Critical behavior of the matrix models generating 3D random volumes

Naoya Umeda*

Department of Physics, Kyoto University, Kyoto 606-8502, Japan

*E-mail address: n_umeda@gauge.scphys.kyoto-u.ac.jp

Abstract

The worldvolume theory of membrane is mathematically equivalent to three-dimensional quantum gravity coupled to matter fields corresponding to the target space coordinates of embedded membrane. We propose a class of models, "triangle-hinge models", that generate three-dimensional random volumes, where each configuration consists of triangles glued together along multiple hinges. Although most of the diagrams represent configurations which are not manifolds, we show that the set of possible diagrams can be drastically reduced such that only (and all of the) three-dimensional manifolds with tetrahedral decompositions appear, by introducing a color structure and taking an appropriate large-N limit. We further introduce matter degrees of freedom to the models by coloring simplices in a way that they have local interactions. We also define unoriented membrane theories in terms of tetrahedral decompositions, and then realize them as triangle-hinge models.

Triangle-hinge models can be redefined so that they give finite free energies without changing the original perturbation theory. We study the phase structure of the models numerically, but the restriction of the diagrams to tetrahedral decompositions makes the action complex-valued, and the complexified model suffers from the sign problem in the large-N limit. We use the generalized Lefschetz thimble method, which is one of the promising methods to solve the sign problem. However, this method is known to suffer from a multimodal problem, and thus we improve it by implementing the parallel tempering method, where the flow time of an antiholomorphic gradient flow is used as a tempering parameter. We then study the phase structure of the complexified triangle-hinge model. Our parallel tempering algorithm works well, but it turns out that a large amount of computational cost is still required to make a definite statement on the phase structure of the model.

In order to reduce the computational cost, it is generally important to carefully choose the set of tempering parameters. We define for a given Markov chain Monte Carlo algorithm a distance between two configurations that quantifies the difficulty of transition in the configuration space. The distance enables us to make the above adjustment in a geometrical way. We also argue that an anti-de Sitter-like geometry naturally emerges when we consider highly multimodal action with a large number of degenerate vacua.

Contents

Chapter 1

Introduction

String theory is a strong candidate for a unified theory including quantum gravity. However, it still does not have a constructive and nonperturbative definition. The main reason is the lack of our understanding on the real fundamental dynamical variables of string theory. In fact, since the advent of D-branes and the discovery of string dualities, the idea has widely spread that the fundamental dynamical variables need not be strings and can be other types of extended objects.

String theory is mathematically equivalent to the two-dimensional gravity theory, where the target space coordinates of a string are expressed as scalar fields in the worldsheet. The roles played by matrix models in noncritical string theory should be suggestive, where the Feynman diagrams of matrix models are interpreted as triangular (or polygonal) discretization of string worldsheets (see [6, 7] for reviews). Furthermore, by introducing the degrees of freedom corresponding to matter fields on worldsheets or by considering a matrix field theory, one can define various kinds of string theory in terms of matrix models [8]. The $1/N$ expansion of matrix models, where N is the size of matrix, corresponds to the genus expansion of string worldsheets as in [9]. Moreover, the double scaling limit enables us to study the nonperturbative aspects of string theory [10, 11, 12, 13, 14, 15] as well as their integrable structure [16, 17, 18].

However, this approach cannot reach to higher-dimensional string theories: Unfortunately there is the obstacle called "$c = 1$ barrier", which means that the continuum theory no longer describes two-dimensional gravity if we consider scalar fields with central charge $c > 1$, because a class of diagrams called "branched polymer" have dominant contributions rather than the diagrams corresponding to smooth two-dimensional surfaces. Because of this obstacle, this approach cannot give nonperturbative definitions of the string theory in the target space of more than two dimension.

If we consider a membrane instead of a string, we may be able to define higher-dimensional

quantum gravity theory nonperturbatively. In the case of matrix models, the collective coordinate of the eigenvalues play the role of a scalar field. This emergence arises from the difference between the degrees of freedom of the metric of the worldsheet (= 4) and the dimension of the diffeomorphism (= 3). On the other hand, in the case of membrane theory, the difference between them (6 and 3 for the three-dimensional worldvolume) suggests that three-dimensional space can emerge in this approach.

