarithm leads to enumeration of *connected* gluings. Thus all the innovations are due to the changed measure.

The changing measure results in appearing of the *weight* of a gluing, and therefore integral (3.19) enumerates weighted gluings, not ordinary ones.

In order to compute this weight, let us find the average $\langle h_{ij}h_{kl}\rangle$ of a monomial $h_{ij}h_{kl}$ of degree two with respect to the new measure.

In coordinates $x_{ii}, x_{ij}, y_{ij},\ 1 \le i < j \le N$ the quadratic form $\operatorname{tr} H^2\Lambda$ looks like

$$\begin{pmatrix} \Lambda_1 & & & & & & \\ & \Lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \Lambda_N & & & \\ & & & & \Lambda_1+\Lambda_2 & & \\ & & & & & \ddots & \\ & & & & & & \Lambda_{N-1}+\Lambda_N \end{pmatrix}.$$

This form is diagonal and we easily compute

$$\langle x_{ii}^2\rangle = \langle h_{ii}^2\rangle = \frac{1}{\Lambda_i},$$

$$\langle x_{ij}^2\rangle = \langle y_{ij}^2\rangle = \frac{1}{\Lambda_i+\Lambda_j} \text{ for } i<j.$$

The result of this computation can be unified in the formula

$$\langle h_{ij}h_{ji}\rangle = \frac{2}{\Lambda_i+\Lambda_j} \tag{3.20}$$

and

$$\langle h_{ij}h_{kl}\rangle = 0 \text{ if } (i,j) \ne (l,k).$$

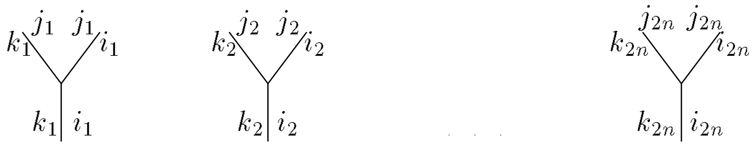


Figure 3.13: Indexed 3-stars

Consider a set of 3-stars with needles marked with pairs of indices, as usual (see Figure 3.13). And consider an arbitrary gluing of these needles. Let us name such an object an *indexed gluing* for the time being. For the standard Gaussian measure the contribution of an indexed gluing depends on the indexing. However the dependence is not too serious: the contribution of an indexed gluing equals 1 if the indexing "agrees" with the gluing, and it vanishes otherwise. The word "agrees" means that all edge sides belonging to one and the same face have the same index. Thus an indexing, which agree with a gluing, is simply an indexing of the faces of the gluing. In order to compute the contribution of a gluing itself we simply summed "ones" over all indexing that agree with the gluing.

For the new measure the situation is different. The contribution of an indexed gluing, where the indexing does not agree with the gluing, is zero as well. However, if the indexing agrees with the gluing, the contribution depends on this indexing.

Consider an indexed gluing with indexing in agree with the gluing. In this case an edge of the gluing corresponds to a pair $\langle h_{ij}h_{ji}\rangle$, where i and j are the indices of faces adjacent to this edge. If an edge has one and the same face with index i on both sides, then it corresponds to the average $\langle h_{ii}^2\rangle$.

Thus, the contribution of an indexed gluing is the product $\prod 2/(\Lambda_i+\Lambda_j)$, taken over all edges of the gluing, and the entry of the product is $2/(\Lambda_i+\Lambda_j)$ iff the corresponding edge is adjacent to the faces i and j.

For example, the contributions of indexed gluings on Figure 3.14 a), b), c) are

$$\frac{2}{\Lambda_i+\Lambda_j}\frac{2}{\Lambda_j+\Lambda_k}\frac{2}{\Lambda_k+\Lambda_i}$$

$$\frac{1}{\Lambda_i}\frac{2}{\Lambda_i+\Lambda_j}\frac{2}{\Lambda_i+\Lambda_k}$$

$$\frac{1}{\Lambda_i}\frac{1}{\Lambda_i}\frac{1}{\Lambda_i}$$

correspondingly.

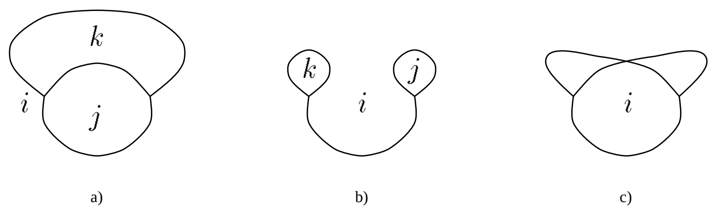


Figure 3.14: Indexed gluings of a pair of 3-stars

The contribution of a gluing is the sum of contributions of all its possible indexings. However, it is more interesting to consider not the contribution of a separate gluing, but the contribution of a class of gluings that provide the surface of the same genus.

3.7.2 Computations

Example 3.7.1 There exist 15 gluings of two 3-stars. Twelve of them give the sphere, while the other three results to the torus. Those giving the sphere are of the type shown on Figure 3.14 a) and b). There are three gluings of the type a) and six gluings of the type b). Fixing the set of indices $\{i, j, k\}$ for the faces and summing over all permutation of this set of indices we obtain the total contribution of all spherical gluings

$$\begin{aligned} 3\cdot\frac{6\cdot 8}{(\Lambda_i+\Lambda_j)(\Lambda_j+\Lambda_k)(\Lambda_k+\Lambda_i)} \quad &+9\cdot\left(\frac{2\cdot 8}{2\Lambda_i(\Lambda_i+\Lambda_j)(\Lambda_i+\Lambda_k)}\right. \\ &+\frac{2\cdot 8}{2\Lambda_j(\Lambda_j+\Lambda_i)(\Lambda_j+\Lambda_k)} \\ &+\left.\frac{2\cdot 8}{2\Lambda_k(\Lambda_k+\Lambda_i)(\Lambda_k+\Lambda_j)}\right) \\ &=\frac{72}{\Lambda_i\Lambda_j\Lambda_k} \qquad (3.21) \end{aligned}$$

Summing over all possible values of the indices $i, j, k = 1, \dots, N$ and taking into account the normalizing factor $\frac{1}{2!}\frac{1}{6^2}$ we obtain the total contribution of all spherical gluings to integral (3.19):

$$c_{0,1} = \frac{1}{2!}\frac{1}{6^2}\sum_{i,j,k=1}^{N} \frac{72}{\Lambda_i \Lambda_j \Lambda_k} = (\operatorname{tr}\Lambda^{-1})^3.$$

Here $c_{g,n}$ stands for "the contribution of all genus g gluings of $2n$ 3-stars".

Recall for conveniency the meaning of the coefficients in the left hand side of (3.21). In the first term the coefficient 3 reflects three different ways to obtain a gluing of type Figure 3.14 a). The coefficient 6 corresponds to six different ways of indexing the faces of such a gluing with indices i, j, k. And the factor $8/(\Lambda_i + \Lambda_j)(\Lambda_j + \Lambda_k)(\Lambda_k + \Lambda_i)$ is the contribution of the corresponding indexed gluing.

In the second term the coefficient 9 reflects nine different ways to obtain a gluing of type Figure 3.14 b). Each of the three terms in brackets corresponds to a pair of indexed gluings (all in all, there are six possibilities, similarly to the preceding case). Two gluings in a pair differ by indices, but the contribution of each gluing in a pair is the same. The rest factor is the contribution of the corresponding indexed gluing.

Example 3.7.2 The gluing on Figure 3.14 c) is much more easy to handle. It is the only gluing type for two 3-stars providing the torus, and there is precisely three gluings of this type. This gluing has only one face, and marking this face with the index i we obtain the resulting contribution $3/(\Lambda_i^3)$. Thus, summing over all possible indexing,

$$c_{1,1} = \frac{1}{2!}\frac{1}{6^2}\sum_{i=1^N} \frac{3}{\Lambda_i^3} = \frac{1}{24}\operatorname{tr}(\Lambda^{-3}).$$

Example 3.7.3 The computational difficulty grows very rapidly and we restrict ourselves with presenting only one more example, which we do not describe in details. Figure **??** shows all possible types of spherical gluings of four 3-stars. Note that the word "spherical" means here not only connected spherical gluings, but also disconnected gluings with four faces. For each gluing type the number of gluings of this type is shown as well as a sample indexing of faces and the contribution of the corresponding index gluing.

Summing up over all indexed gluings with the set of indices $\{i, j, k, l\}$ gives the total contribution

$$\frac{2025}{4} \frac{\Lambda_i^2 \Lambda_j^2 \Lambda_k^2 + \Lambda_i^2 \Lambda_j^2 \Lambda_l^2 + \Lambda_i^2 \Lambda_k^2 \Lambda_l^2 + \Lambda_j^2 \Lambda_k^2 \Lambda_l^2}{\Lambda_i^3 \Lambda_j^3 \Lambda_k^3 \Lambda_l^3},$$

and summing over all indices $i, j, k, l = 1, \ldots, N$ taking into account the normalizing factor $\frac{1}{4!}\frac{1}{6^4}$ we obtain

$$c_{0,2} = \frac{25}{1536} \operatorname{tr}(\Lambda^{-1}) \operatorname{tr}(\Lambda^{-3}).$$

Exercise 3.7.4 Find $c_{-1,2}$.

Answer. $\frac{1}{288} \operatorname{tr}(\Lambda^{-1})^2$.

3.7.3 Properties of the Kontsevich model

In this section we describe very briefly exciting connections of the Kontsevich model with other models and theories. This model was invented as a bridge between two other models of two-dimensional quantum gravity. Our brevity can be partially explained by a huge amount of facts and related topics involved in these models, and probably another book would prove insufficient to give a full description of all of them.

Sample calculations in the previous section show that the contribution $c_{g,n}$ of all genus g gluings of $2n$ 3-stars is a monomial in $\operatorname{tr}(\Lambda^{-1}), \operatorname{tr}(\Lambda^{-3}), \ldots$. A proof of this statement was the main combinatorial difficulty in the proof of Kontsevich theorem. It is more convenient, however, to use a slightly different normalized infinite sequence of independent variables: $t_0 = \operatorname{tr}(\Lambda^{-1})/1!!, t_1 = \operatorname{tr}(\Lambda^{-3})/3!!, \ldots, t_i = \operatorname{tr}(\Lambda^{-(2i+1)}/(2i+1)!!, \ldots$

Theorem 3.7.5 [16] *Integral (3.19) is a power series in variables $t_0, t_1, \ldots$ with rational coefficients.*

Let us denote this series by $K(t_0, t_1, \ldots)$. We have computed already few first terms of this series:

$$K(t_0, t_1, \ldots) = \log(1 + t_0^3 + \frac{1}{24} t_1 + \frac{25}{1536} t_0 t_1 + \frac{1}{288} t_1^2 + \ldots) \qquad (3.22)$$

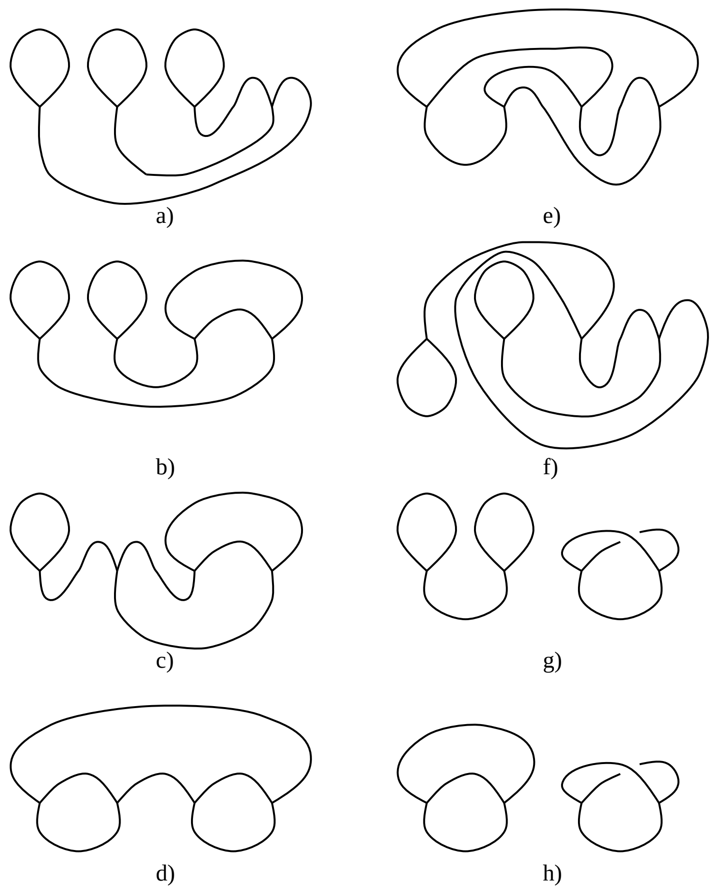


Figure 3.15: Types of four 3-stars gluings with Euler characteristic 2. The number of gluings and a sample weight of a colored gluing for each type are:
a) $\#=3^4\cdot 2^3$; $w=(8\lambda_1^3(\lambda_1+\lambda_2)(\lambda_1+\lambda_3)(\lambda_1+\lambda_4))^{-1}$
b) $\#=3^5\cdot 2^2$; $w=(4\lambda_1^2(\lambda_1+\lambda_2)^2(\lambda_1+\lambda_3)(\lambda_1+\lambda_4))^{-1}$
c) $\#=3^5\cdot 2^3$; $w=(2\lambda_1(\lambda_1+\lambda_2)^2(\lambda_1+\lambda_3)(\lambda_1+\lambda_4)(\lambda_2+\lambda_3)^2)^{-1}$
d) $\#=3^5\cdot 2$; $w=((\lambda_1+\lambda_2)(\lambda_1+\lambda_3)(\lambda_1+\lambda_4)^2(\lambda_2+\lambda_4)(\lambda_3+\lambda_4))^{-1}$
e) $\#=3^4\cdot 2$; $w=((\lambda_1+\lambda_2)(\lambda_1+\lambda_3)(\lambda_1+\lambda_4)(\lambda_2+\lambda_3)(\lambda_2+\lambda_4)(\lambda_3+\lambda_4))^{-1}$
f) $\#=3^5\cdot 2^2$; $w=(4\lambda_1\lambda_3(\lambda_1+\lambda_2)(\lambda_1+\lambda_3)^2(\lambda_3+\lambda_4))^{-1}$
g) $\#=3^4\cdot 2$; $w=(16\lambda_1\lambda_4^3(\lambda_1+\lambda_2)(\lambda_1+\lambda_3))^{-1}$
h) $\#=3^3\cdot 2$; $w=(8\lambda_4^3(\lambda_1+\lambda_2)(\lambda_2+\lambda_3)(\lambda_1+\lambda_3))^{-1}$
Total: $2025(\lambda_1^2\lambda_2^2\lambda_3^2+\lambda_1^2\lambda_2^2\lambda_4^2+\lambda_1^2\lambda_3^2\lambda_4^2+\lambda_2^2\lambda_3^2\lambda_4^2)/(4\lambda_1^3\lambda_2^3\lambda_3^3\lambda_4^3)$

The first one of the two other models of the two-dimensional quantum gravity is the Universal one-matrix model explained in Section 3.4.3. It is described by another power series in infinite set of variables $t_0, t_1, \dots$. The statement of the first part of Kontsevich theorem is now easy.

Theorem 3.7.6 [16] *The power series describing the universal one-matrix model and the Kontsevich model coincide.*

The proof, however, is not that easy. It is based on the observation that both power series satisfies the famous "Korteweg–de Vries hierarchy of partial differential equations". This hierarchy has been extensively studied since late seventies. It is known that solutions of this hierarchy with coinciding "initial conditions" coincide. Coincidence of the initial conditions was easy to verify. The integral of the universal model was also known to satisfy the "Korteweg–de Vries hierarchy". The proof of the same statement for the Kontsevich model made use of its understanding as a multidimensional analog of the classical Airy function.

The other two-dimensional quantum gravity model is related to the topology of the "moduli space of complex curves" (this is the space of equivalence classes of complex curves modulo isomorphisms). In this case it is also possible to describe the model as a power series in infinite number of indeterminates. The coefficients of the series are "intersection numbers" of some cohomology classes on this space. The connection to this model was established through a construction of the "combinatorial model of the moduli space". The last construction is based on an observation by Mamford, and its variant was heavily used by Penner.

This construction is extremely interesting by its connection with embedded graphs, and we would like to describe it in few more words.

The base of this construction is a totally unexpected statement::

There exists a natural correspondence between embedded graphs with metric and complex structures on curves.

To be more precise, an *embedded graph with a metric* is just an embedded graph with given lengths of all its edges. Let us denote by $\mathcal{M}_{g,n}^{\mathrm{comb}}$ the space of all the graphs of genus g and with n faces supplied with a metric. This space carries natural topology.

On the other hand, consider the compactified moduli space $\overline{\mathcal{M}}_{g,n}$ of complex structures on curves of genus g with n marked points. This space carries a natural topology (in fact, one can show that it is compact) and a natural structure of a complex space with singularities. The complex dimension of this space is $3g - 3 + n$.

Associate with each marked point of a curve a positive real number r_i and consider such an object as a point in the topological space $\overline{\mathcal{M}}_{g,n} \times \mathbf{R}_+^n$. The precise statement looks like the following.

Theorem 3.7.7 *There exists a natural homeomorphism between the spaces* $\overline{\mathcal{M}}_{g,n} \times \mathbf{R}_+^n$ *and* $\mathcal{M}_{g,n}^{\mathrm{comb}}$.

This natural homeomorphism is established by means of the *Jenkins–Strebel quadratic differentials*.

3.8 Appendix: Logarithm of a formal power series

In this Appendix we present an alternative proof of the Wick formula based on a formula for the logarithm of a power series.

The Wick formula is a very particular case of a formula well known in the probability theory, which expresses the moments of a probability distribution in terms of its semi-invariants. The latter one, in its turn, is a "probabilistic interpretation" of the formula of logarithm of a power series. For more details see, for example, [31], Chapter 2 "Semi-invariants and combinatorics".

We use the multi-index notation:

$$\begin{aligned}
\alpha &:= (\alpha_1, \ldots, \alpha_n), \quad \alpha_i \in \mathbf{N}, \\
|\alpha| &:= \alpha_1 + \ldots + \alpha_n, \\
\alpha! &:= \alpha_1! \times \ldots \times \alpha_n!, \\
t^\alpha &:= t_1^{\alpha_1} \ldots t_n^{\alpha_n}, \quad \text{where} \quad t = (t_1, \ldots, t_n), \\
\frac{\partial^{|\alpha|} f}{\partial t^\alpha} &:= \frac{\partial^{|\alpha|} f}{\partial t_1^{\alpha_1} \ldots \partial t_n^{\alpha_n}}.
\end{aligned}$$

Let

$$f(t) = 1 + \sum_{|\alpha|>0} \frac{m_\alpha}{\alpha!} t^\alpha$$

be the Taylor series for a function $f(t)$ satisfying $f(0) = 1$; here

$$m_\alpha = \left.\frac{\partial^{|\alpha|} f}{\partial t^\alpha}\right|_{t=0}.$$

Let the following be the Taylor series for $\log f(t)$:

$$\log f(t) = \sum_{|\alpha|>0} \frac{s_\alpha}{\alpha!} t^\alpha.$$

The formula below expresses the coefficients m_α in terms of s_α.

Proposition 3.8.1 *The Taylor coefficients m_α of a function f are expressed in terms of the Taylor coefficients s_α of $\log f$ by the following formula:*

$$m_\alpha = \sum_{k=1}^{|\alpha|} \frac{1}{k!} \sum_{\substack{\beta^1 + \ldots + \beta^k = \alpha, \\ |\beta^i| > 0}} \frac{\alpha!}{\beta^1! \ldots \beta^k!} \prod_{i=1}^{k} s_{\beta^i}.$$

Here β^i are also multi-indices; the interior sum is taken over all ordered *k-tuples of multi-indices* (but the factor $1/k!$ "kills the order").