Tensor models [19, 20, 21] or group field theory [22, 23] are natural generalizations of matrix models to three (and higher) dimensions. For three-dimensional models, the perturbative expansion generates random tetrahedral decompositions of three-dimensional objects. Unlike the matrix model approach to string theory, however, there are some obstacles: One is that the randomly generated tetrahedral decompositions are not always manifolds or not even pseudomanifolds. Another obstacle is that the dynamical variables of these models are not matrices but rank-three tensors, which do not have such good analytic properties in contrast to matrices. In particular, tensors cannot be diagonalized as matrices can, and an analogue of the saddle point method has not been found yet. Recently the situation was improved by the colored tensor models (see, e.g., [24] for a review). It is shown that the colored tensor models admit a large-N expansion and the leading contributions represent higher-dimensional sphere [25, 26]. Moreover, it is claimed that one can take a double scaling limit in the tensor models [27, 28]. Thus, the colored tensor models give a fascinating formulation of higher dimensional quantum gravity. Nevertheless, the analytic treatment of tensor models is still not so easy as that of matrix models.

We proposed in a paper [1] a new class of models which generate three-dimensional random volumes (which we call "triangle–hinge models"), by regarding each random diagram as a collection of triangles glued together along multiple hinges as in [29]. Our models have real symmetric matrices as the dynamical variables and are characterized by semisimple associative algebras $\mathcal{A}$. Although most of the diagrams represent configurations which are not manifolds,[1] we show that the set of possible diagrams can be drastically reduced such that only (and all of the) three-dimensional manifolds with tetrahedral decompositions appear, by introducing a color structure and taking an appropriate limit of parameters existing in the models. Since our models are written with matrices, there should be a chance that various techniques in matrix models can be applied and the dynamics of random volumes can be understood more analytically. Furthermore, our models have a novel strong-weak duality which interchanges the roles of triangles and hinges when $\mathcal{A}$ is a matrix ring. This duality may suggest the analytic solvability of the models.

We gave a general prescription to introduce matter degrees of freedom to triangle–hinge

[1]By a manifold we always mean a closed combinatorial manifold, which is a collection of tetrahedra whose faces are identified pairwise and each of whose vertices has a neighborhood homeomorphic to three-dimensional ball B^3. See, e.g., [30] for the rigorous definition.

models in [2]. This is achieved by setting the defining associative algebra $\mathcal{A}$ to be a tensor product of the form $\mathcal{A}_{\rm grav} \otimes \mathcal{A}_{\rm mat}$ and by modifying the interaction terms appropriately. The matter fields thus obtained have local interactions, since a colored tetrahedron can only interact with the neighbors. We can assign colors not only to tetrahedra but also to simplices of arbitrary dimensions. By this prescription we expect that we can introduce gauge fields coupled to gravity in the continuum theory. We also gave a prescription to gauge the orientation of the gluing of tetrahedra in [3]. By regarding each triangle as representing a propagation of an open membrane of disk topology, we introduce a local worldvolume parity transformation which inverts the orientation of triangle, and define unoriented triangle–hinge models by gauging the transformation. Unlike two-dimensional cases, this local transformation generally relates a manifold to a nonmanifold, but still is a well-defined manipulation among tetrahedral decompositions.

The free energy of the triangle–hinge models diverges even if we consider the highest order term of the potential as even. However, we find that we can redefine the model so that the models give finite free energy without changing the original perturbation series. This gives us the nonperturbative definition of the model, and thus will give the nonperturbative definition of the membrane theory. Furthermore, this enables us to study the phase structure of the models by the Monte Carlo simulation. We then perform the Monte Carlo simulation to a simplified triangle–hinge model, and show that the third-order phase transition occurs. However, this critical point is quite similar to that of ordinary matrix models, and this suggests that the continuum theory does not describe three-dimensional gravity. This would be because we do not restrict the diagrams to tetrahedral decompositions in the simplified model, and we should study the triangle–hinge models with the restriction. However, unfortunately, if we restrict the diagram and redefine the models to give finite free energy simultaneously, the action becomes complex-valued. Therefore we must take care of the sign problem.

There have been many proposals to circumvent the sign problem. One of the approaches that are currently under intense study is the use of the integration over Lefschetz thimbles [31] (see also [32, 33, 34, 35, 36, 37, 38, 39]), which is called the Lefschetz thimble method. This method suppresses the oscillation of integrand by changing integration manifold to a set of Lefschetz thimbles, on each of which the imaginary part of action is constant. This change of variables can be achieved by using an antiholomorphic gradient flow. However, this method suffers from a multimodal problem which is caused by infiinitely high potential barriers. Recently, Alexandru et al. [40] made a very interesting proposal to consider configurations on a manifold that is obtained from the original configuration manifold by a finite amount of flow time (this method is called the generalized Lefschetz thimble method). This algorithm made a great success in various models [40, 41], but reducing the amount of flow time may also reduce the effectiveness against the sign problem, and one does not know

a priori whether the chosen flow time avoids both the sign problem and the multimodal problem simultaneously. We proposed in [4], as a versatile tools for Monte Carlo calculations of models with complex actions, a Lefschetz thimble algorithm where the flow time is used as an auxiliary variable for the highly multimodal distribution. There, two separated modes at large flow times (where the sign problem no longer exists) are then connected by passing through configurations at small flow times (where the original sign problem exists but the multimodality is expected to be mild). Since the sample average is taken only with respect to the largest flow time, we need not worry about the sign problem at small flow times although we take into account configurations there.