Remark 3.8.2 The above formula is valid for formal power series as well as for analytical functions.

Proposition 3.8.3 *Let μ be a probability distribution in $\mathbf{R}^n$* (an arbitrary one, not necessarily Gaussian)*, and let*

$$\varphi(t) = <e^{i(t,x)}> = \int_{\mathbf{R}^n} e^{i(t,x)} d\mu(x)$$

be its characteristic function. If φ is analytic in a neighborhood of 0, then all the moments $m_\alpha := <x^\alpha>$ exist, and

$$\varphi(t) = 1 + \sum_{|\alpha|>0} \frac{m_\alpha}{\alpha!} (it)^\alpha.$$

The coefficients s_α of the series

$$\log \varphi(t) = \sum_{|\alpha|>0} \frac{s_\alpha}{\alpha!}(it)^\alpha$$

are called *semi-invariants* of the distribution μ, so the formula of Proposition 3.8.1 expresses the moments of a probability distribution in terms of its semi-invariants.

The formulas and computations involving semi-invariants are often much simpler than those involving moments. A Gaussian distribution is a striking example of this phenomenon, since for the characteristic function (3.5), its logarithm is no more an infinite series but a quadratic polynomial

$$\log \varphi(t) = -\frac{1}{2}(Ct, t),$$

so $s_\alpha \neq 0$ implies $|\alpha| = 2$, and for such an α we have $s_\alpha = m_\alpha = c_{ij}$ (here $\alpha_i = \alpha_j = 1$ if $i \neq j$, and $\alpha_i = 2$ if $i = j$, while all the other components of the multi-index α are equal to 0).

The proof of the Wick formula requires one more technical result.

Lemma 3.8.4 *Let μ be a Gaussian measure in $\mathbf{R}^n$, and let $A : \mathbf{R}^n \to \mathbf{R}^k$ be a linear operator. Then A induces a measure ν in $\mathbf{R}^k$, which is also Gaussian.*

Proof. Let C be the covariance matrix of μ, and let $x, t \in \mathbf{R}^n$, $y, s \in \mathbf{R}^k$, and $y = Ax$. Then

$$\varphi_\nu(s) := <e^{i(s,y)}> = <e^{i(s,Ax)}> = <e^{i(A^*s,x)}> = \varphi_\mu(A^*s),$$

where $A^* : \mathbf{R}^k \to \mathbf{R}^n$ is the operator adjoint to A. What remains is to substitute A^*s for t in $\varphi_\mu(t) = \exp\{-\frac{1}{2}(Ct,t)\}$, which gives the covariance matrix of ν equal to ACA^*.

Remark 3.8.5 In the above lemma, the dimension k may be less than n, equal to n or greater than n, and the operator A may be of an arbitrary rank.

Return now to the Wick formula. Let $y_1 = f_1(x), \ldots, y_{2k} = f_{2k}(x)$. According to Lemma 3.8.4 the vector $(y_1, \ldots, y_{2k})$ has a Gaussian distribution, and our goal is to compute, for this distribution, the moment $<y_1 \dots y_{2k}>$ which corresponds to the multi-index $\alpha = (1, 1, \dots, 1)$. Now we can apply the formula of Proposition 3.8.1 and split α into sums of multi-indices β^i, of which survive only those of the form $(0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$, having only two non-zero entries.

The Wick formula is proved.

3.9 Historical and bibliogrphical remarks

The idea of applying matrix integrals to enumeration of embedded graphs was first implemented in [4]. In [20] it was rediscovered for the purpose of solving the enumeration problem for polygon gluings. Note that A. Lehman investigated this problem from the purely combinatorial point of view and obtained partial results in early 70's (unpublished).

The equivalence of two 2-dimensional gravity models was conjectured by E. Witten in [39]. The conjecture was based on explicit calculations of the first few terms of both series, as well on the ensurance that "there is only one gravity". It was proved shortly after by M. Kontsevich [16]. However, essential combinatorial parts of the proof still seem to be rather complicated and probably require a different approach.

The "coloring approach" to enumeration of "fixed-cursal" embedded graphs was suggested by V. Kazakov [30].

Bibliography

[1] V. I. Arnold *Plane Curves: Their Invariants, Perestroikas and Classifications*, in: Advances in Soviet Mathematics, vol. 21, 1994, 33–91

[2] **Bauer M., Itzykson C.** Triangulations. – In [34], p. 179–236.

[3] **Bessis D.** A new method in the combinatorics of the topological expansion. – *Comm. Math. Phys.*, 1979, vol. **69**, 147–163.

[4] D. Bessis, C. Itzykson, J. B. Zuber *Quantum Field Theory Techniques in Graphical Enumerations*, Adv. Appl. Math., vol. 1 (1980), pp.109–157

[5] **Boulatov D. V., Kazakov V. A., Kostov I. K., Migdal A. A.** Analytical and numerical study of a model of dynamically triangulated random surfaces. – *Nuclear Physics B*, 1986, vol. **275** [FS17], 641–686.

[6] **Brézin E., Itzykson C., Parisi G., Zuber J.-B.** Planar diagrams. – *Commun. Math. Phys.*, 1978, vol. **59**, 35–51.

[7] **Brézin E., Itzykson C., Parisi G., Zuber J.-B.** Planar diagrams. – *Commun. Math. Phys.*, 1978, vol. **59**, 35–51.

[8] **Brézin E., Kazakov V. A.** Exactly solvable field theories of closed strings. – *Physics Letters B*, 1990, vol. **236**, no. 2, 144–150.

[9] **Chadha S., Mahoux G., Mehta M. L.** A method of integration over matrix variables: II. – *J. Phys. A: Math. Gen.*, 1981, vol. **14**, 579–586.

[10] P.Deligne, D. Mamford *The irreducibility of the space of curves of given genus*, Publ. Math I. H. E. S., vol 36 (1969), pp. 75–109

[11] **Di Francesco Ph., Ginsparg P., Zinn-Justin J.** $2D$ gravity and random matrices. – *Phys. Rep.*, 1995, vol. **254**, 1–133.

[12] **Di Francesco Ph., Golinelli O., Guitter E.** Meander, folding and arch statistics. – Preprint, June 1995, 82 pp.

[13] **Di Francesco Ph., Itzykson C.** A generating function for fatgraphs. – Preprint, 1992, 27 pp.

[14] **Di Francesco Ph., Itzykson C., Zuber J.-B.** Polynomial averages in the Kontsevich model. – *Commun. Math. Phys.*, 1993, vol. **151**, 193–219.

[15] **Douglas M. R., Shenker S. H.** Strings in less than one dimension. – *Nuclear Physics B*, 1990, vol. **335**, 635–654.

[16] W. Feller *An Introduction to Probability Theory and its application*, vol. II, John Wiley & Sons, Inc., (1971)

[17] I. M. Gelfand, M. A. Naimark *Unitarnye predstavleniya klassicheskikh grupp*, [in Russian], Izd. AN SSSR, Moskva-Leningrad, 1950

[18] M. B. Green, J. H. Schwarz, E. Witten, *Superstring Theory*, Vol. 1. Introduction, Vol. 2. Loop amplitudes, anomalies and phenomenology. Cambridge University Press, Cambridge–New York, 1988.

[19] **Gross D. J., Migdal A. A.** A nonperturbative treatment of two-dimensional quantum gravity. – *Nuclear Physics B*, 1990, vol. **340**, 333–365.

[20] J. Harer, D. Zagier *The Euler Characteristic of the Moduli Space of Curves*, Inv. Math., vol. 85 (1986),pp. 457–485

[21] **Harary F., Palmer E. M.** Graphical Enumeration. – Academic Press, New York and London, 1973.

[22] **'t Hooft G.** A planar diagram theory for strong interactions. – *Nuclear Physics B*, 1974, vol. **72**, 461–473.

[23] **Itzykson C., Zuber J.-B.** The planar approximation. II. – *J. Math. Phys.*, 1980, vol. **21**, no. 3, 411–421.

[24] **Itzykson C., Zuber J.-B.** Matrix integration and combinatorics of modular groups. – *Commun. Math. Phys.*, 1990, vol. **134**, 197–207.

[25] **Itzykson C., Zuber J.-B.** Combinatorics of the modular group II. The Kontsevich integrals. – *Intern. J. of Modern Physics A*, 1992, vol. **7**, no. 23, 5661–5705.

[26] **Jackson D. M.** Counting cycles in permutations by group characters, with an application to a topological problem. – *Trans. Amer. Math. Soc.*, 1987, vol. **299**, no. 2, 785–801.

[27] **Kazakov V. A.** Bosonic strings and string field theories in one-dimensional target space. – Preprint, October 1990, 49 pp.

[28] M. Kontsevich *Intersection Theory on the Moduli Space of Curves and the Matrix Airy Function*, Comm. Math. Phys., vol. 147 (1992), pp. 1–23

[29] **Kostov I. K., Mehta M. L.** Random surfaces of arbitrary genus: exact results for $D = 2$ and -2 dimensions. – *Physics Letters B*, 1987, vol. **189**, no. 1-2, 118–124.

[30] **Lando S. K., Zvonkin A. K.** Meanders. – *Selecta Mathematica Sovietica*, 1992, vol. **11**, no. 2, 117–144.

[31] **Malyshev V. A., Minlos R. A.** Gibbs Random Fields. – Kluwer, Dordrecht, 1991.

[32] M. L. Mehta *Random Matrices*, Academic Press, New York and London, 1967.

[33] **Mehta M. L.** A method of integration over matrix variables. – *Comm. Math. Physics*, 1981, vol. **79**, 327–340.

[34] **Schneps L., ed.** The Grothendieck Theory of Dessins d'Enfants. – London Math. Soc. Lecture Notes Series, vol. 200, Cambridge Univ. Press, 1994.

[35] **Simon B.** The $P(\varphi)_2$ Euclidian (Quantum) Field Theory. – Princeton Univ. Press, 1974.

[36] **Walsh T. R. S., Lehman A. B.** Counting rooted maps by genus. I – *J. of Combinat. Theory B*, 1972, vol. **13**, 192–218.

[37] H. Weyl *Zur Theorie der Darstellung der einfachen kontinuierlichen Gruppen. (Aus einem Schreiben an Herrn I.Schur)*, Sitzungsberichte der Preußischen Akademie der Wissenschaften zu Berlin, 338-345 (1924); republished in: [H. Weyl *Gesammelte Abhadlungen*, Band II, Springer–Verlag, Berlin – Heidelberg – New York, pp. 453–460 (1968)]

[38] H. Weyl *The Classical Groups*, Princeton, 1946.

[39] E. Witten *Two Dimensional Gravity and Intersection Theory on the Moduli Space*, Surveys in Diff. Geom., vol. 1 (1991), pp. 243–310

Chapter 4

Vassiliev knot invariants

4.1 Introduction

4.2 Knot invariants and origins of chord diagrams

4.2.1 Extending knot invariants to singular knots

Consider an embedding of the circle S^1 into $\mathbf{R}^3$, i.e., a smooth mapping $u : S^1 \to \mathbf{R}^3$ without self-intersections and with a nonvanishing tangent vector. *Isotopy* provides a natural equivalence relation for such embeddings.

Two embeddings $u_1, u_2 : S^1 \to \mathbf{R}^3$ are considered to be *isotopy equivalent* if there exists an orientation preserving diffeomorphism $g : \mathbf{R}^3 \to \mathbf{R}^3$ such that the diagram

$$\begin{tikzcd} & \mathbf{R}^3 \arrow[d, "g"] \\ S^1 \arrow[ur, "u_1"] \arrow[dr, "u_2"'] & \mathbf{R}^3 \end{tikzcd}$$

is commutative.

Definition 4.2.1 A *knot* is an isotopy class of embeddings $S^1 \to \mathbf{R}^3$.

Further we will make no distinction between a knot and an embedding $S^1 \to \mathbf{R}^3$ representing this knot.

Definition 4.2.2 A *knot invariant* is just a function on knots.

We usually consider knot invariants with values in an arbitrary commutative associative ring K with unit.

A knot is usually presented by its diagram – a generic plane projection of the knot. At double points of the projection overpasses and underpasses of the two branches are shown, and an orientation of the knot is given.

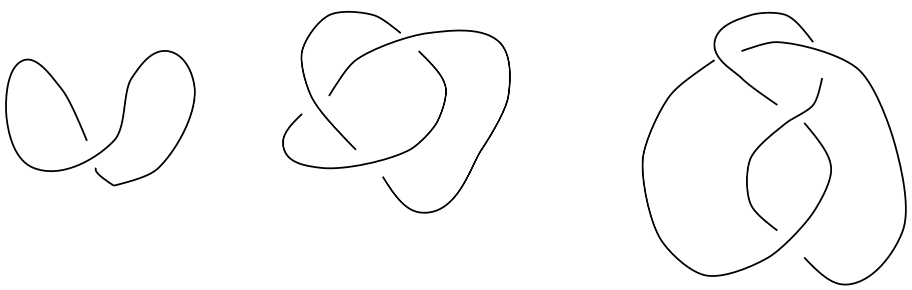


Figure 4.1: Diagrams of various knots

The main problem in the theory of knots is "How to distinguish knots". The problem may have different modifications. One of them is:

"Given two knots, find a knot invariant such that the two knots coincide if and only if the values of the invariant on the knots coincide."

Various types of knot invariants have been introduced since the beginning of the study of knots. Few of them are explained later in this section.

The main idea of constructing Vassiliev knot invariants is to consider "singular knots", not only smooth ones.

Definition 4.2.3 *A singular knot* is (an isotopy class of) mappings $u : S^1 \to \mathbf{R}^3$ such that

1. $u^{'}(t) \neq 0$ for all $t \in S^1$;

2. all points of selfintersection of the image $u(S^1)$ are ordinary double points with transversal selfintersection (i.e., two tangent vectors to the knot at a double point are not collinear).

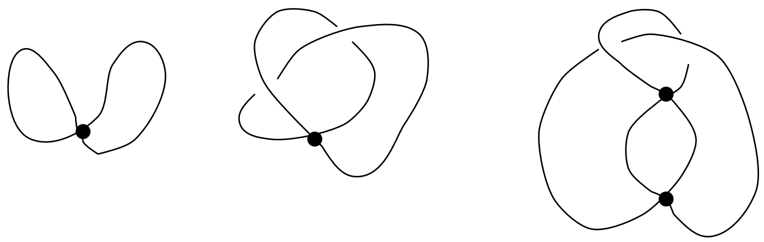
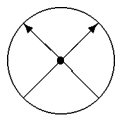


Figure 4.2: Singular knots

Consider a small neighborhood of a double point of a singular knot. In such a neighborhood each singular knot looks like two intersecting segments:

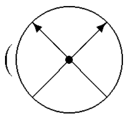
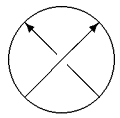
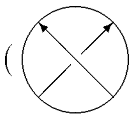


Arrows denote orientation of the knot corresponding to the (counterclockwise) orientation of the initial circle S^1.

A singular point can be "resolved" by a small deformation of the knot in two different directions – the positive and the negative one (see Fig. 4.3).

Figure 4.3: Positive (left) and negative (right) resolutions of a singular point

Theorem 4.2.4 *Let v be a knot invariant. Then it has a (unique) extension to singular knots satisfying the following relation:*

v () $=$ v () $-$ v ()

The three pictures are meant to show three (singular) knots, coinciding everywhere outside a small ball, where they look as on the pictures.

Proof. We will proceed by induction over the number of double points in a singular knot.

Consider a knot with one singular point. The two resolutions of the point are uniquely determined, and we just determine the value of a knot invariant on the singular knot according to the defining rule, which provides the base of induction.

Suppose a knot invariant v is extended to all singular knots with less than n double points. Let us prove that it may be extended to any singular knot u with n double points.

Let $x, y \in \mathbf{R}^3$ be two double points of the singular knot u. We are going to show that the extension of the invariant v onto u does not depend on which of the points x or y is resolved.

Consider four possible ways of resolving both singular points x and y (see Fig. 4.4).

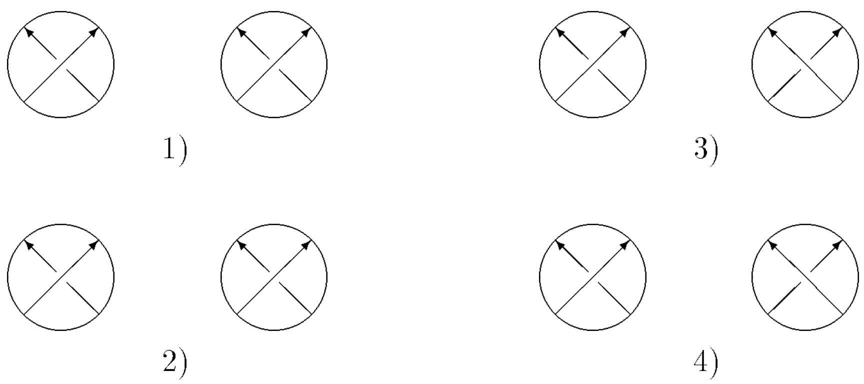


Figure 4.4: Four possible resolutions of two singular points

Then the prolongation of the invariant v to the singular knot u should be equal to $v(1) - v(2) - v(3) + v(4)$ independently on what the order of resolutions is.

The theorem is proved.

4.2.2 Invariants of finite order

We are now able to introduce the main notion of the theory.

Definition 4.2.5 A knot invariant is called *a Vassiliev invariant of order no greater than* n, if it vanishes on all singular knots with more than n double points.

Vassiliev invariants of order no greater than n form a K-module, that we denote by $\mathcal{V}_n$. The union $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1 \cup \mathcal{V}_2 \cup \ldots$ forms a K-module $\mathcal{V}$ of finite order invariants with filtration $\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \ldots \subset \mathcal{V}$.

Example 4.2.6 Constant functions on nonsingular knots provide Vassiliev invariants of order ≤ 0. On the other side, each invariant from $\mathcal{V}_0$ should be a constant on nonsingular knots, i.e. $\mathcal{V}_0 \cong K$.

Indeed, introduce an "elementary transformation" of a knot – passing through a singular knot with an additional double point. Such a transformation does not change the value of any invariant of order ≤ 0, and any knot may be obtained from any other knot by a series of elementary transformations.

Example 4.2.7 Vassiliev invariants of order no greater than 1 also come from constant functions on knots. Indeed, consider a singular knot with one double point. After a series of elementary transformations it may be transformed to a "decomposed" knot with one singular point (see Fig. 4.5).

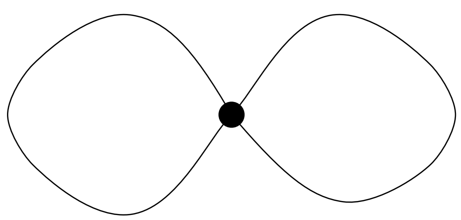


Figure 4.5: A decomposed knot with one singular point

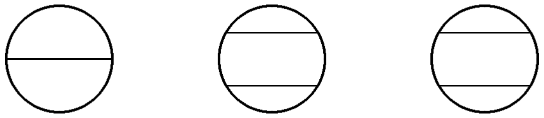


Figure 4.6: Chord diagrams of singular knots from Figure 4.2

These elementary transformations do not change the value of an invariant of order ≤ 1, and for the final knot on the picture it should be equal to zero: both resolutions give one and the same knot. Thus, $\mathcal{V}_1 = \mathcal{V}_0 \cong K$.

Now we are going to associate a chord diagram with any singular knot. A Vassiliev invariant of order no greater than n determines, in fact, a function on chord diagrams with n chords.

Definition 4.2.8 Let $u : S^1 \to \mathbf{R}^3$ be a singular knot with n double points $x_1, \ldots, x_n \in \mathbf{R}^3$. The *chord diagram of the singular knot* u is the circle S^1 with n pairs of points, which are the pre-images $\{u^{-1}(x_1), \ldots, u^{-1}(x_n)\}$, shown as chords.

Obviously, any chord diagram can be realized as a chord diagram of a singular knot.

Theorem 4.2.9 *The value of a Vassiliev invariant of order no greater than n on a singular knot with precisely n double points depends only on the chord diagram of the knot, but not on the knot itself.*

The proof follows from the following

Lemma 4.2.10 *Let two singular knots have the same chord diagram. Then one of them can be obtained from the other one by a series of elementary transformations, i.e., deformations of a knot passing through an additional double point.*

Proof. Indeed, let u_1, u_2 be two singular knots with n double points, which have identical chord diagrams. By an appropriate isotopy the knot u_2 can be transformed so that

a) the corresponding double points of u_2 coincide with those of u_1;

b) the corresponding arcs of u_2 coincide with those of u_1 in a small neighborhood of each double point;

c) the arcs of u_1 and u_2 have no common points outside these small neighborhoods.

Now take two subsequent double points of u_2 and an arc of u_2 connecting these points. This arc can be transformed into the corresponding arc of u_1 by a series of elementary transformations. Due to the defining condition the value of v does not change under such deformations, whence the required assertion.

Consider a singular knot u with n double points. Resolving these double points in all possible ways one obtains 2^n different (nonsingular) knots $u_{\epsilon_1,\dots,\epsilon_n}$. Here $\epsilon_i = 1$ if the i-the double point is resolved in positive direction, $\epsilon_i = -1$ otherwise.

Exercise 4.2.11 Verify that a knot invariant v is of order no greater than $n-1$ if and only if

$$\sum_{(\epsilon_1,\dots,\epsilon_n)} (-1)^{\epsilon_1\dots\epsilon_n} v(u_{\epsilon_1,\dots,\epsilon_n}) = 0$$

for any singular knot u with n double points.

The values of the invariant do not thus depend on the specific behavior of the knotted curve being completely determined by the corresponding chord diagram, i.e., by the order in which double points are encountered while moving along the curve. Moreover, if the restrictions of two invariants of order no greater than n to the set of all singular knots with precisely n double points coincide, then their difference is an invariant of order no greater than $n-1$. This means that an element of $\mathcal{V}_n/\mathcal{V}_{n-1}$ determines a function on chord diagrams of order n, and this mapping is a monomorphism.

As an immediate corollary we obtain the following

Statement 4.2.12 *The linear space $\mathcal{V}_n$ is finite-dimensional for each $n = 0, 1, 2, \ldots$.*

Indeed, the dimension of the quotient space $\mathcal{V}_n/\mathcal{V}_{n-1}$ does not exceed the number of chord diagrams of order n. This dimension is, in fact, essentially smaller, as we will see below.

4.2.3 Deducing 1-term and 4-term relations for invariants

Although each invariant of order $\leq n$ determines a function on chord diagrams, not each function can be obtained in this way. Indeed, we have seen, for example, that each invariant of order ≤ 1 comes from a constant function. It follows that each function on a chord diagram with one chord coming from invariants of order ≤ 1 must vanish.

Thus, there are some restrictions on functions on chord diagrams coming from finite order knot invariants.

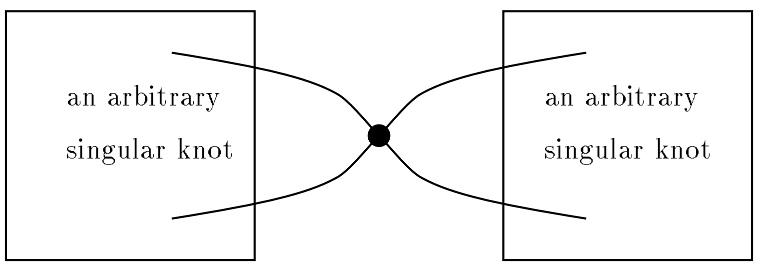
Theorem 4.2.13 *Any function f on chord diagrams of order n coming from a knot invariant of order no greater than n satisfies two types of relations:*

1. *1-term relations: if a chord diagram contains a chord that does not intersect any other chord, then f vanishes on this diagram;*

2. *4-term relations.*

Proof.

1. A series of elementary transformations transforms a singular knot with the chord diagram having a chord, which does not intersect any other chord, into a knot of the following type

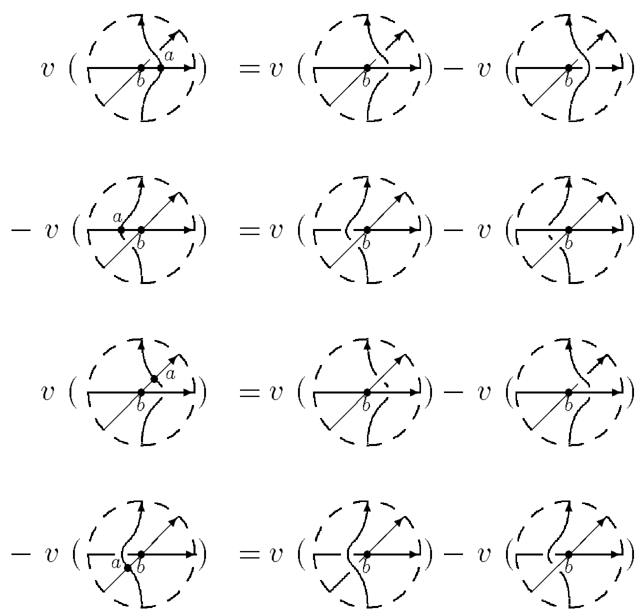


It is obvious that the type of a singular knot resolved in the distinguished point does not depend on the resolution.

2. Consider two chords taking part in a 4-term relation. One of them is a "moving" chord, we denote it by a, the other one is a "stable" one, we denote it by b.

   Each of the chords a and b corresponds to a double point of the knot. Let us draw four pictures of the knots corresponding to four terms of the 4-term relation. The right hand side presents two resolutions of the double point a in each knot.

   Summing all four relations we see, that all terms in the right hand side annihilate each other pairwise, and thus we obtain the 4-term relation.



A natural question arises: whether there exist other restrictions for knot invariants of finite order that do not follow from 1-term and 4-term relations? In the case of knot invariants with values in a field $\mathbf{k}$ of characteristic zero the answer is given by a theorem due to M. Kontsevich:

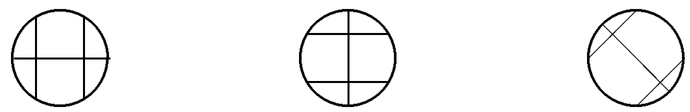


Figure 4.7: Three pictures of one and the same chord diagram

Figure 4.8: Two different chord diagrams

Theorem 4.2.14 *Every function on chord diagrams of order n with values in a field* $\mathbf{k}$ *of characteristic zero satisfying 1-term and 4-term relations can be obtained from a knot invariant of order no greater than n.*

We give the proof of the theorem in Section 4.10.

4.3 The bialgebra of chord diagrams

4.3.1 Chord diagrams and arc diagrams

Consider a circle and put $2n$ points at the vertices of a regular $2n$-gon inscribed in the circle. Split these vertices arbitrarily into n pairs and draw chords connecting the two vertices of each pair. The result will be a chord diagram of order n.

Two chord diagrams are considered equivalent iff one of them may be obtained from the other one by a rotation of the circle.

A reflection generally yields a different chord diagram.

Definition 4.3.1 *A chord diagram of order n* (or *a chord diagram with n chords*) is an oriented circle with a distinguished set of n pairs of points, regarded up to orientation preserving diffeomorphisms of the circle.

Of course, these chords may be drawn as straight line or curve segments whose actual shape is irrelevant; what matters is the way they bind their endpoints into pairs.

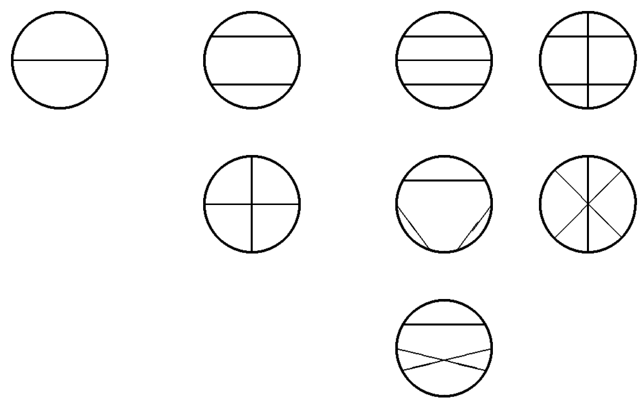


Figure 4.9: All chord diagrams of order 1, 2, 3

Figure 4.9 presents the list of all chord diagrams with 1, 2 and 3 chords.

Sometimes it is more convenient to represent a chord diagram as an arc diagram. To obtain an arc diagram from a chord diagram one must choose an arbitrary point on the circle (different from the ends of all chords) and "cut" the circle at this point ("push the point to infinity"). The circle becomes a straight line, and chords become arcs.

Of course, the representation of a chord diagram as an arc diagram depends on the choice of the point at infinity. One chord diagram usually has $2n$ different arc representations. But in the case when the chord diagram has a rotational symmetry, the number of different arc representations decreases. The cyclic group of order $2n$ of rotations of a regular polygon acts naturally on the set of arc diagram representations of a chord diagram.

On the contrary, an arc diagram uniquely determines the corresponding chord diagram.

Definition 4.3.2 An *arc diagram of order* n (or *an arc diagram with* n *arcs*) is an oriented line with a distinguished set of n pairs of points regarded up to orientation preserving diffeomorphisms of the line.

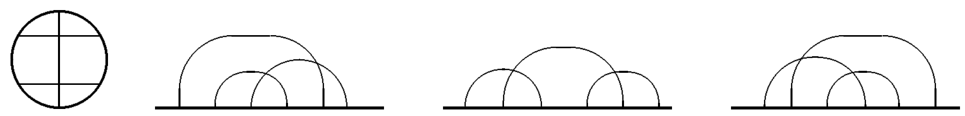


Figure 4.10: A chord diagram and three different arc representations of it

We will usually draw an arc diagram as a horizontal line with arcs lying in the upper half-plane.

4.3.2 4-term relation

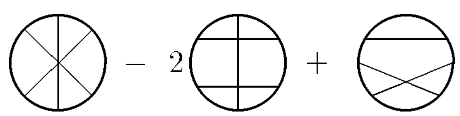
4-term relation is, in fact, a set of equivalence relations on linear combinations of chord diagrams. Four diagrams take part in each such relation. In each of the four diagrams two of the chords are selected. The other chords in all four diagrams coincide. These common chords are irrelevant, and we will not draw them explicitly.

We admit thus the following

Agreement on drawing chord diagrams. If a segment of the circle is drawn as a solid arc, then all the chords whose ends belong to the segment are shown explicitly. If a segment is drawn as a dotted arc, then ends of arbitrary non-drawn chords may belong to the segment.

We will abide by a similar agreement for drawing arc diagrams.

Let $\mathcal{C}_n$ denote the $\mathbf{Z}$-module generated by all chord diagrams of order n as free generators. The elements of $\mathcal{C}_n$ are formal linear combinations of chord diagrams with integer coefficients. For example,



We denote by $\mathcal{A}_n$ the $\mathbf{Z}$-module generated by all arc diagrams of order n as free generators.

Let c_1, c_2, c_3, c_4 be four chord diagrams of order $n \geq 2$ (see Fig. 4.11) while all the chords adjacent to the dotted circle segments are the same in all four diagrams.

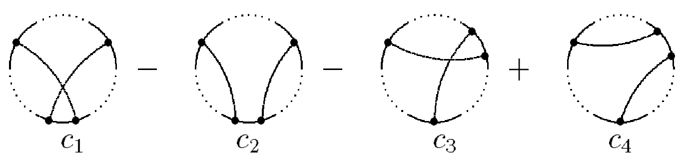


Figure 4.11: A 4-term element in $\mathcal{C}_n$

Definition 4.3.3 The element

$$c_1 - c_2 - c_3 + c_4 \in \mathcal{C}_n$$

is called a *4-term element*.

A 4-term element $a_1 - a_2 - a_3 + a_4 \in \mathcal{A}_n$ has the appearance shown in Fig. 4.12

Figure 4.12: A four-term element in $\mathcal{A}_n$

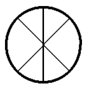
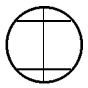
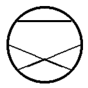
Consider the submodule $\mathcal{CF}_n \subset \mathcal{C}_n$ spanned by all 4-term elements in $\mathcal{C}_n$ and the submodule $\mathcal{AF}_n \subset \mathcal{A}_n$ spanned by all 4-term elements in $\mathcal{A}_n$. We set $\mathcal{CF}_0 = \mathcal{CF}_1 = \mathcal{AF}_0 = \mathcal{AF}_1 = 0$.

Exercise 4.3.4 *Verify that*

1. $\mathcal{AF}_2$ *is generated by the element*  .

2. $\mathcal{CF}_2 = 0$.

3. $\mathcal{CF}_3$ *is generated by the element*

 $-$  $+$  .

Theorem 4.3.5 *There exists a natural isomorphism*

$$\mathcal{A}_n/\mathcal{AF}_n \cong \mathcal{C}_n/\mathcal{CF}_n$$

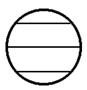
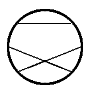
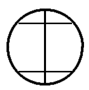
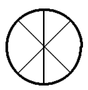
given by mapping the class of an arc diagram $[a] \in \mathcal{A}_n/\mathcal{AF}_n$ *to the class* $[c] \in \mathcal{C}_n/\mathcal{CF}_n$ *of the chord diagram* c *represented by the arc diagram* a*.*

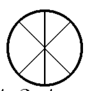
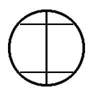
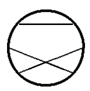
The proof will be given below.

Definition 4.3.6 The quotient module

$$\mathcal{M}_n = \mathcal{C}_n/\mathcal{CF}_n \cong \mathcal{A}_n/\mathcal{AF}_n$$

is said to be the *reduced module of chord diagrams of order* n.

For example, $\mathcal{M}_3$ is a free **Z**-module with 3 generators, given by equivalence classes of the elements [], [], [], while the equivalence class of the element [] may be expressed as

[] = 2[] — []

according to Exercise 4.3.4.

Thus calculations in $\mathcal{M}$ are just ordinary linear algebra calculations where chord diagrams are used instead of letters.

Exercise 4.3.7 Write out a basis for $\mathcal{M}_4, \mathcal{M}_5$.

From now on we will omit square brackets, thus making no difference in the notation for chord (or arc) diagrams and their equivalence classes in the module of chord diagrams.

Proof of Theorem 4.3.5. We must prove that two arc representations of one and the same chord diagram are equivalent modulo the 4-term relations.

Consider an arc diagram a of order n. Pick a point on the line different from all the ends of arcs of a. The point will be one of two ends of an additional $(n+1)$th arc. The second end of the additional arc may occupy one of $2n+1$ places since the arcs of a split the line into $2n+1$ segments. Let a_k denote the arc diagram of order $n+1$ obtained from a by setting the second end of the additional arc on the kth place starting from the left $(k = 0, 1, \ldots, 2n)$.

We must prove that $a_0 - a_{2n} = 0$. Rewrite the left hand side of the equality as

$$a_0 - a_{2n} = a_0 - a_1 + a_1 - a_2 + a_2 - \ldots - a_{2n-1} + a_{2n-1} - a_{2n}.$$

There are $(2n-1)\cdot 2+2 = 4n$ terms in the right-hand side of the previous relation. We split them into n groups, four terms in each group. Such a group is determined by an arc in a: if the arc's ends separate segments $k, k+1$ and $m, m+1$, then the group will be $a_k - a_{k+1} + a_m - a_{m+1}$. Such a group is a 4-term element, and so it equals zero. The theorem is proved.

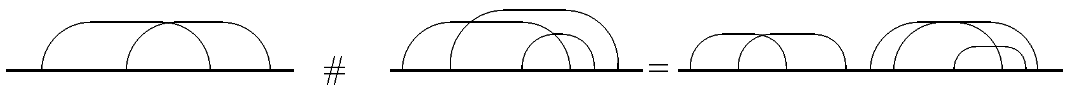
4.3.3 Multiplying chord diagrams

Two arc diagrams may be concatenated, producing a new arc diagram. Considering chord diagrams modulo 4-term relations allows one to descend concatenation of arc diagrams to a new operation in the reduced module

$$\mathcal{M} = \mathcal{M}_0 \oplus \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \ldots$$

of chord diagrams. We call the operation *multiplication* and define it on classes of chord diagrams (or, equivalently, on classes of arc diagrams) as follows.

The *concatenation* $a_1 \# a_2$ of arc diagrams a_1, a_2 simply consists in drawing both diagrams one after another:



For two arc diagrams $a_1 \in \mathcal{A}_m, a_2 \in \mathcal{A}_n$, the product $a_1 a_2 \in \mathcal{A}_{m+n}$ is given by $a_1 a_2 = a_1 \# a_2$. The product of two chord diagrams is defined as the product of their arbitrary arc representations.

Theorem 4.3.8 1. *The multiplication of chord diagrams is a well-defined operation.*

2. *This multiplication may be extended to a bilinear operation*

$$m : \mathcal{M} \otimes \mathcal{M} \to \mathcal{M},$$

respecting the grading

$$m : \mathcal{M}_k \otimes \mathcal{M}_m \to \mathcal{M}_{k+m}$$

making it a commutative associative graded algebra over $\mathbf{Z}$ *with unit. The unit is given by the equivalence class of the empty diagram.*

Note that the unit in the algebra $\mathcal{M}$ over $\mathbf{Z}$ may be understood as a homomorphism $e : \mathbf{Z} \to \mathcal{M}$ (we simply identify the unit with $e(1)$). Since multiplication in $\mathcal{M}$ respects grading we obviously get $e(1) \in \mathcal{M}_0$.

Proof.

1. Follows from Theorem 4.3.5.

2. The arc diagrams $a_1 \# a_2$ and $a_2 \# a_1$ obviously represent the same chord diagram. We omit the purely technical verification of other axioms of algebras.

We would like to stress that in spite of the non commutativity of *arc* diagram concatenation, multiplication of *chord* diagrams is commutative.

4.3.4 A bialgebra structure

Besides multiplication, one more operation may be introduced in the reduced module of chord diagrams. This operation, called *comultiplication*, is a unary one, and it maps $\mathcal{M}$ into its tensor square:

$$\mu : \mathcal{M} \to \mathcal{M} \otimes \mathcal{M}.$$

The tensor square inherits a natural grading from $\mathcal{M}$, and the comultiplication respects this grading in the following sense:

$$\mu : \mathcal{M}_n \to \mathcal{M}_0 \otimes \mathcal{M}_n + \mathcal{M}_1 \otimes \mathcal{M}_{n-1} + \ldots + \mathcal{M}_n \otimes \mathcal{M}_0.$$

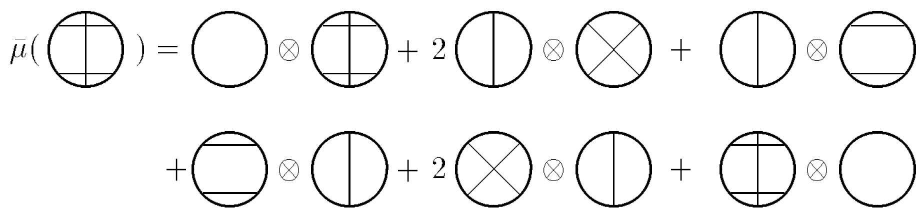
First we define the operation $\bar{\mu} : \mathcal{C} \to \mathcal{C} \otimes \mathcal{C}$.