We perform the Monte Carlo simulation to the triangle–hinge model with the restriction to tetrahedral decompositions, by using the generalized Lefschetz thimble method with the parallel tempering algorithm. Our parallel tempering algorithm works well, but it turns out that a large amount of computational cost is still required to make a definite statement on the phase structure of the model.

In order to reduce the computational cost of the tempering method, it is important to choose the set of tempering parameters. As a first step to obtain a general prescription to tune the parameters, we introduce in [5] for a given Markov chain Monte Carlo (MCMC) algorithm a distance between two configurations that quantifies the difficulty of transition from one configuration to the other configuration. We argue that the distance takes a universal form for the class of algorithms which generate local moves in the configuration space. We make explicit calculations of distance for the Langevin algorithm, and show that it has desired and expected properties as distance which quantifies the extent of separation between two configurations. We further show that the distance for a multimodal distribution gets dramatically reduced from a large value by the introduction of a tempering method. The introduction of such distance opens a way to investigate relaxation processes in an MCMC algorithm in terms of the geometry of the configuration space itself (which should not be confused with the geometry of the set of probability distributions). As an example, we argue that, when the original distribution is highly multimodal with large number of degenerate vacua, our distance can be regarded as a geodesic distance with respect to an anti-de Sitter-like (AdS-like) metric. Such a geometrical viewpoint enables us to adjust the tempering parameter in a purely geometrical way; the adjustment can be carried out by requiring that the resulting geodesic distance is minimized.

This thesis consists of three parts. Part I is the review of triangle–hinge models. First, in Chap. 2, we briefly review the matrix model approach to two-dimensional gravity theory, and tensor models which is an natural generalization of the approach to three-dimensional gravity. In Chap. 3 we define triangle–hinge models [1]. In Chap. 4, as extensions of the triangle–hinge models, we explain a the prescription to introduce matter degrees of freedom

[2], and further define unoriented models [3]. In Part II we make a numerical study of triangle–hinge models and give a versatile prescription to take care of the sign problem. In Chap. 5 we make a numerical study of a simplified triangle–hinge models, and show that there is a third-order critical point. We also argue that this critical point do not correspond to three-dimensional gravity. In Chap. 6, after reviewing the generalized Lefschetz thimble method [40], we implement the parallel tempering algorithm for this method [4]. We also review a promising approach [42] using the deep learning, although it seems not to be inefficient in practical use yet. In Chap. 7, as an exercise for the application of the above method to the triangle–hinge models, we study the one-matrix models with complex coefficients. There, we find that a new class of solution, which appears only for complex coefficients, and it actually realizes in some parameter region. In Chap. 8 we study the triangle–hinge models with the restriction to tetrahedral decompositions. Part III is on a tool to adjust parameters in MCMC algorithms. In Chap. 9 we define for a given MCMC algorithm a distance between two configurations that quantifies the difficulty of transition from one configuration to the other configuration [5]. Chap. 10 is devoted to conclusion and outlook for future work.

Part I

Triangle–hinge models

Chapter 2

Discrete approach to quantum gravity

In Part I, we review our work [1, 2, 3], where we define a class of models which generate three-dimensional random volumes, and study the extension of the models. In this chapter, we first review a discretized approach to two-dimensional quantum gravity, and its natural generalization to three-dimensional quantum gravity.

2.1. Discrete approach to two-dimensional quantum gravity

In this section, we briefly review the idea of the random triangulation of two-dimensional surface, and its realization by matrix models [8].

The idea of this approach is (roughly speaking) as following: First, we assume that path integral of metric which appear in two-dimensional quantum gravity is equivalent to summing up all different "shape" of two-dimensional (closed) surface. Then, if we discretize two-dimensional surface as triangular decomposition, the discretize version of the path integral is summing up all different triangular decomposition. So if we construct a model which generate triangular decomposition dynamically, more specifically, a model where each Feynman diagram has one-to-one correspondence to a triangular decomposition, the model with specific Boltzmann weight should describe discretized quantum gravity. We see this idea and the realization of this idea in Sect. 2.1.1.