On a chord diagram c of order n, the comultiplication $\bar{\mu}$ acts as follows. Let $V(c)$ denote the set of all chords of the diagram c. Let $J \subset V(c)$ be a subset of $V(c)$. The subset J determines a chord diagram c_J formed by all chords belonging to the set J. We set

$$\bar{\mu}(c) = \sum_{J \subset V(c)} c_J \otimes c_{V(c)-J},$$

where the sum is taken over all 2^n subsets of the set of n chords.

For example,



Exercise 4.3.9 Prove that the operation $\bar{\mu} : \mathcal{C} \to \mathcal{C} \otimes \mathcal{C}$ descends to a linear operation $\mu : \mathcal{M} \to \mathcal{M} \otimes \mathcal{M}$. In other words, prove that if one replaces a chord diagram by a linear combination of chord diagrams equivalent to the original one modulo the 4-term relation, the result of comultiplication will be equivalent to the original one modulo the 4-term relation.

The operation μ admits a *counit*, i.e., a homomorphism $\epsilon : \mathcal{M} \to \mathbf{Z}$ given by the natural projection $\epsilon : \mathcal{M} \to \mathcal{M}_0 \cong \mathbf{Z}$.

The algebra and coalgebra structures in $\mathcal{M}$ allow one to include investigation of $\mathcal{M}$ into a general framework of graded bialgebras. Let us give formal definitions.

Let $\mathbf{k}$ be a commutative associative ring with unit. The ring $\mathbf{k}$ may be regarded as a graded algebra over itself if we set $\mathbf{k} = \mathbf{k} \oplus 0 \oplus 0 \oplus \ldots$. From another point of view, $\mathbf{k}$ may be regarded as a graded coalgebra over itself with the comultiplication $\mu : a \mapsto a \otimes a$ for $a \in \mathbf{k}$. The unit and the counit are represented by the identical homomorphism $\mathbf{k} \to \mathbf{k}$.

Definition 4.3.10 Let $E_0, E_1, E_2, \ldots$ be finitely generated free $\mathbf{k}$-modules, $E_0 \cong \mathbf{k}$. The graded module $E = E_0 \oplus E_1 \oplus E_2 \oplus \ldots$ is called *a bialgebra*, if it carries

1. a graded algebra structure $m : E \otimes E \to E, m(E_l \otimes E_n) \subset E_{l+n}$ with a unit $e : \mathbf{k} \to E$, and

2. a graded coalgebra structure $\mu : E \to E \otimes E, \mu(E_n) \subset E_0 \otimes E_n + E_1 \otimes E_{n-1} + \ldots + E_n \otimes E_0$ with a counit $\epsilon : E \to \mathbf{k}$

satisfying the following properties:

1. e is a graded coalgebra homomorphism;

2. ϵ is a graded algebra homomorphism;

3. μ is a graded algebra homomorphism (or, equivalently, m is a graded coalgebra homomorphism).

Exercise 4.3.11 Verify that the reduced module $\mathcal{M}$ of chord diagrams forms a bialgebra over $\mathbf{Z}$ with respect to the previously introduced operations. It is sufficient to prove that the comultiplication is a homomorphism for the algebra structure of $\mathcal{M}$.

Further we will consider only even commutative and even cocommutative bialgebras referring to them simply as to commutative bialgebras.

Definition 4.3.12 A bialgebra E is called *even commutative*, if $m(a \otimes b) = m(b \otimes a)$ for any pair of elements $a, b \in E$. It is called *even cocommutative*, if for any element $a \in E$ the result of comultiplication is a symmetric element of $E \otimes E$, i.e. $\mu(a) = \sum b_i \otimes c_i = \sum c_i \otimes b_i$.

It is clear from the definitions that $\mathcal{M}$ is even commutative and even cocommutative.

Example 4.3.13 *The polynomial bialgebra.*

Let $X = \{x_1, x_2, \ldots\}$ be a set of commuting variables. Let x_i be a variable of (integer) weight w_i. We suppose that for any positive integer there is only finite set of variables of this weight. The $\mathbf{Z}$-module E_n is freely generated by monomials

$$x_1^{q_1} \dots x_k^{q_k}$$

with the weight of the monomial given by

$$n = w_1 q_1 + \dots + w_k q_k.$$

In other words, E_n is the module of "quasi homogeneous polynomials of order n" in the variables X. Setting $E_0 = \mathbf{Z}$, we get a graded algebra $E = E_0 \oplus E_1 \oplus E_2 \oplus \dots$ over $\mathbf{Z}$. The natural embedding $\mathbf{Z} \to E_0$ provides E with the unit.

The coproduct for a monomial $y_1 \dots y_m \in E$ ($y_i \in X$) is given by the formula

$$\mu(y_1 \dots y_m) = 1 \otimes y_1 \dots y_m + y_1 \otimes y_2 \dots y_m + y_2 \otimes y_1 y_3 \dots y_m + \dots + y_1 \dots y_m \otimes 1.$$

The polynomial bialgebra is in a sense a standard bialgebra: any commutative bialgebra over a field of characteristic zero is isomorphic to a polynomial bialgebra (see the next subsection).

4.3.5 Structure theorem for the bialgebra $\mathcal{M}$

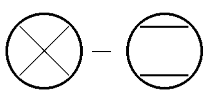
A special class of elements plays an important role in understanding the structure of bialgebras. These elements are "simpler" than the others.

Definition 4.3.14 An element $p \in E$ is called *primitive* if $\mu(p) = 1 \otimes p + p \otimes 1$.

Note that a nonzero element $p \in E_0$ cannot be a primitive element. Indeed, $\mu(p) = p \cdot \mu(1) = p \cdot 1 \otimes 1$, while $1 \otimes p + p \otimes 1 = 2p \cdot 1 \otimes 1$.

Exercise 4.3.15 Prove that an element p of the polynomial bialgebra is primitive if and only if it is a linear polynomial.

Exercise 4.3.16 Verify that any primitive element in $\mathcal{M}_2$ is a multiple of

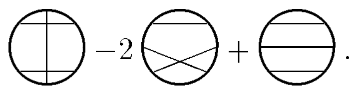
 .

Statement 4.3.17 *All primitive elements of any bialgebra form a graded submodule $P \subset E$, $\quad P = P_1 \oplus P_2 \oplus \dots$, where $P_k = P \cap E_k$.*

The proof is a direct verification.

Exercise 4.3.18 Find a basis in the module of primitive elements in $\mathcal{M}_3$.

Answer. The module $\mathcal{P}_3$ of primitive elements is generated, for example, by the element



Of course, primitive elements do not form neither a subalgebra, nor a subcoalgebra in E.

Primitive elements generate a subalgebra in E. In the case of an even commutative and even cocommutative bialgebra E over a field $\mathbf{k}$ of characteristic zero, this subalgebra is a polynomial algebra, and it coincides with E itself. This statement is a special case of the Milnor-Moor structure theorem for bialgebras.

Let us fix a basis $\{p_{k1}, \dots, \}$ in each space of primitive elements $\mathcal{P}_k \subset \mathcal{M}_k$. Denote by $S(\mathcal{P})$ the polynomial bialgebra in these elements.

Theorem 4.3.19 *([18]) Over a field $\mathbf{k}$ of characteristic zero the bialgebras $\mathcal{M}$ and $S(\mathcal{P})$ are isomorphic.*

The dimensions of the spaces of primitive elements $\mathcal{P}_n \otimes k$ were computed by Vassiliev [22] and Bar-Natan [3] for the first few values of the grading:

n	0	1	2	3	4	5	6	7	8	9
$\dim \mathcal{P}_n$	0	1	1	1	2	3	5	8	12	18

The dimensions of the spaces $\mathcal{M}_n$ can be easily calculated from that of $\mathcal{P}_n$. Indeed, the following relation holds:

Statement 4.3.20

$$m(t) = \frac{1}{(1 - p_1 t)(1 - p_2 t^2)(1 - p_3 t^3)\dots}$$

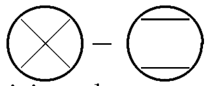
where $m(t) = 1 + m_1 t + m_2 t^2 + \dots$ is the generating function for the dimensions of the spaces $\mathcal{M}_n$ and p_n is the dimension of the space $\mathcal{P}_n = \mathcal{P} \cap \mathcal{M}_n$.

For the first values of this grading the statement gives

n	0	1	2	3	4	5	6	7	8	9
$\dim \mathcal{M}_n$	1	1	2	3	6	10	19	33	60	104

Comparing $\dim \mathcal{M}_9 = 104$ with the expected number of chord diagrams (exceeding $17!!/18$) one should note the dramatic decrease in the number of degrees of freedom.

4.3.6 Primitive elements of bialgebra of chord diagrams

How can one find the primitive element  in $\mathcal{M}_2$? What must be done to construct one or more primitive elements in $\mathcal{M}_8$, for example? This subsection is devoted to answering such questions.

In the case of a polynomial bialgebra, the space of primitive elements coincides with that of linear polynomials in the generators. The problem of writing out the primitive elements in other bialgebras is mainly connected with the fact that we know a set of generators of the algebra, which are not primitive, and we do not know *a priori* their representation as polynomials in primitive elements. This is the case for the algebra $\mathcal{M}$.

There exists a natural projection of the bialgebra $\mathcal{M}$ onto the space of primitive elements, which allows one to construct a primitive element knowing only an element of $\mathcal{M}$. We shall write out an explicit formula for this projection.

Let $V = V(c)$ denote the set of chords of a chord diagram c. For a subset $V_1 \subset V(c)$, let (V_1) denote the chord diagram formed by all chords from V_1. Set

$$p(c) = (V) - 1! \sum_{V_1 \sqcup V_2 = V} (V_1)(V_2) + 2! \sum_{V_1 \sqcup V_2 \sqcup V_3 = V} (V_1)(V_2)(V_3) - \dots \qquad (4.1)$$

where the sums are taken over all disjoint partitions of the set V into nonempty subsets.

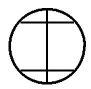
Theorem 4.3.21 *For each diagram c the element $p(c)$ is primitive.*

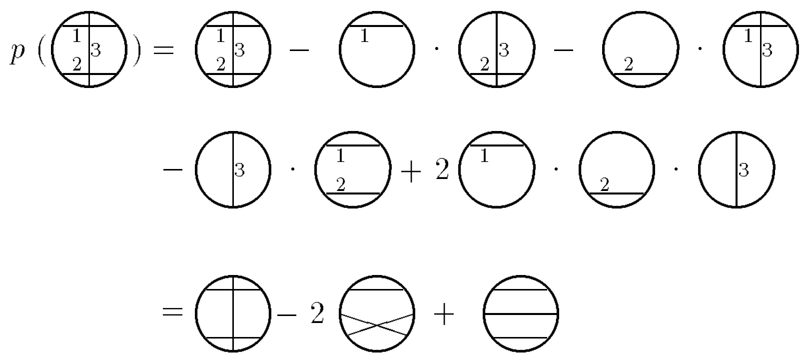
The next subsection will be devoted to a proof of this theorem.

Corollary 4.3.22 *The module of chord diagrams is generated by its primitive elements over* $\mathbf{Z}$.

Indeed, for a chord diagram c, one has $c = p(c) + \ldots$, where dots denote a polynomial in primitive elements of smaller order with integer coefficients.

Remark 4.3.23 It follows from the results of sec. 4.5 that the element $p(c)$ is zero if and only if the diagram c is decomposable.

Example 4.3.24 For the chord diagram , we have



which is a primitive element in $\mathcal{M}_3$.

4.3.7 The bialgebra of partitions

We associate with each positive integer n a bialgebra $\mathcal{B}(n)$, which we call the *bialgebra of partitions*.

Algebra structure The definition proceeds as follows. Let $N_n = \{1, 2, \ldots, n\}$. Consider all non-empty subsets $V \subset N_n$ as commuting variables. We prefer to denote such a subset as a sequence of elements in brackets: $(1), (12)$, and so on. As an algebra, $\mathcal{B}(n)$ is a polynomial algebra in these $2^n - 1$ variables.

The *order* of a variable $V \subset N_n$ is just the number of elements in V. This order makes the algebra $\mathcal{B}(n)$ a graded algebra $\mathcal{B}(n) = \mathcal{B}_0(n) \oplus \mathcal{B}_1(n) \oplus \ldots$, where $\mathcal{B}_0(n) \cong \mathbf{Z}$ (the generator of $\mathcal{B}_0(n)$ is the empty subset $()$), $\mathcal{B}_k(n)$ is the free $\mathbf{Z}$-module freely generated over $\mathbf{Z}$ by all monomials of order k.

For example, for $n = 3$,

$$
\begin{aligned}
\mathcal{B}_1(3) &= \ <(1),(2),(3)> \\
\mathcal{B}_2(3) &= \ <(12),(13),(23),(1)(2),(1)(3),(2)(3),(1)(1),(2)(2),(3)(3)> \\
\mathcal{B}_3(3) &= \ <(123),(1)(12),(1)(13),(1)(23),\ldots,(1)(1)(1),(1)(1)(2),\ldots>
\end{aligned}
$$

A monomial in $\mathcal{B}(n)$ will be called *a partition*.

The empty partition $e = ()$ is the unit with respect to this multiplication.

Coalgebra structure Let $V \subset N_n$ and let (V) denote a partition of V, consisting of one set, namely V itself. We set

$$\mu((V)) = e \otimes (V) + (V) \otimes e + \sum (V_1) \otimes (V_2),$$

where $V = V_1 \sqcup V_2$ is an arbitrary partition of V into a disjoint union of two nonintersecting subsets, and the sum is taken over all such partitions.

For other types of partitions, the comultiplication is extended as an algebra structure homomorphism.

Statement 4.3.25 *The module $\mathcal{B}(n)$ with the multiplication and comultiplication thus introduced is a graded (even) commutative and cocommutative bialgebra.*

Primitive elements

Theorem 4.3.26 *For an arbitrary subset $V \subset N_n$ the element*

$$p(V) = (V) - 1!\sum(V_1)(V_2) + 2!\sum(V_1)(V_2)(V_3) - 3!\sum(V_1)(V_2)(V_3)(V_4) + \ldots \tag{4.2}$$

is a primitive element in $\mathcal{B}(n)$.

Corollary 4.3.27 *The submodule $\mathcal{P}(n) \subset \mathcal{B}(n)$ of primitive elements is generated by elements of type* (4.2). *The bialgebra $\mathcal{B}(n)$ is isomorphic to the symmetric bialgebra of $\mathcal{P}(n)$.*

Remark 4.3.28 Primitive elements of $\mathcal{B}(n)$ are concentrated in the first n orders of grading.

Proof of Theorem 4.3.26. It is clearly sufficient to prove the theorem for the case $V = N_n$.

The coefficient at the term

$$(V_1)\dots(V_k)\otimes(V_{k+1})\dots(V_{k+m})$$

in $\mu(N_n)$ is uniquely determined by two numbers k and m. Denote this coefficient by a_{km}. We set $a_{00}=0$. We have

$$a_{km} = (k+m-1)! - \frac{(k+m-2)!km}{1!} + \frac{(k+m-3)!k(k-1)m(m-1)}{2!} - \dots$$

Indeed, the term we are considering can be obtained by comultiplying a term with $(k+m-j)$ subsets in partition for $j=0,1,\dots,\min(k,m)$. Such a term may be chosen by picking j subsets from $(V_1),\dots,(V_k)$ and j subsets from $(V_{k+1}),\dots,(V_{k+m})$ and taking their pairwise union. We must divide by $j!$, since the partitions are unordered.

Multiplying and dividing by $k!m!$, we obtain the following formula for a_{km}:

$$\begin{aligned} a_{km} &= k!m!\left(\frac{(k+m-1)!}{0!k!m!} - \frac{(k+m-2)!}{1!(k-1)!(m-1)!} + \dots\right) \\ &= k!m!\sum_{j=0}^{\min(k,m)} \frac{(k+m-1-j)!}{j!(k-j)!(m-j)!} \end{aligned}$$

Consider the exponential generating function for the sequence of coefficients a_{km} and make the following transformations

$$\begin{aligned} \sum_{k,m=0}^{\infty} \frac{a_{km}}{k!m!}x^k y^m &= \sum_{k,m=0}^{\infty} x^k y^m \sum_{j=0}^{\min(k,m)} (-1)^j \frac{(k+m-1-j)!}{j!(k-j)!(m-j)!} \\ &= \sum_{j=0}^{\infty}(-1)^j \frac{1}{j!} \sum_{k,m=j}^{\infty} \frac{(k+m-1-j)!}{(k-j)!(m-j)!} x^k y^m \\ &= \sum_{j=0}^{\infty}(-1)^j \frac{1}{j!} y^j \sum_{k=j}^{\infty} \frac{x^k}{(k-j)!} \sum_{m'=0}^{\infty} \frac{(k-1+m')!}{m'!} y^{m'} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty}(-1)^j \frac{1}{j!} y^j \sum_{k=j}^{\infty} \frac{x^k}{(k-j)!}\left(\frac{1}{1-y}\right)^{(k-1)} \\
&= \sum_{j=0}^{\infty}(-1)^j \frac{1}{j!} y^j x^j \left(\frac{\partial}{\partial y}\right)^{j-1} \sum_{k'=0}^{\infty} \frac{\left(x\frac{\partial}{\partial y}\right)^{k'}}{k'!}\frac{1}{1-y} \\
&= \sum_{j=0}^{\infty}(-1)^j \frac{1}{j!} y^j x^j \left(\frac{\partial}{\partial y}\right)^{j-1} \frac{1}{1-y-x} \\
&= -\sum_{j=0}^{\infty} \frac{\left(-xy\frac{\partial}{\partial y}\right)^{j}}{j!} \log(1-x-y) \\
&= -\log(1-x-y+xy) \\
&= -\log(1-x)-\log(1-y)
\end{aligned}
$$

The upper index in parentheses denotes, as usual, the corresponding derivative. Note that

$$
\left(\frac{1}{1-y}\right)^{(-1)} = -\log(1-y).
$$

In this calculation we used the fact that the exponent of the derivative is a shift twice.

Thus, the coefficient at $(V_1)\dots(V_k)\otimes(V_{k+1})\dots(V_{k+m})$ in $\mu(N_n)$ is 1 if and only if the left or the right hand side of this tensor product is empty, i.e. $k=0$ or $m=0$, and it vanishes otherwise. This means precisely that $\mu(N_n)$ is primitive, and the theorem is thus proved.

We are now able to prove Theorem 4.3.21.

Proof of Theorem 4.3.21. Let c be a chord diagram with n chords. Associate with c a homomorphism $m_c\colon \mathcal{B}(n) \to \mathcal{M}$ in the following way. Enumerate the chords of c by the numbers $\{1,\dots,n\}$ arbitrarily, so that each chord has the unique number. The partition (V) maps to the chord diagram $m_c(V)$ consisting of the chords numbered by elements of V. The partition $(V_1)\dots(V_k)$ maps to the chord diagram $m_c(V_1)\dots m_c(V_k)$.

The mapping m_c obviously respects both multiplication and comultiplication and thus extends to a homomorphism of bialgebras. Primitive elements map to primitive elements, which proves the theorem.

4.4 Invariants and preinvariants

4.4.1 A bialgebra structure on the module $\mathcal{V}$ of knot invariants

The multiplication and comultiplication in the module of chord diagrams do not appear by an accident. They just mimic corresponding (dual) operations in knot invariants.

Consider the module $\mathcal{V}$ of finite-order invariants with values in a ground ring K. This module is, in fact, a K-algebra with respect to pointwise multiplication of invariants:

$$m(v_1, v_2)(u) = v_1 v_2(u) = v_1(u) v_2(u).$$

The algebra is a filtered algebra:

Statement 4.4.1 $\mathcal{V}_m \mathcal{V}_n \subset \mathcal{V}_{m+n}$.

Proof. Exercise for the reader.

Hint. Use exercise 4.2.11.

A *concatenation* (or *connected sum*) of knots is a natural operation of gluing two knots together respecting their orientation, see Fig. 4.13.

We denote the concatenation of two knots u_1, u_2 by $u_1 \# u_2$.

The concatenation supplies the algebra $\mathcal{V}$ with a *comultiplication*. For $v \in \mathcal{V}$ we set

$$\mu(v)(u_1 \otimes u_2) = v(u_1 \# u_2).$$

The comultiplication makes $\mathcal{V}$ a filtered coalgebra:

Statement 4.4.2

$$\mu(\mathcal{V}_n) \subset \mathcal{V}_0 \otimes \mathcal{V}_n + \mathcal{V}_1 \otimes \mathcal{V}_{n-1} + \ldots + \mathcal{V}_n \otimes \mathcal{V}_0.$$

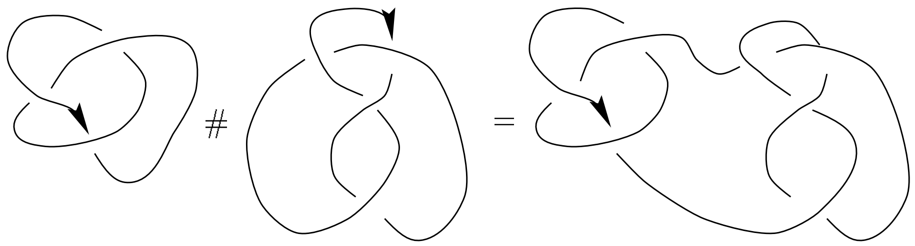


Figure 4.13: Concatenation of knots

Proof. Exercise for the reader.

Now let us associate with the filtered module $\mathcal{V}$ the corresponding graded module

$$\operatorname{gr}\mathcal{V} = \mathcal{V}_0 \oplus \mathcal{V}_1/\mathcal{V}_0 \oplus \mathcal{V}_2/\mathcal{V}_1 \oplus \dots .$$

Theorem 4.4.3 *The operations of multiplication $m : \mathcal{V}\otimes\mathcal{V} \to \mathcal{V}$ and comultiplication $\mu : \mathcal{V} \to \mathcal{V}\otimes\mathcal{V}$ descend to multiplication $m : \operatorname{gr}\mathcal{V}\otimes\operatorname{gr}\mathcal{V} \to \operatorname{gr}\mathcal{V}$ and comultiplication $\mu : \operatorname{gr}\mathcal{V} \to \operatorname{gr}\mathcal{V}\otimes\operatorname{gr}\mathcal{V}$ making the K-module $\operatorname{gr}\mathcal{V}$ a graded bialgebra with the unit $e\colon K \to \mathcal{V}_0$ and the counit $\epsilon : \mathcal{V}_0 \to K$ given by the natural isomorphism mappings.*

Proof. Exercise for the reader.

Now Kontsevich theorem 4.2.14 can be reformulated as follows:

Theorem 4.4.4 *Over a field $\mathbf{k}$ of characteristic zero, the bialgebra $\operatorname{gr}\mathcal{V}$ is isomorphic to the dual bialgebra $\mathcal{N}_0^* \oplus \mathcal{N}_1^* \oplus \mathcal{N}_2^* \oplus \dots$ (having the natural structure of a graded bialgebra).*

4.4.2 Renormalization

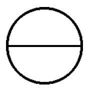
Functions on chord diagrams originating from Vassiliev knot invariants satisfy two types of relations, 1-term relations and 4-term relations. And up to now we considered chord diagrams modulo only 4-term relations. The reason is that the 1-term relations are much less important than 4-term ones.

To be more precise, let $\mathcal{M}^*$ be the module of functions on $\mathcal{M}$, and $\mathcal{N}^* \subset \mathcal{M}^*$ – the submodule of functions, also satisfying 1-term relations.

Theorem 4.4.5 *There exists a natural projection* $\mathrm{Rn}^* : \mathcal{M}^* \to \mathcal{N}^*$*, i.e.,* $(\mathrm{Rn}^*)^2 = \mathrm{Rn}^*$*;* $\mathrm{Rn}^* \|_{\mathcal{N}^*} = \mathrm{id}$*.*

This projection is called the *renormalization operator*, and it may be written out explicitly. Having the renormalization operator in mind, the study of functions satisfying only 4-term relations is equivalent to the study of functions satisfying both 4-term and 1-term relations.

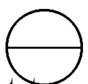
Proof. Let $X_n : \mathcal{M}_n \to \mathcal{M}_n$ be the *chord separating operator in order* n defined by the formula

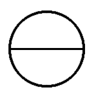
$$X_n\colon c \mapsto$$  $$\cdot \sum_{i=1}^{n} c_i.$$

Here c is a n-chord diagram, c_i is a $(n-1)$-chord diagram obtained from c by removing the i-th chord.

Introduce the operator $\mathrm{Rn} : \mathcal{M} \to \mathcal{M}$ by the following formula

$$\mathrm{Rn}_n = \prod_{i=1}^{n} \left(1 - \frac{X_n}{i}\right).$$

In the case when the ground ring $\mathbf{k}$ is a field of characteristic zero, $\mathcal{M}$ is isomorphic to the polynomial algebra in primitive elements, and Rn is equivalent to setting  $= 0$ in each polynomial. In order to prove the theorem it is sufficient to prove two lemmas.

Lemma 4.4.6 *The kernel of* Rn *contains the submodule*  $\cdot\, \mathcal{M} \subset \mathcal{M}$.

Proof. Suppose a chord diagram c has $k \geq 1$ isolated chords. Then $(1 - X_n/k)(c)$ is a linear combination of diagrams with at least $k+1$ isolated chords, and we can proceed by induction.

Lemma 4.4.7 *The operator* Rn_n *is the identity on* $\mathcal{P}_n \subset \mathcal{M}_n$ *for* $n \geq 2$*.*

Indeed, the chord separating operator X_n maps $\mathcal{P}_n$ to zero.

We finish this subsection with a proposition describing properties of the renormalization operator, which can be verified directly.

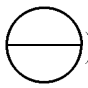
Statement 4.4.8 *1. The operator $X: \mathcal{M} \to \mathcal{M}$ is a derivation, i.e.,*

$$X(c_1c_2) = X(c_1)c_2 + c_1X(c_2)$$

.

2. *The operator $\mathrm{Rn}: \mathcal{M} \to \mathcal{M}$ is a bialgebra homomorphism.*

3. *The operator $\mathrm{Rn}_n: \mathcal{M}_n \to \mathcal{M}_n$ can also be given by the formula*

$$\mathrm{Rn}: c \mapsto \sum_J (-$$  $$)^{n-|J|} \cdot c_J,$$

where $J \subset V(c)$ runs over all subsets of chords of the diagram c, the diagram c_J being formed by all chords from J, and $|J|$ being the number of chords in J.

4.4.3 Preinvariants

As we have seen from the previous subsection, investigation of finite order knot invariants is equivalent, in a sense, to investigation of linear functions on $\mathcal{M}$, i.e., of the module $\mathcal{M}^*$.

Definition 4.4.9 Elements of $\mathcal{M}^* = \mathrm{Hom}(\mathcal{M}, K)$ are called *preinvariants* (with values in K). A preinvariant is called *multiplicative*, if it is a homomorphism of ring structures in $\mathcal{M}$ and K.

Modules $\mathcal{M}_k^*$ for $k = 0, 1, 2, \ldots$ of preinvariants of order k are naturally embedded in $\mathcal{M}^*$. Indeed, each preinvariant of order k can be extended to an element in $\mathcal{M}^*$ simply by setting it equal to zero on $\mathcal{M}_i$ for $i \neq k$. Note that $\mathcal{M}^*$ does not coincide with $M_0^* \oplus \mathcal{M}_1^* \oplus \mathcal{M}_2^* \ldots$.

We are now ready to present some simple examples of preinvariants. Large series of examples are presented in sections 4.5 and 4.6. Using renormalization, it is possible to construct an invariant from each preinvariant.

We usually define values of a preinvariant either on chord or arc diagrams, meaning that it can be extended to a linear function on $\mathcal{M}$.

Example 4.4.10 The number of chords in the diagram is obviously a preinvariant. It is a constant function on $\mathcal{M}_n$, and thus is of no interest.

Example 4.4.11 The number of chord intersections is a preinvariant. Indeed, it is obvious that interchanging the ends of two neighboring chords we change the number of intersections by one.

Exercise 4.4.12 Prove that the renormalized number of intersections is identically equal to zero.

Example 4.4.13 Consider a chord diagram c as a graph with rotations. The vertices of the graph are ends of the chords. The rotation at each end of each chord is determined by the standard orientation of the plane.

Let $f(c)$ be the number of faces of the corresponding gluing (see Fig. 4.14).

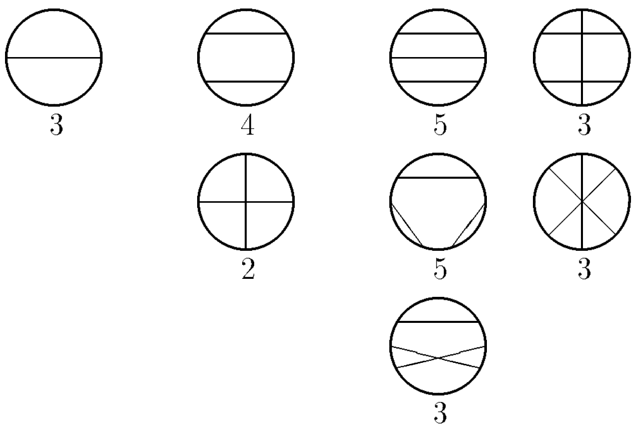


Figure 4.14: Number of faces for all chord diagrams of order 1, 2, 3

Let $F : \mathbf{N} \to K$ be any fixed function with values in an arbitrary ring K. Then the mapping $F \circ f$ provides a preinvariant with values in K.

Indeed, consider the four terms of a 4-term element with doubled chords (see Fig. 4.15). In the chord diagram c_1 with doubled chords the tails (a, d), (b, e), (c, f) are glued pairwise. The same pairs of tails are glued in the diagram c_3. Therefore the numbers of faces in the diagrams c_1 and c_3 coincide. Similar argument is valid for the pair c_2 and c_4. We see that the 4-term relation splits into a pair of 2-term relations. Composition with an arbitrary function does not change the situation.

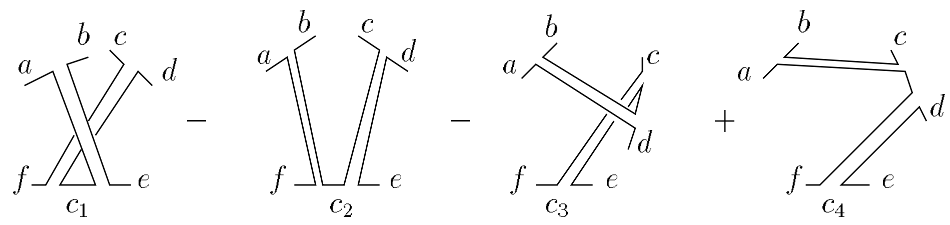


Figure 4.15: Number of faces in the 4-term element

Example 4.4.14 Associate to any chord diagram c its *intersection graph* $\Gamma(c)$. The vertices of the graph correspond to chords of the diagram, and two vertices are connected by an edge iff the corresponding chords intersect. Let $\chi : G \mapsto \chi_G \in \mathbf{Z}[t]$ be the mapping that takes each graph G to its chromatic polynomial. Then $\chi \circ \Gamma$ is a preinvariant with values in $\mathbf{Z}[t]$. The proof is given in section 4.5.

4.4.4 Constructing knot invariants from diagram invariants. Actuality table

A diagram invariant f of order n originates from a knot invariant F of order $\leq n$. We would like to compute the value of F on a knot without singularities. Of course, f does not determine F uniquely. It determines F only up to addition of a knot invariant of order $\leq n-1$.

Given a diagram invariant f we know values of F on each singular knot with n or more singularities. We must extend F to knots with less than n singularities.

Let us proceed by induction and determine F first on knots with $n-1$ singularities.

Consider an arbitrary basis in the module $\mathcal{N}_{n-1}$ presented by chord diagrams $c_1, \ldots, c_k$ of order $n-1$. Fix a set $u_1, \ldots, u_k$ of singular knots with $n-1$ singular points such that the chord diagram of a knot u_i coincides with c_i for $i = 1, \ldots, k$.

First of all, we are free to choose values of the invariant F on knots $u_1, \dots, u_k$. Values of F on other knots with $n-1$ double points are uniquely determined by this data.

Now let $u : S^1 \to \mathbf{R}^3$ be a singular knot with $n-1$ singular points. The chord diagram $c(u)$ of the knot u is equivalent modulo 4-term relations to a linear combination of chord diagrams $c_1, \dots, c_k$. Consider a sequence of "4-term transformations" realizing this equivalence. Each "4-term transformation" is a substitution of the linear combination of three singular knots instead of the original one.

At the end of the sequence one obtains a linear combination of singular knots with specified chord diagrams.

Now let u_1' be a singular knot whose chord diagram coincides with c_1. Through a sequence of elementary transformations (see sec. 4.2.2) one obtains the knot u_1 from u_1'. The increment of F during each elementary transformation is known since F is known on each knot with n double points.

Thus we have calculated the value of F on each knot with $n-1$ double points. We may continue the process and calculate F on each knot.

For any value of n "basic knots" with n double points may be chosen independently on the precise invariant we are calculating. On Fig. 4.16 one may see a sample set of basic knots up to the order 4.

The values of an invariant F on basic knots form the *actuality table* for F.

Problem 4.4.15 A basic singular knot with given chord diagram may be chosen arbitrarily. It seems naturally, however, that some possibilities are more preferable than the others. Is there a regular way to define "the most preferable" basic knot with given chord diagram?

4.4.5 Vassiliev knot invariants and other knot invariants

This subsection is devoted to examples of some classical knot invariants that were known long before Vassiliev introduced finite order invariants. All these invariants are functions on knots without singularities.

Knot invariants are usually defined for plane projections of knots, and then one must prove that they are well defined. For such proofs we refer the

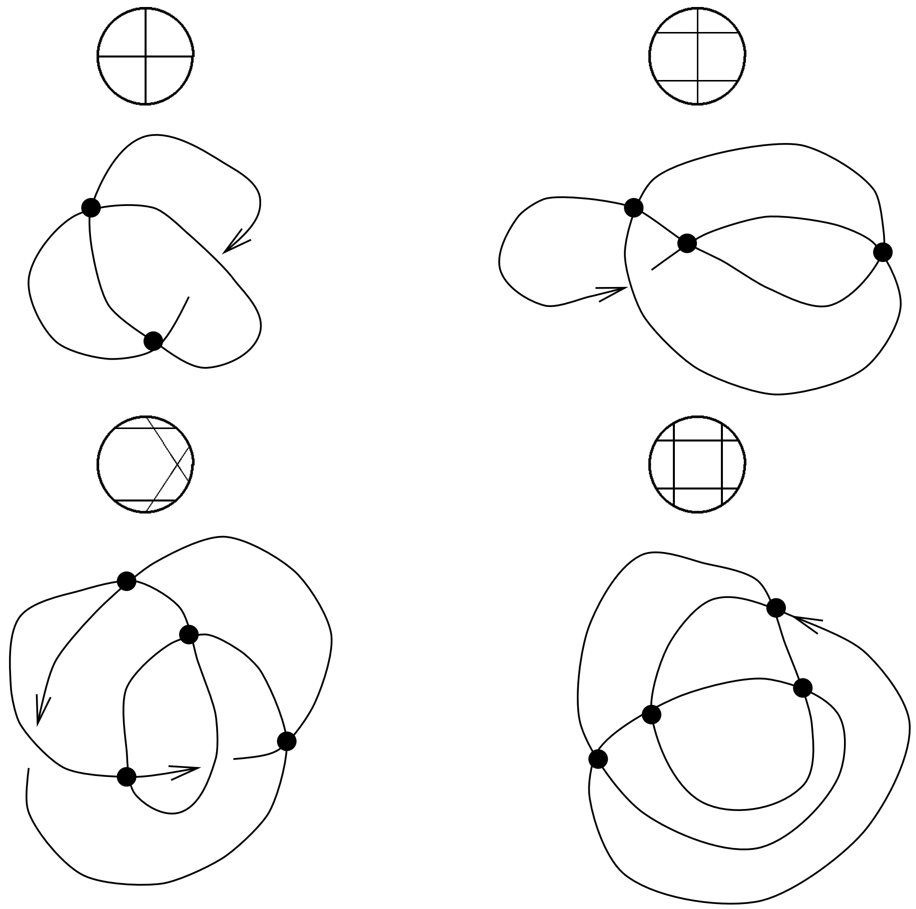


Figure 4.16: Sample basic singular knots and their chord diagrams

reader to [12]. But sometimes the invariance of the function follows directly from the definition, as in the following two examples.

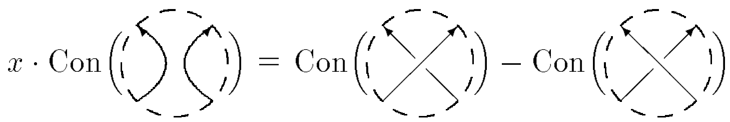
Example 4.4.16 The *genus* of a knot u is the minimal genus of a surface embedded in $\mathbf{R}^3$ and such that the boundary of the surface coincides with u.

Example 4.4.17 The *fundamental group* of the compliment of a knot in $\mathbf{R}^3$ is a knot invariant.

For large classes of knot invariants it is proved already that they may be expressed in terms of Vassiliev invariants. The words “may be expressed” should be made precise in each case separately. We hope, however, that the two following examples clearly explain the meaning of these words.

Example 4.4.18 The *Conway polynomial* is a knot invariant with values in the ring $\mathbf{Z}[x]$ in one variable x. In fact, the Conway polynomial is defined on *links*, not only on knots. A *link* is an isotopy class of nonsingular embeddings of a finite set of circles in $\mathbf{R}^3$.

Consider a generic plane projection of a knot (or a link), i.e. a projection with only transversal double points of selfintersection. We define the Conway polynomial by the recurrent rule:

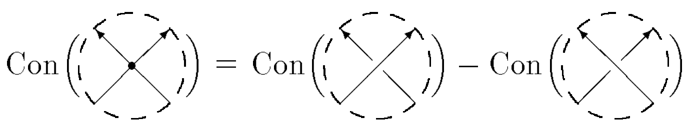


and the initial value Con(trivial knot) $= 1$.

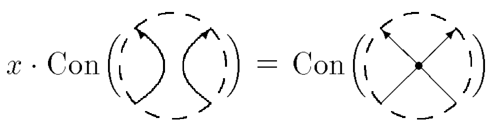
This definition allows one to calculate the value of the Conway polynomial knowing a generic projection of the knot. The definition is very close to that of Vassiliev invariants so there is nothing strange in the fact, that the coefficients of Conway polynomial are Vassiliev invariants.

Statement 4.4.19 *Let c_n be the knot invariant with values in $\mathbf{Z}$ equal to the coefficient at x^n of the Conway polynomial. Then c_n is a Vassiliev knot invariant of order $\leq n$.*

Indeed, extend the Conway polynomial to singular knots according to the rule



So by the defining recurrent rule we have



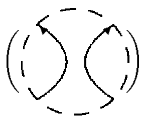
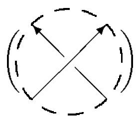
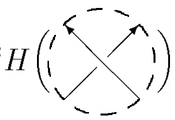
and the coefficient c_n obviously vanishes on singular knots with more than n double points.

Thus each coefficient c_n determines a preinvariant of order n.

Exercise 4.4.20 According to Example 4.4.13 the number of cells in a graph with rotations determined by a chord diagram is a preinvariant. Prove that the value of c_n on a chord diagram is equal to 1, if there are only two cells, and it vanishes otherwise.

Remark 4.4.21 Although the coefficients of the Conway polynomial are Vassiliev invariants, the polynomial itself is not a Vassyliev invariant. Indeed, the sequence of coefficients has increasing degree.

Example 4.4.22 The *HOMFLY polynomial* $H = H(N,q)$ is a knot (and link) invariant. It depends on two variables, denoted usually by N and $q^{1/2}$. Similarly to the Conway polynomial, the HOMFLY polynomial is defined by relation

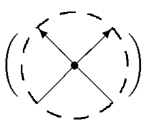
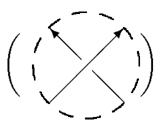
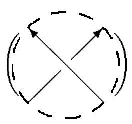
$(q^{1/2} - q^{-1/2})H$  $= q^{N/2}H$  $- q^{-N/2}H$ 

and the initial value $H(\text{trivial knot}) = 1$.

Substitute $q = q(x) = 1 + q_1x + q_2x^2 + \ldots$ into the HOMFLY polynomial and consider it as a power series in variable x. Then we have

Statement 4.4.23 *The coefficient $h_n = h_n(N)$ of the power series $H = H(N, q(x))$ is an invariant of order $\leq n$.*

Proof Indeed, the extension of the HOMFLY polynomial to singular knots is given by the equation

H  $=$ H  $-$ H  $= x\tilde{H}$

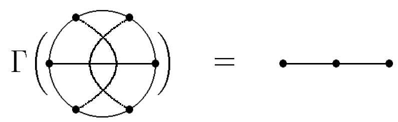
and we can argue as in the case of Conway polynomial.

4.5 Constructing invariants via intersection graphs

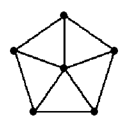
4.5.1 Intersection graph of a chord diagram

Definition 4.5.1 The *intersection graph* $\Gamma(c)$ of a chord diagram c is a graph whose vertices correspond to the chords of c and two vertices are connected by an edge iff the corresponding chords intersect. (Two chords, a and b, are said to intersect if their endpoints a_1, a_2 and b_1, b_2 appear in interchanging order a_1, b_1, a_2, b_2 along the circle.)

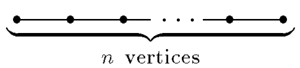
For example,



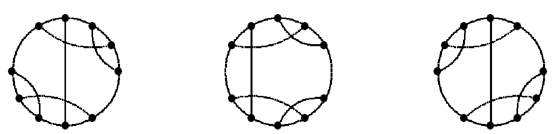
Note that not every graph can be obtained as the intersection graph of a diagram. The simplest example appears in degree 6 and is given by the graph



Different diagrams can have one and the same intersection graph. For example, consider the graph a_n:



For $n = 5$ there are three diagrams having a_5 as intersection graph:



The intersection graph of a chord diagram carries a lot of information from the diagram. Many chord diagram invariants depend, in fact, on the intersection graph of the diagram. We will show also how it can be used for constructing a large family of chord diagram invariants.

4.5.2 Tutte functions for graphs

The theory of weighted graph invariants we are going to present in the next subsection is very close to the classical theory of Tutte functions. Here we give a brief description of the theory of Tutte functions for graphs. A complete description can be found in [21].

Let G be a graph. We allow graphs having loops and multiple edges. A *loop* is an edge with coinciding ends. A *link* is an edge in G that is not a loop.

Definition 4.5.2 The *deletion* of a link e is the operation on the graph G making it into the graph G_e' with the link e removed.

The *contraction* of a link e is the operation on G making it into the graph G_e'' with the link e contracted (i.e., the link e is removed from the graph, its ends become one and the same vertex). Other links connecting these two vertices become loops.

Loops may not be deleted or contracted.

Definition 4.5.3 A function f on graphs is called *a Tutte function* if it satisfies *Tutte relations*

$$f(G) = f(G_e') + f(G_e'')$$

for every graph G and each link e in it.

Example 4.5.4 It is easy to check that $(-1)^{|V(G)|}\chi(G)$, where $|V(G)|$ is the number of vertices in G and χ is the chromatic polynomial, is a Tutte function.

Each of the operations of deletion and contraction decreases the number of links in the graph at least by one. It follows that each Tutte function is completely determined by its values on graphs without links. Such a graph is the disjoint union of graphs consisting of a vertex and a certain number of loops. Tutte proved the converse statement.

Theorem 4.5.5 [21] *A Tutte function may take arbitrary values on graphs without links, and it is uniquely determined by these values.*

In the context of this chapter, the Tutte theorem states that graphs with one vertex and a number of loops "generate a bialgebra of graphs", and this bialgebra is isomorphic to the polynomial bialgebra $Z[s_1, s_2, s_3, \ldots]$ having one generator of each order. We are not going to make these statements more precise, preferring to explain all constructions for weighted graphs, which are more closely related to chord diagrams.

4.5.3 The bialgebra of weighted graphs

Assigning to every chord diagram its intersection graph, $c \mapsto \Gamma(c)$, we obtain a mapping Γ from the set of all chord diagrams to the set of graphs.

We have seen that in the module generated by all chord diagrams modulo the four term relations there is a natural bialgebra structure, and this structure is important in the study of Vassiliev invariants.

It is thus a natural idea to define a bialgebra structure in the module spanned by all graphs in such a way that the mapping Γ would become a homomorphism of bialgebras.

Now there are obvious operations on graphs that mimic the multiplication and comultiplication of chord diagrams. The product of two graphs is just their disjoint union, while the coproduct of a graph G is obtained as the sum of all terms $g' \otimes g''$, where g' and g'' are two complete subgraphs of G whose vertex sets constitute a splitting of the vertex set of G. This transforms the free module generated by all graphs into a bialgebra $\mathcal{G}$.

Of course, the bare mapping Γ defines no algebra homomorphism because it is not compatible with the four term relations. To obtain a well-defined homomorphism, we have to replace the target algebra $\mathcal{G}$ by a quotient $\mathcal{G}/S$, where S contains S_0, the ideal generated by the images of all four term relations. The problem is to find an ideal $S \supset S_0$ which allows a lucid description in terms of graphs and which is small enough so that the quotient $\mathcal{G}/S$ remains substantial.

Quite remarkably, a relation similar to the Tutte relation in the algebra freely generated by *weighted graphs* becomes homogeneous, and the graded structure of the free algebra descends to the quotient modulo these relations.

In what follows, we call this quotient algebra the *algebra of weighted graphs* and denote it by $\mathcal{W}$. The algebra of weighted graphs is an example of an ample quotient algebra of $\mathcal{M}$ having one primitive generator in every degree.

Another and probably more important application of the algebra $\mathcal{W}$ is that it yields a series of preinvariants for chord diagrams.

Definition 4.5.6 A *weighted graph* is a graph G without loops and multiple edges given together with a mapping $w : V(G) \to \mathbf{N}$ called the *weight*, which assigns a positive integer to each vertex of the graph. The *weight* $w(G)$ *of the graph* G is the sum of weights of all its vertices, $w(G) = \sum_{v \in V(G)} w(v)$.

Ordinary graphs without loops and multiple edges can be treated as weighted graphs with weights of all vertices equal to 1.

We will use two natural operations on weighted graphs with a distinguished edge: *deletion* and *contraction*.

If e is an edge of the graph G, then the new graph G'_e is obtained from G by removing the edge e. The weights of the vertices do not change. This is what we call *deletion*.

The *contraction* G''_e of an edge e is defined as follows:

1. the edge e is contracted into a vertex A of the new graph G''_e;
2. if multiple edges arise each of them is replaced by a single edge;
3. the weight $w(A)$ of the vertex A is set equal to the sum of weights of the two ends of the edge e in G; weights of other vertices do not change.

Definition 4.5.7 The *weighted element* is the expression

$$G - G'_e - G''_e,$$

where G is an arbitrary weighted graph and e is its arbitrary edge.

We set $\mathcal{W}_n$ to be the $\mathbf{Z}$-module generated by all weighted graphs of weight n modulo all weighted elements. We set $\mathcal{W} = \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \dots$ with $\mathcal{W}_0 = \mathbf{Z}$.

Each weighted graph is equivalent modulo 4-term elements to a linear combination of weighted graphs without edges.

Example 4.5.8 Subsequent deletion and contraction of edges provide the following decomposition for a given graph, see Fig. 4.17.

The question is: whether this decomposition depends on the sequence of edges chosen? It may be checked for any given graph that the answer is negative. The complete proof for all weighted graphs is given by the structure theorem in the next subsection.

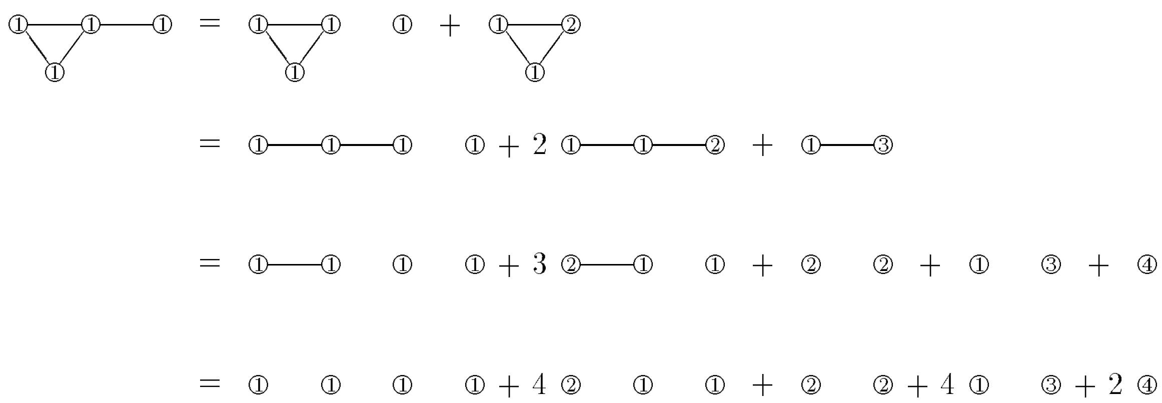


Figure 4.17: A decomposition of a given graph

Definition 4.5.9 The *bialgebra of weighted graphs* is the module $\mathcal{W}$ over $\mathbf{Z}$ with the following operations of multiplication and comultiplication:

1. multiplication
$$m : \mathcal{W}_n \otimes \mathcal{W}_m \to \mathcal{W}_{m+n}$$
comes from the disjoint union of graphs;

2. comultiplication
$$\mu : \mathcal{W}_n \to \mathcal{W}_0 \otimes \mathcal{W}_n \ \oplus \ \mathcal{W}_1 \otimes \mathcal{W}_{n-1} \ \oplus \ \ldots \ \oplus \ \mathcal{W}_n \otimes \mathcal{W}_0$$
is defined on generators as follows; consider a weighted graph G; let $J \subset V(G)$ be a subset in the vertices of G and let $J' = V(G) - J$ be its complement; we set
$$\delta([G]) = [\sum_J G_J \otimes G_{J'}].$$
The sum is taken over all subsets J, and G_J is the complete subgraph of G with vertices in J;

3. the unit is represented by the empty graph;

4. the counit is the coefficient of the empty graph in the canonical expansion of an element.

We omit a purely technical verification of the correctness of this definition.

Definition 4.5.10 A *weighted graph invariant* is a linear function on $\mathcal{W}$ with values in a ring K. A weighted graph invariant is called *multiplicative* if it is a homomorphism of ring structures.

We are going to prove further that each weighted graph invariant generates a preinvariant for chord diagrams.

4.5.4 Structure theorem for weighted graphs

Theorem 4.5.11 *The submodule of primitive elements of $\mathcal{W}_n$ is isomorphic to $\mathbf{Z}$ and freely generated by the graph (n) with one vertex of weight n. The mapping*

$$\iota : (n) \mapsto s_n$$

descends to an isomorphism of bialgebras

$$\iota : \mathcal{W} \to \mathbf{Z}[s_1, s_2, \ldots].$$

The order of the variable s_n is set to n.

Remark 4.5.12 The theorem gives a description of all weighted invariants since ι is a universal weighted invariant. This means that an arbitrary weighted invariant can be obtained from ι by substituting certain appropriate values for the variables s_k.

Proof. Put

$$\iota(G) = \sum_{\gamma} (-1)^{b_1(\gamma)} \prod_{\gamma_i} s_{w(\gamma_i)},$$

where the sum is taken over all spanning subgraphs γ of the graph G and the product is taken over all connected components γ_i of γ. Here $b_1(\gamma)$ is the first Betti number of the graph γ, i.e., the number of independent circuits in γ. (A *spanning* subgraph is a subgraph whose set of vertices coincides with that of the whole graph. Thus, the spanning subgraphs of a graph G are in one-to-one correspondence with the subsets in the set of edges of G.)

Now we extend ι by linearity to the Z-module freely generated by all weighted graphs. We claim that this mapping descends to the quotient algebra $\mathcal{W}$ and defines an isomorphism between $\mathcal{W}$ and $\mathbf{Z}[s_1, s_2, \ldots]$ (also denoted by ι in the sequel).

In order to prove the theorem we must verify that the mapping ι

1. satisfies the weighted relation;

2. is a homomorphism of bialgebras;

3. agrees with the grading in $\mathcal{W}$ and in the polynomial ring, where the degree of every s_k is k; and

4. that it is one-to-one.

1. Let G be an arbitrary weighted graph and let e be an arbitrary edge of G. We are going to prove that

$$\iota(G) = \iota(G'_e) + \iota(G''_e).$$

There exists a natural one-to-one correspondence between subgraphs of G'_e and those subgraphs of the graph G that do not contain the edge e. Thus we have

$$\sum_{\gamma'_e} t(\gamma'_e) = \iota(G'_e),$$

where the sum in the left hand side is taken over all spanning subgraphs γ'_e of the graph G that do not contain the edge e and $t(\gamma'_e)$ denotes the corresponding product.

We are going to prove that the other part of the sum for $\iota(G)$ is equal to $\iota(G''_e)$. Let γ'' be a spanning subgraph of the graph G''_e. Let b be an edge of the graph γ'' that is twice covered by the edges of graph G in the process of contraction. The preimage of b under the contraction thus consists of two edges that form a triangle together with the edge e. A spanning subgraph γ of G which is contracted to γ'' and contains the edge e may either contain any of the two preimages of the edge b or both of them (see Fig. 4.18).

So the spanning subgraph γ'' of G''_e corresponds to 3^k spanning subgraphs of G containing e, where k is the number of twice covered edges in γ''.

Mark an edge in each triangle in G that contains e. We have three possibilities for the preimage of each twice covered edge. Two of these possibilities

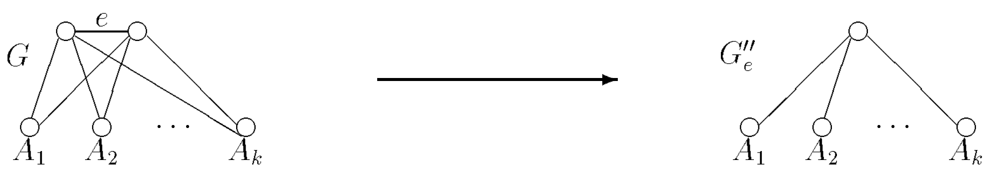


Figure 4.18: Triangles double covering edges in the graph with contracted edge

correspond to one and the same first Betti number of the covering graph. The third one gives a graph whose first Betti number differs from this value by one. Thus the products for two of the three possibilities mutually cancel, and we obtain a one-to-one correspondence between spanning subgraphs of G''_e and spanning subgraphs of G that contain e and a marked edge in each triangle containing e. So we have

$$\sum_{\gamma_e} t(\gamma_e) = \iota(G''_e),$$

where the sum is taken over all spanning subgraphs of G containing the edge e.

2. The multiplicativity of ι follows from a straightforward calculation.

3. This is evident if the grading of every s_k is set to k.

4. Since the polynomial algebra is free, the assignment $s_n \mapsto \textcircled{\scriptsize n}$ defines a homomorphism $\mathbf{Z}[s_1, s_2, s_3, \ldots] \to \mathcal{W}$ which is obviously inverse to ι.

4.5.5 Constructing Vassiliev invariants from weighted invariants

Definition 4.5.13 A function f defined on the set of weighted graphs of weight n and satisfying the weighted relation

$$f(G) - f(G') - f(G'') = 0$$

will be called a *weighted invariant of order* n.

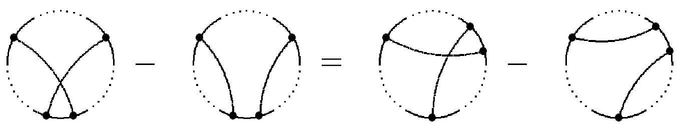
In other words, a weighted invariant of order n is just an element of $\mathcal{W}_n^*$.

Let c be a chord diagram. Its intersection graph $\Gamma(c)$ can be regarded as a weighted graph with all weights set to 1. This yields a homomorphism of free $\mathbf{Z}$-modules.

Theorem 4.5.14 *The mapping Γ descends to a homomorphism of bialgebras $\mathcal{M} \to \mathcal{W}$. The dual mapping $\Gamma^* : \mathcal{W}^* \to \mathcal{M}^*$ determines a subbialgebra of weighted preinvariants in the bialgebra of preinvariants.*

Proof.

We must prove that the mapping Γ satisfies the four-term relation. Write this relation in the form:



Denote the intersection graph of the first term in the left hand side by G_1. It contains two vertices A and B_1 corresponding to the two selected chords. These vertices are connected by an edge. The intersection graph of the second term is obtained from G_1 by deleting the edge AB_1. We denote this graph by $G_1^{'}$.

In the algebra $\mathcal{W}$ the following relation holds:

$$G_1 - G_1' = G_1'',$$

where G_1'' is the result of contracting the edge AB_1 in G_1.

In the right hand side we similarly have:

$$G_2 - G_2' = G_2'',$$

where G_2'' is obtained by contracting the edge AB_2 in G_2.

To accomplish the proof, it is sufficient to show that the two weighted graphs G_1'' and G_2'' coincide. In fact, there exists a natural one-to-one correspondence between the two sets of vertices $V(G_1'')$ and $V(G_2'')$. This correspondence preserves the weight function (the weight is equal to 1 for every vertex but one, for which it equals two). The edges of the two graphs are also in a one-to-one correspondence. Two chords different from A, B_1, B_2 intersect in G_1'' if and only if they intersect in G_2''. Suppose a chord C in G_1

intersects B_1 but does not intersect B_2 in G_2; this means that C intersects A in both diagrams. Thus the vertex C in both graphs G_1'' and G_2'' is connected to the new vertex by an edge, and we obtain a natural one-to-one correspondence between the edges of graphs G_1'' and G_2'' as well.

4.5.6 The forest subalgebra

Among primitive elements in the module $\mathcal{M}_n$ of chord diagrams of order n some are "simpler" than the others. These simplest primitive elements come from "simpler" chord diagrams. The most simple chord diagrams that are not products of diagrams of less order, are "tree diagrams".

Definition 4.5.15 A chord diagram is called *a tree diagram*, iff its intersection graph is a tree. The subalgebra $\mathcal{F}$ in $\mathcal{M}$ generated by all tree diagrams is called *the forest subalgebra.*

We present the following statement without proof.

Theorem 4.5.16 *The forest subalgebra $\mathcal{F} \subset \mathcal{M}$ is isomorphic to the bialgebra of weighted graphs, and the mapping*

$$\Gamma : \mathcal{F} \to \mathcal{W}$$

provides a natural isomorphism of the bialgebras.

4.6 Constructing invariants from associative algebras

4.6.1 Free associative algebras

Let $\mathcal{B}_m =< x_1, \dots, x_m >$ be the free associative algebra generated by $x_1, \dots, x_m$ as free generators. We are going to associate with each positive integer m a preinvariant b_m. The ring of values of the preinvariant will be described later.

Consider a mapping $\hat{b}_m : \mathcal{A}_n \to \mathcal{B}_m$ of the module of arc diagrams of order n into the algebra $\mathcal{B}_m$. The mapping $\hat{b}_m$ is constructed as follows.

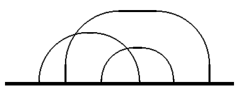
Let $a \in \mathcal{A}_n$ be an arc diagram. Number each arc of the diagram a with one of the numbers $1, ..., m$. Denote such a numbering by $\pi : V(a) \to \{1, ..., m\}$ ($V(a)$ is the set of chords in a). We associate with the diagram a and the numbering π an element $\bar{b}_m(a, \pi)$: by writing the letter $x_{\pi(v)}$ on both ends of each arc $v \in V(a)$ and reading these letters from left to right. The word thus obtained will be the element $\bar{b}_m(a, \pi)$.

We set

$$\hat{b}_m(a) = \sum_{\pi} \bar{b}_m(a, \pi).$$

The sum is taken over all m^n possible numberings $\pi : V(a) \to \{1, \ldots, m\}$.

For example, for $m = 2$ and the diagram



one has

$$\hat{b}_2(a) = x_1^6 + x_1^2 x_2 x_1 x_2 x_1 + x_1 x_2 x_1^3 x_2 + x_2 x_1^2 x_2 x_1^2$$
$$+ x_1 x_2^2 x_1 x_2^2 + x_2 x_1 x_2^3 x_1 + x_2^2 x_1 x_2 x_1 x_2 + x_2^6.$$

Remark 4.6.1 Obviously, the image of $\hat{b}_m$ lies in the symmetric part of $\mathcal{B}_m$.

Introduce a scalar product on the linear part of the algebra $\mathcal{B}_m$ simply by setting $(x_i, x_j) = \delta_{ij}$. Thus the orthogonal group $O(m)$ acts on the linear part of the algebra $\mathcal{B}_m$. This action may be naturally extended to an action of $O(m)$ on the algebra $\mathcal{B}_m$ itself. The action preserves the grading of elements in $\mathcal{B}_m$.

Theorem 4.6.2 *The mapping $\hat{b}_m$ is invariant under the action of the orthogonal group $O(m)$ on $\mathcal{B}_m$.*

Proof. The group $O(m)$ is generated by subgroups $O(2)$ acting by rotating coordinate planes. That is why it is sufficient to prove the Theorem for the action of the group $O(2)$ on a pair of letters x_1, x_2, i.e., it is sufficient to prove the Theorem only for the mapping $\hat{b}_2 \colon \mathcal{A} \to \mathcal{B}_2$.

Now let a be an arc diagram. The image $\hat{b}_2(a)$ is a sum of monomials. Consider the $O(2)$-change of variables $x_1 = ax_1' + bx_2'$, $x_2 = cx_1' + dx_2'$. We need to check that each monomial of the type $(x_1')^{\alpha_1}(x_2')^{\beta_1}\dots$ appears in the transformed image with coefficient 1 if and only if the monomial $(x_1)^{\alpha_1}(x_2)^{\beta_1}\dots$ appears in $b_2(a)$. Otherwise the monomial must have coefficient 0.

That is indeed the case. If there is no monomial of the above type in $\hat{b}_2(a)$, then there exists a pair $\dots x_1' \dots x_2' \dots$ on two ends of one and the same arc. Such a monomial comes from two types of monomials $\dots x_1 \dots x_1 \dots$ and $\dots x_2 \dots x_2 \dots$ after substitutions, and we take ax_1' times dx_2' from the first monomial and cx_1' times bx_2' from the second monomial. Thus the coefficient is a multiple of $ad + bc = 0$ since our matrix is an orthogonal one.

Similarly, if there exists a monomial of the above type in the original image $\hat{b}_2(a)$, then all letters x_1' as well as x_2' are splitted into pairs corresponding to the ends of the arc. Each pair $\dots x_1 \dots x_1 \dots$ introduces a multiplier in the coefficient of the type $(a^2 + b^2) = 1$, while the pair $\dots x_2 \dots x_2 \dots$ provides a multiplier $(c^2 + d^2) = 1$ and the product equals 1.

The mapping $\hat{b}_m$ descends to a linear mapping (which we also denote by $\hat{b}_m$) $\hat{b}_m : \mathcal{A} \to \mathcal{B}_m$. The image $\hat{b}_m(\mathcal{A})$ forms a subalgebra in $\mathcal{B}_m$. The image $\hat{b}_m(\mathcal{AF}) \subset \hat{b}_m(\mathcal{A})$ forms a two-sided ideal.

Theorem 4.6.3 *The quotient mapping* $b_m : \mathcal{M} = \mathcal{A}/\mathcal{AF} \to \hat{b}_m(\mathcal{A})/\hat{b}_m(\mathcal{AF})$ *is a preinvariant.*

The proof follows directly from the definition.

Of course, having an arbitrary associative algebra B with a system $x_1, \dots, x_m$ of generators and a quotient mapping $\mathcal{B}_m \to B$ one may construct a composite preinvariant with values in an appropriate quotient of a subalgebra in B.

The preinvariant b_m is a complicated one for calculations. Each of the preinvariants b_m is more complicated than the previous one due to the natural embedding $\mathcal{B}_m \subset \mathcal{B}_{m+1}$. The sequence of mappings b_m carries the most important information from a chord diagram.

4.6.2 Universal enveloping algebras for Lie algebras

A special case of the construction in the previous subsection is given by the universal enveloping algebra of a Lie algebra.

Let H be a Lie algebra. For simplicity we restrict ourselves to complex Lie algebras. Consider a basis $\{x_1, \ldots, x_m\} \subset H$. The mapping $\hat{b}_m : \mathcal{A} \to \mathcal{B}_m$ descends to a mapping $b_m^H : \mathcal{A} \to UH$ into the universal enveloping algebra UH of H. The mapping b_m^H depends, of course, on the choice of the basis in H.

Suppose now that H admits a nondegenerate bilinear bi-invariant form $(\cdot,\cdot)$. The word "bi-invariant" means that $([x,y],z) = (x,[y,z])$ for each triple $x,y,z \in H$. Let $\{x_1, \ldots, x_m\}$ be an orthonormal basis with respect to this form, i.e., $(x_i, x_j) = \delta_{ij}$.

Theorem 4.6.4 *When $\{x_1, \ldots, x_m\}$ is an orthonormal basis in H we have*

1. $\mathcal{AF} \subset \operatorname{Ker} b_m^H$;

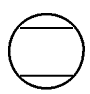
2. $\operatorname{Im} b_m^H$ *is a subset of the center* $ZUH \subset UH$;

3. b_m^H *does not depend on the choice of the basis.*

Thus, having a Lie algebra H with nondegenerate bilinear bi-invariant form, the two mappings $\hat{b}_m^H$ and b_m^H coincide and b_m^H determines a preinvariant with values in the center of the universal enveloping algebra of H.

Example 4.6.5 For the only chord diagram with one chord, the value of the pre-invariant b_m^H is $x_1^2 + x_2^2 + \ldots + x_m^2$. The last element in ZUH is called *the Casimir element* for arbitrary Lie algebra H; we will denote it by C_H.

There are two chord diagrams of order 2. One has

$b_m^H($  $) = C_H^2$

since this diagram is the square of the one-chord diagram.

This chord diagram has two different arc representations. These two different representations give the relation:

$$x_1^2x_2^2 + x_1^2x_3^2 + \ldots = x_1x_2^2x_1 + x_1x_3^2x_1 + \ldots \tag{4.3}$$

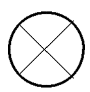
Since we have

$$\begin{aligned} x_i^2x_j^2 - x_ix_j^2x_i + x_j^2x_i^2 - x_jx_i^2x_j &= x_i^2x_j^2 - x_ix_jx_ix_j + x_ix_jx_ix_j - x_ix_j^2x_i \\ &\quad + x_j^2x_i^2 - x_jx_ix_jx_i + x_jx_ix_jx_i - x_jx_i^2x_j \\ &= x_i[x_i, x_j]x_j - x_j[x_i, x_j]x_i \\ &\quad + x_ix_j[x_i, x_j] - x_jx_i[x_i, x_j] \end{aligned} \tag{4.4}$$

it follows that relation (4.3) can be rewritten as

$$\sum_{i \neq j} x_i[x_i, x_j]x_j - x_j[x_i, x_j]x_i = -\sum_{i \neq j}[x_i, x_j]^2.$$

The last equality allows one to calculate the value of the invariant b_m^H on the second diagram of order 2 for arbitrary Lie algebra H:

$b_m^H($  $) = \sum x_ix_jx_ix_j$

and

$$\begin{aligned} \sum x_ix_jx_ix_j &= \sum x_ix_jx_ix_j - \sum x_i^2x_j^2 + \sum x_i^2x_j^2 \\ &= -\sum_{i \neq j} x_i[x_i, x_j]x_j + C_H^2 \\ &= \sum_{i \neq j}[x_i, x_j]^2 + C_H^2 \end{aligned} \tag{4.5}$$

The simplest noncommutative Lie algebra with a nondegenerate bilinear form is the simple Lie algebra $\mathrm{sl}_2(\mathbf{C})$. The algebra is generated by three elements x, y, z such that $[x, y] = z, [y, z] = x, [z, x] = y$. The bilinear form is given by the relations $(x, x) = (y, y) = (z, z) = 1; (x, y) = (y, z) = (z, x) = 0$.

The center of the universal enveloping algebra coincides with the polynomial algebra in the Casimir element, $C_{\mathrm{sl}_2} = x^2 + y^2 + z^2$. Thus we obtain a preinvariant with values in the ring of polynomials in one variable. The value of (4.5) becomes $C_{\mathrm{sl}_2}^2 + C_{\mathrm{sl}_2}$.

Corollary 4.6.6 *Let $R : H \to \mathrm{GL}_N$ be a representation of the Lie algebra H, $\hat{R} : UH \to \mathrm{GL}_N$ be the corresponding representation of the universal enveloping algebra,* $\mathrm{Tr} : GL_N \to \mathbf{C}$ *be the trace mapping. Then the composition mapping $\frac{1}{N}\,\mathrm{Tr} \circ \hat{R} \circ b_m^H : \mathcal{A} \to \mathbf{C}$ provides a preinvariant with values in $\mathbf{C}$.*

Proof of the theorem.

1. Fix a numbering on the set of arcs that do not participate in the 4-term relation. The part of the 4-term element value corresponding to this numbering looks like

$$\begin{aligned}\Delta = &\sum_{i,j} w_0 x_i x_j w_1 x_i w_2 x_j w_3 - \sum_{i,j} w_0 x_j x_i w_1 x_i w_2 x_j w_3 \\ &- \sum_{i,j} w_0 x_i w_1 x_j x_i w_2 x_j w_3 + \sum_{i,j} w_0 x_i w_1 x_i x_j w_2 x_j w_3 \qquad (4.6)\end{aligned}$$

where w_0, w_1, w_2, w_3 are fixed words in the alphabet $\{x_1, \dots, x_m\}$ corresponding to the given numbering of the arcs. Transforming the expression (4.6) one obtains

$$\begin{aligned}\Delta &= \sum_{i,j} w_0 [x_i, x_j] w_1 x_i w_2 x_j w_3 + \sum_{i,j} w_0 x_i w_1 [x_i, x_j] w_2 x_j w_3 \\ &= \sum_{i,j,k} C_{i,j}^k w_0 x_k w_1 x_i w_2 x_j w_3 + \sum_{i,j,l} C_{i,j}^l w_0 x_i w_1 x_l w_2 x_j w_3\end{aligned}$$

or, changing indices $i \to l, k \to i$ in the first sum,

$$\begin{aligned}\Delta &= \sum_{i,j,l} C_{l,j}^i w_0 x_i w_1 x_l w_2 x_j w_3 + \sum_{i,j,l} C_{i,j}^l w_0 x_i w_1 x_l w_2 x_j w_3 \\ &= \sum_{i,j,l} (C_{l,j}^i + C_{i,j}^l) w_0 x_i w_1 x_l w_2 x_j w_3 \\ &= 0\end{aligned}$$

Here $C_{i,j}^k$ are the structure constants for the basis $\{x_1, \ldots, x_m\}$, i.e., $[x_i, x_j] = \sum_k C_{ij}^k x_k$ in the Lie algebra H. The last sum equals zero since for an orthonormal basis structure constants are anticommutative under exchange of upper and left lower indices: $C_{l,j}^i = -C_{i,j}^l$.

2. We want to prove that $x_j b_m^H(a) = b_m^H(a) x_j$ for each j, say for $j = 1$, and each arc diagram a. We will proceed step by step moving the letter x_1 through the element $b_m^H(a)$. Fix an arc and consider "the contribution" of this arc:

$$\begin{aligned}
\Delta(a) &= \sum_i w_0 x_1 x_i w_1 x_i w_2 - \sum_i w_0 x_i x_1 w_1 x_i w_2 \\
&\quad - \sum_i w_0 x_i w_1 x_1 x_i w_2 + \sum_i w_0 x_i w_1 x_i x_1 w_2 \\
&= \sum_{i,k} C_{i,1}^k w_0 x_k w_1 x_i w_2 + \sum_{i,k} C_{k,1}^i w_0 x_i w_1 x_k w_2 \\
&= 0
\end{aligned}$$

Thus, the contribution of each arc equals zero, as required.

3. Follows from Theorem 4.6.3.

4.7 Vassiliev invariants of links

4.7.1 Chord diagrams for links = maps

4.7.2 Bialgebra of maps

4.8 Historical and bibliographic remarks

We do not present a full bibliography. Further references may be found in [3], for example. The reader should take into account that the first version of [3] has been widely distributed since August 1992.

Vassiliev invariants for knots were introduced by V.A.Vassiliev in 1989 [22]. The construction is a special case of a general construction describing the topology of the complement to the discriminant. In this special case one considers the space of all smooth mappings $u: S^1 \to \mathbf{R}^3$. The discriminant is formed by mappings with singularities. The compliment to the discriminant is a set of disjoint components. The components are in one-to-one correspondence with knots.

Similar construction may be applied to wide class of situations, including links, high-dimensional knots, spaces of polynomials, and so on, see Vassiliev's book [24].

For the situation of knots Vassiliev introduced, in fact, chord diagrams and wrote out 1-term and 4-term relations.

The theory became popular after it was explained by V.I.Arnold (see [2]).

The proof of the fact that most of previously known knot invariants can be expressed in terms of Vassiliev knot invariants has been accomplished by J.Birman and X.S.Lin (see [6]) and D.Bar-Natan (see [3]).

In [16] M.Kontsevich constructed a universal Vassiliev invariant with the help of an integral (the Kontsevich integral). It has been proved thus, that 1-term and 4-term relations are determining relations for Vassiliev knot invariants over a field of characteristic zero. See also the description of the Kontsevich integral in [3]. Kontsevich stated in [16] that the coefficients of the integral are rational numbers, but the detailed proof was first accomplished in [17].

The Hopf algebra structure on the module of chord diagrams was introduced by M.Kontsevich and D.Bar-Natan (see [3]).

The notion of the intersection graph of a chord diagram is an old one (see [14]). It appeared in the context of Vassiliev invariants in [8, 9, 10]. The bialgebra of weighted graphs, the homomorphism from the bialgebra of chord diagrams into the bialgebra of weighted graphs, the structure theorem for the bialgebra of weighted graphs are introduced in [10].

D. Bar-Natan [3] was the first one who constructed preinvariants from Lie algebras with degenerate bilinear bi-invariant form and given representation. The construction was then simplified and generalized by M.Kontsevich [16].

The papers [2, 4, 7, 8, 19] are also well suited for the first acquaintance with the subject. For more sophisticated topics we refer the reader to [3, 16, 24].

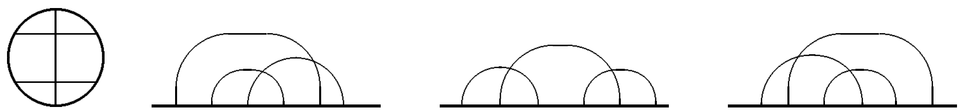


Figure 4.19: Indecomposable (left) and decomposable (right) arc diagrams

4.9 Appendix: enumeration of arc diagrams

The problem of enumerating *chord* diagrams is complicated. Its solution would include the investigation of the action of the rotation group and Polya's theory machinery. In fact, what we need is the enumeration of chord diagrams modulo 4-term elements (see in this section below). But such an enumeration is still a problem. This subsection presents, however, a complete and self–contained enumeration of "indecomposable" arc diagrams.

The problem of enumerating *all* the *arc* diagrams is trivial. An obvious calculation gives the number of different arc diagrams with n arcs to be equal to $(2n-1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)$, i.e., the number of splittings of $2n$ points into pairs.

Definition 4.9.1 An arc diagram is called *decomposable* if the set of its arcs may be split into a disjoint union of two nonempty subsets such that no arc from the first subset intersects an arc from the second subset. Otherwise the diagram is called *indecomposable*.

Let $a_{2n} = (2n-1)!!$ be the number of all arc diagrams of order n and

$$a(t) = a_0 + a_2t^2 + a_4t^4 + \ldots = 1 + t^2 + 3t^4 + 15t^6 + \ldots$$

be the corresponding generating function. Denote by

$$n(s) = n_0 + n_2s^2 + n_4s^4 + \ldots = 1 + s^2 + s^4 + 4s^6 + 27s^8 + \ldots$$

the generating function for indecomposable diagrams.

Statement 4.9.2 *Generating functions $a(\cdot)$ and $n(\cdot)$ are related by the Lagrange equality*

$$a(t) = n(ta(t)). \tag{4.7}$$

Remark 4.9.3 According to the Lagrange theorem, relation (4.7) provides an explicit way to calculate the number of indecomposable arc diagrams of arbitrary order. Here are a few first values:

k	0	1	2	3	4	5	6	7	8	9
n_k	1	1	1	4	27	248	2,830	38,232	593,859	10,401,712

One may easily obtain a direct proof of the statement, but we prefer to include it in general framework of framed sequences. In fact, the theorems of this type appear in various situations explaining thus the importance of the Lagrange theorem.

The proof is based on a general theorem on relations between the generating function for "framed sequences" and the generating function for "indecomposable framed sequences".

Definition 4.9.4 *A framed n-sequence* θ is the set $N_n = \{1, 2, \ldots, n\}$ with the permutation group S_θ acting on N_n (i.e. a pair $\theta = \{N_n, S_\theta\}$).

Let Θ_n be the set of all framed n-sequences, $\Theta = \Theta_0 \sqcup \Theta_1 \sqcup \Theta_2 \sqcup \ldots$ being the set of all framed sequences.

An insertion of a framed n-sequence θ *into* a framed m-sequence η *at the place* k is the framed $(n+m)$-sequence $\xi = \theta \triangleright_k \eta \in \Theta_{n+m}$ obtained by including n elements after the element k of the set N_m, with the groups S_θ and S_η acting independently on corresponding subsets.

Let $L \subset \Theta$ be a set of framed sequences.

Definition 4.9.5 An element $\theta \in L$ is called *indecomposable with respect to L*, if it cannot be represented as $\theta = \theta_1 \triangleright_k \theta_2$ with both θ_1, θ_2 nonempty, $\theta_1, \theta_2 \in L$.

A set L is called *closed* (under insertion) iff

1. $\emptyset \in L$;

2. For any $\theta \in L, \eta \in L \cap \Theta_n$ the framed set $\theta \triangleright_k \eta$ is contained in L for all $k = 0, \ldots, n$.

3. If $\theta \in L$ and θ is of the form $\theta = \theta_1 \triangleright \theta_2, \theta_1 \in L$, then $\theta_2 \in L$.

Let now l_k denote the number of framed k-sequences in L and $l(t) = l_0 + l_1 t + l_2 t^2 + \ldots$ be the corresponding generating function. Let n_k denote the number of indecomposable framed k-sequences in L and $n(s) = n_0 + n_1 s + n_2 s^2 + \ldots$ be the corresponding generating function.

Theorem 4.9.6 *If L is closed under insertion, then functions $l(\cdot)$ and $n(\cdot)$ are related by the Lagrange equality:*

$$l(t) = n(tl(t)).$$

Proof. Each decomposable framed sequence from L has a decomposition into a set of indecomposable framed sequences from L. Such a decomposition is unique.

Indeed, let us take a framed sequence and two indecomposable subsequences in it. These two subsequences cannot overlap. Otherwise their intersection would be a framed subsequence in each of them, which means that at least one of them would be decomposable.

Now, for each framed sequence θ from L an indecomposable framed subsequence ξ containing the element 1 is well defined. Thus, each framed sequence $\theta \in L$ may be obtained from an indecomposable sequence $\xi \in L$ by inserting some sequences from L after the first, the second, ... elements of ξ. This means that framed sequences from L with the first indecomposable sequence of length k are enumerated by the function $n_k t^k l^k(t)$. Summing over all k we obtain the required statement.

Now in order to prove the theorem for arc diagrams, we must make a framed sequence from an arc diagram. Each arc diagram of order n becomes a framed $2n$-sequence if we mark points on the line by the numbers $1, 2, \ldots, 2n$ and generate the group of substitutions by all transpositions, exchanging ends of each arc. The set of all arc diagrams is obviously closed under insertions, which ends the proof.

We present without proof one more formula for the sequence $\{n_k\}$.

Statement 4.9.7 *The sequence $\{n_k\}$ of the numbers of indecomposable arc diagrams satisfies the following recurrent relation:*

$$n_{k+1} = k(s_0 s_k + s_1 s_{k-1} + \ldots + s_k s_0).$$

4.10 Appendix: Kontsevich integral

The Kontsevich's proof of the theorem is based on the construction of the universal invariant. He introduced a mapping taking each knot to an element of the bialgebra $\mathcal{N}$, bialgebra of chord diagrams modulo 1-term and 4-term relations. We denote this mapping by $\tilde{Z} : k \mapsto \tilde{Z}(k) \in \mathcal{N}$. This mapping is a universal Vassiliev invariant in the following sense: for any multiplicative Vassiliev knot invariant with values in a commutative associative ring K with unit there exists a ring homomorphism $\mathcal{N} \to K$ such that the diagram

$$\begin{array}{lll} \mathcal{K} & & \\ \downarrow \tilde{Z} & \searrow & \\ N & \longrightarrow & K \end{array}$$

is commutative.

4.10.1 Construction of the Kontsevich integral

Let us represent the 3-dimensional real space $\mathbf{R}^3$ as the direct product $\mathbf{R}^3 \cong \mathbf{R} \times \mathbf{C}$. We think of the real line as the vertical line and of the complex line as the horizontal plane. Consider a knot $k : S^1 \to \mathbf{R}^3$. The horizontal projection of the image of the knot on the real line determines the *height function* $h : S^1 \to \mathbf{R}$ on the knot. The vertical projection determines the (complex) coordinate $z \in \mathbf{C}$.

Definition 4.10.1 The knot k is called a *Chebyshev knot* if the restriction of the height function h on k has only nondegenerate critical points, and only two different critical values (all the maxima are mutually equal and all the minima are mutually equal).

For a Chebyshev knot k let $t_{\min}, t_{\max}$ denote the maximum and the minimum of the height function on k. Choose n points on the segment $[t_{\min}, t_{\max}]$, $t_{\min} < t_1 < t_2 < \ldots < t_n < t_{\max}$.

Let x_i, y_i be two different points on the knot $k(S^1)$ with the height $h(x_i) = h(y_i) = t_i$. If t_i has $2m$ preimages on the knot k then there are $m(2m-1)$ ways to choose such a pair. The set of pairs $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ determines a chord diagram of order n, where the ends of the i-th chord are given by the preimages $(k^{-1}(x_i), k^{-1}(y_i))$. Thus, such a choice determines an element $c((x_1, y_1), \ldots, (x_n, y_n)) \in \mathcal{N}_n$. This element is continiously extended to close noncritical values of $(t_1, \ldots, t_n)$.

For a set $t_{\min} < t_1 < t_2 < \ldots < t_n < t_{\max}$ consider the element $C(t_1, \ldots, t_n) \in \mathcal{N}_n$ equal to the alternating sum

$$C(t_1, \ldots, t_n) = \sum (-1)^{u(c)} c((x_1, y_1), \ldots, (x_n, y_n)) \in \mathcal{N}_n,$$

where the sum is taken over all $m(2m-1)$ possible ways to choose the set of pairs (x_i, y_i). Here $u(c)$ is the number of “decreasing points”, i.e., the number of looking down tangent vectors to the knot k in the points x_i, y_i.

Denote z-coordinates of the points x_i, y_i by z_i, z_i' correspondently.

Definition 4.10.2 The *Kontsevich integral of order* n is the expression

$$Z_n(k) = \int C(t_1, \ldots, t_n) \frac{d(z_1 - z_1')}{z_1 - z_1'} \ldots \frac{d(z_n - z_n')}{z_n - z_n'}. \tag{4.8}$$

where the integral is taken over the set of all n-tuples $t_1 < \ldots < t_n$ of regular values of the height function.

Theorem 4.10.3 *1. The integral $Z_n(k)$ is well defined, i.e., it absolutely converges for any Chebyshev knot k.*

2. The element $Z_n(k) \in \mathcal{N}_n$ is invariant under transformations of the knot k in the class of Chebyshev knots. (Such a transformation does not change the number of critical points of k).

Set now $Z_0(k) = 1 \in \mathcal{N}_0$, $Z(k) = \sum_{n=0}^{\infty} Z_n(k)$. Remark that there is no problem with the convergence of this sum since each term belongs to its own grading.

Definition 4.10.4 *The* Kontsevich invariant *is the fraction*

$$\tilde{Z}(k) = \frac{Z(k)}{Z^{q/2}(k_0)},$$

where k_0 is a trivial embedding with 4 critical points, and $q/2$ is half the number of critical points in the knot k.

Remark 4.10.5 For $n = 0$ and any knot k the zero-order part $Z_0(k)$ is the unit element in the bialgebra $\mathcal{N}$. The expression $1/Z(k) \in \mathcal{N}$ thus makes sense.

Denote by $\tilde{Z}_n(k)$ the natural projection of $\tilde{Z}(k)$ to the submodule $\mathcal{N}_0 \oplus \ldots \mathcal{N}_n \subset \mathcal{N}$.

Theorem 4.10.6

1. *$\tilde{Z}_n(\cdot)$ is a Vassiliev invariant of order no greater than n with values in $\mathcal{N}_0 \oplus \ldots \oplus \mathcal{N}_n$.*

2. *The invariant $\tilde{Z}_n(\cdot)$ is the universal Vassiliev invariant of order no greater than n, i.e., for any knot invariant $v_n \in \mathcal{V}_n$ with values in a ring K there exists a homomorphism $g : \mathcal{N} \to K$ such that the diagram*

$$\begin{array}{lll} \mathcal{K} & & \\ \downarrow \tilde{Z}_n & \searrow & \\ N & \xrightarrow{g} & K \end{array}$$

is commutative.

3. *Considered over a field $\mathbf{k}$ of characteristic zero, the images of the Kontsevich invariant generate $\mathcal{N}$.*

4. *For any knot k the coefficients of the Kontsevich integral $Z(k)$ are rational numbers.*

5. *The element $\tilde{Z}_n(k) \in \mathcal{N}$ is a group-like element of the bialgebra, i.e., $\mu(\tilde{Z}_n(k)) = \tilde{Z}_n(k) \otimes \tilde{Z}_n(k)$.*

4.10.2 The Knizhnik-Zamolodchikov connection and the Kontsevich integral

Iterated integrals

The iterated integral form of the universal invariant is not by chance. Iterated integrals naturally appear as solutions of ordinary differential equations.

Consider a smooth manifold M, a 1-form ω on M and a path $x : [a,b] \to M$. Let h be the function $h : [a,b] \to \mathbf{R}$ satisfying the following differential equation

$$\frac{dh}{dt} = \omega(\dot{x}(t))h(t), \qquad t \in (a,b)$$

and the initial condition $h(a) = 1$. Then h may be represented as the iterated integral

$$h(t) = 1 + \sum_n \int_{a<t_1<\ldots<t_n<t} \omega(x(t_n)) \wedge \ldots \wedge \omega(x(t_1)), \qquad t \in [a,b].$$

Indeed, differentiating the last expression we obtain the same expression times $\omega(\dot{x}(t))$.

In the case when ω is a 1-form with values in associative algebra K the solution h of the equation is as well the function with values in K. The *curvature* F_ω of the *connection* ω is the 2-form $F_\omega = d\omega + \omega \wedge \omega$. In the case when the connection is flat (i.e. $F_\omega = 0$) the *holonomy* $h(t) = h(x,\omega)$ depends only on the point $x(t)$, but not on the path x (in a given homotopy class).

The formal Knizhnik–Zamolodchikov connection

Let $\tilde{\mathbf{C}}^n = \mathbf{C}^n - \Delta$ be the space of n-tuples of pairwise different complex numbers. The *Knizhnikov–Zamolodchikov connection* ω_n is a connection on $\tilde{\mathbf{C}}^n$ with values in an associative algebra $\mathcal{N}^{n,n}$.

Consider $2n$ vertical lines; the first n lines are oriented upwards, the other n lines are oriented downwards. A *chord* is a segment connecting two of these lines; a $2n$*-strands chord diagram* is a finite set of chords (see Figure). "Composition" determines the natural multiplication of two chord diagrams.

The algebra $\mathcal{N}^{n,n}$ is the algebra freely generated by $2n$-strands chord diagrams modulo natural 4-term relations.

Denote by ω_{ij} the chord diagram with one chord connecting the i-th and the j-th strands. We set

$$\omega_n = \sum_{1\le i<j\le n} s_i s_j \omega_{ij} \frac{d(z_i - z_j)}{z_i - z_j},$$

where $s_i = 1$ if the i-th strand looks upwards, and $s_i = -1$ otherwise.

Proposition 4.10.7 *The formal Knizhnik–Zamolodchikov connection is flat.*

Proof. The form ω_n is exact, therefore $d\omega_n = 0$. Further, we have

$$\omega_n \wedge \omega_n = \sum_{i<j;i'<j'} \omega_{ij}\omega i'j' \frac{d(z_i - z_j)}{z_i - z_j} \wedge \frac{d(z_{i'} - z_{j'})}{z_{i'} - z_{j'}}.$$

The entries of the last sum may be groupped into three groups.

If the pair $\{i,j\}$ coincides with the pair $\{i',j'\}$ then the external product equals zero.

If the intersection of $\{i,j\}$ with $\{i',j'\}$ is empty then ω_{ij} and $\omega_{i'j'}$ commute and the pair

$$\omega_{ij}\omega i'j' \frac{d(z_i - z_j)}{z_i - z_j} \wedge \frac{d(z_{i'} - z_{j'})}{z_{i'} - z_{j'}}.$$

and

$$\omega_{i'j'}\omega ij \frac{d(z_{i'} - z_{j'})}{z_{i'} - z_{j'}} \wedge \frac{d(z_i - z_j)}{z_i - z_j}.$$

The last case is the most interesting one. Suppose we consider the strands $1,2,3$ and all of them look upwards (the general case is treated similarly). We are going to prove that the sum of six entries

$$\omega_{12}\omega_{23} \frac{d(z_1 - z_2)}{z_1 - z_2} \wedge \frac{d(z_2 - z_3)}{z_2 - z_3} + (permutations) = 0.$$

Indeed, the last sum is equal to

$$(\omega_{12}\omega_{23} - \omega_{23}\omega_{12} \frac{d(z_1 - z_2)}{z_1 - z_2} \wedge \frac{d(z_2 - z_3)}{z_2 - z_3} + (cyclicpermutations) = (\omega_{12}\omega_{23} - \omega_{23}\omega_{12})(\frac{d(z_1 - z_2)}{z_1 - z_2} \wedge \frac{d(z_2 - z_3)}{z_2 - z_3} + \frac{d(z_2 - z_3)}{z_2 - z_3} \wedge \frac{d(z_3 - z_1)}{z_3 - z_1} + \frac{d(z_3 - z_1)}{z_3 - z_1} \wedge \frac{d(z_1 - z_2)}{z_1 - z_2 3})$$

since all the coefficients are equal to one another because of the 4-term relations. The last form equals zero thet may be checked by direct computation (the Arnold's identity).

4.10.3 Proofs

Proof of Theorem 1

The singularities in the integral 4.8 arise in the points, where $z_i - z'_i$ is close to zero. There are two different possibilities.

If there are no other chord ends on the small segment $[x_i, y_i] \subset S^1$ then the chord (x_i, y_i) does not intersect any other chord. In this case the element $c((x_1, y_1), \ldots, (x_n, y_n)) \in \mathcal{N}_n$ equals zero.

If there is another chord end on the segment $[x_i, y_i]$ then the integral 4.8 may be upper estimated by the converging integral

$$\int dt_1 \frac{dt_2}{t_2} \ldots \frac{dt_n}{t_n}.$$

The last integral is converging since one may take it first over the variable t_1.

Consider now a deformation of the knot k preserving the number of critical points. Without lost of generality we may suppose that the critical points are fixed. The knot may be then

Proof of Theorem 2

1. Proof first that the function $\tilde{Z}$ does not change its value while adding a pair of new critical points. Indeed, additional chords may be either

   "long" or "short" ones. The long chords are the chords with one end on a new strand and the other end on an old strand. The short chords are chords with both ends on new strands.

   Suppose a chord diagram contains a long chord. Let t_k be the maximal value of the height function corresponding to a long chord, x_k be the end on the old strand, y_k be the end on the new strand. The chord diagram with y_k changed for the other new strand is the same chord diagram, and its coefficient in the Kontsevich integral has the opposite sign. These two diagrams thus annihilate each other, and chord diagrams with long chords do not change the value of $Z(k)$.

Short chords form their own chord diagram for the trivial knot embedding k_0 with two pairs of critical points. That is why adding of an additional pair of critical points multiply the value of $Z(k)$ by $Z(k_0)$.

2. Any invariant of order no greater than n may be extended to chord diagrams of order n.

3. Let c be a chord diagram of order n. Consider a singular knot realizing the diagram c and a Chebyshev embedding of this knot. The Kontsevich invariant $\tilde{Z}$ may be extended to this singular knot. We are going to prove that the value of $\tilde{Z}$ on this knot is c plus terms of order greater than n. In order to do this it is sufficient to prove the same fact for the Kontsevich integral.

   Consider 2^n (nonsingular) knots that may be obtained by all possible resolutions of this knot. For a chord diagram $c((x_1, y_1), \ldots, (x_m, y_m))$ containing a "long" chord x_k, y_k the alternating sum in the integral vanishes. We are thus interested in diagrams containing only "short" chords. If such a diagram contains "short" chords not for each singular point, the corresponding alternating sum vanishes as well. Therefore the minimal number of chords in a chord diagram with a nonzero coefficient is n and this chord diagram should coincide with the diagram c.

   For such a diagram the coefficient may be calculated explicitely. It equals the product of n residues of the forms $d(z_i - z_i')/(z_i - z_i')$ in double points, i.e., to 1.

4. This is a complicated result following from results of Drinfeld.

5. Since the set of group-like elements is closed with respect to the product, it is sufficient to prove this statement for any value of the Kontsevich integral $Z_n(k)$. For the integral it follows directly from the definition of the comultiplication in the bialgebra of chord diagrams.

Bibliography

[1] J. W. Alexander, *Topological invariants of knots and links*, Trans. Amer. Math. Soc. **30**, (1928), 275–306.

[2] V. I. Arnold, *The Vassiliev Theory of Discriminants and Knots*, in: First European Congress of Mathematics, Paris, July 6-10, 1992; A. Joseph, ... ed. – Basel; Boston; Berlin; Birkhäuser, pp. 3–30 (1994)

[3] D. Bar-Natan, *On Vassiliev knot invariants*, Topology, Vol. 34, No. 2, 423–472 (1995)

[4] J. Birman, *New points of view in knot theory*, Bull. Amer. Math. Soc., **28**, (April 1993), 253–287

[5] J. Birman, *Braids, links and mapping class group*, 1984

[6] J. Birman and X.-S. Lin, *Knot polynomials and Vassiliev invariants*, Invent.Math. **111**, (1993), 225–270.

[7] S. V. Chmutov, S. V. Duzhin, *An upper bound for the number of Vassiliev knot invariants*, Journal of Knot Theory and Ramifications, Vol. 3, No. 2, 141–152 (1994)

[8] S. V. Chmutov, S. V. Duzhin, S. K. Lando, *Vassiliev knot invariants I. Introduction*, Advances in Sov. Math., vol. 21. pp. 117–126 (1994)

[9] S. V. Chmutov, S. V. Duzhin, S. K. Lando, *Vassiliev knot invariants II. Intersection Graph Conjecture for Trees*, Advances in Sov. Math., vol. 21. pp. 127–134 (1994)

[10] S. V. Chmutov, S. V. Duzhin, S. K. Lando, *Vassiliev knot invariants III. Forest algebra and weighted graphs*, Advances in Sov. Math., vol. 21. pp. 135–146 (1994)

[11] J. H. Conway, *An enumeration of knots and links and some of their algebraic properties*, Computational Problems in Abstract Algebra (J.Leech, ed.), Pergamon Press, New York, 1970, 329–358.

[12] R. H. Crowell, R. H. Fox, *Introduction to knot theory*, Ginn, 1963

[13] L. Geissinger, *Hopf algebras of symmetric functions and class functions*, — in: "Combinatoire et représentations du groupe symétrique", Lect. Notes in Math., 579, 168–181, Springer-Verlag, 1977

[14] M. C. Golumbic, *Algorithmic Graph Theory*, Academic Press, New York, 1980

[15] L. H. Kauffman, *On knots*, Princeton University Press, Princeton, 1987

[16] M. Kontsevich, *Vassiliev's Knot Invariants*, in: Advances in Soviet Math., v. 16, part 2, 1993, 137-150

[17] Le Tu Quoc Thang, Murakami Jun, *The universal vassiliev-Kontsevich invariant for framed oriented links*, preprint MPIM-Bonn, 93-89

[18] J. Milnor, J. Moore, *On the structure of Hopf algebras*, Ann. Math. **81** (1965), 211–264.

[19] A. B. Sossinsky, *Feynman diagrams and Vassiliev invariants*, IHES preprint, Paris, February 1992.

[20] M. E. Sweedler *Hopf algebras*, W.A.Benjamin, Inc., 1969

[21] W. T. Tutte. *Graph Theory,* Addison-Wesley, 1984

[22] V. A. Vassiliev, *Cohomology of knot spaces*, Theory of Singularities and Its Applications (ed. V. I. Arnold), Advances in Soviet Math., vol. 1, AMS, 1990

[23] V. A. Vassiliev, *Compliments of discriminants of smooth maps: Topology and Applications,*, AMS, 1992

[24] V. A. Vassiliev, *Invariants of knots and complements of discriminants*, in: Developments in Mathematics, The Moscow School, V. Arnold, M. Monastyrsky, eds., Chapmann & Hall, 1993, 194–250

[25] A. Zelevinsky, *Representations of Finite Classical Groups. A Hopf Algebra Approach*, Lect.Notes in Math., **869**, Springer-Verlag, Berlin – Heidelberg – New York, 1981