These models have a great success due to its analytic properties, especially such that they can be solved analytically in the large-N limit by using saddle point method. We briefly see some of analytic properties in Sect. 2.1.2 and in Sect. 2.1.3.

2.1.1. Random triangulation of two-dimensional surface

In this subsection, we illustrate the idea of random triangulation. In order to do this, we first discretize the action of two-dimensional Euclidean gravity, and next consider the models which dynamically generate the Boltzmann weights.

Let us consider the simplest action of string theory

$$Z = \sum_{\text{topologies}} \int [\mathcal{D}g_{\alpha\beta}][\mathcal{D}X] e^{-S} \tag{2.1.1}$$

$$S = \frac{1}{4\pi} \int d^2\sigma \sqrt{g} \Big(4\lambda + \phi R + \frac{1}{\alpha'} g^{\alpha\beta} \partial_\alpha X \partial_\beta X \Big), \tag{2.1.2}$$

where the summation is taken over all topologies of two-dimensional surfaces. Let us consider the discretization of two-dimensional surfaces by triangular decomposition as in Fig. 2.1. In this discretization, the surface is flat on every face of triangles, and the curvature

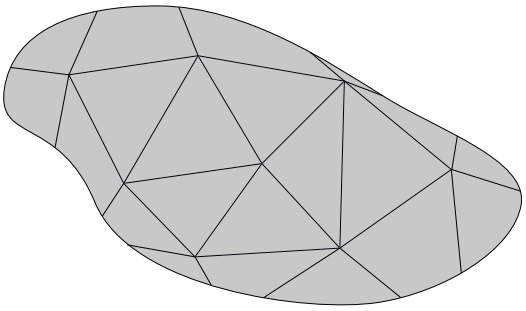


Figure 2.1: An example of triangular decompositions of two-dimensional surface.

concentrates at vertices: If q_v triangles meet at the vertex v, there is a conical singularity corresponding to the deficit angle $2\pi(1 - q_v/6)$, and this gives the curvature (as we will see below). Since the (continuous) path integral corresponding to the metric can be regarded as the summation over all "shapes" of two-dimensional surfaces, the partition function can be given as the sum of all different triangular decompositions. The scalar field $X(\sigma)$ can then be represented as scalar degrees of freedom assigned on the center of each triangles. The discretized version of partition function Z can then be given as the sum of all different triangular decompositions with proper Boltzmann weights. The continuum partition function should be given in the limit $a \to 0$, where a is the length of a side of triangles.[1]

The Boltzmann weight with the action (2.1.2) can be rewritten in terms of triangular decompositions. Since the first term of the action (2.1.2) represents the area of two-dimensional surfaces, this corresponds to the number of triangles in a triangular decomposition. Con-

[1] Although the curvature concentrates at vertices, it can become continuous in the limit since the vertices distribute enough densely.

cretely, the first term can be rewritten as

$$\int d^2\sigma\sqrt{g} \to \frac{\sqrt{3}}{4}a^2 \times (\#\text{ of triangles}). \tag{2.1.3}$$

The second term represents Euler characteristics:

$$\frac{1}{4\pi}\int d^2\sigma\sqrt{g}R = 2(1-h), \tag{2.1.4}$$

where h is the genus of the surface. Actually, if we consider a parallel displacement of a vector u^α along a circle in Fig. 2.2, the vector u'^α after the displacement can be written as

$$\begin{aligned} u'^\alpha &= u^\alpha - \epsilon_1^\gamma \epsilon_2^\delta R_{\gamma\delta}{}^\alpha{}_\beta u^\beta \\ &= u^\alpha - R_{12}{}^\alpha{}_\beta u^\beta \\ &\Rightarrow u'^1 + iu'^2 = e^{iR_{1212}}(u^1 + iu^2). \end{aligned} \tag{2.1.5}$$

Then the scalar curvature corresponds to the deficit angle as

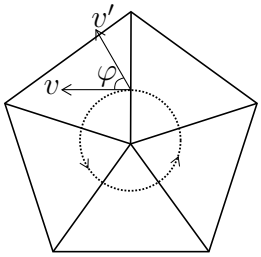


Figure 2.2: A parallel displacement of a vector around a vertex. The vector u^α moves along the dotted circle, and becomes u'^α after the displacement.

$$R = 2R_{1212} = 4\pi(1 - \frac{q_v}{6})\delta^{(2)}(\sigma - \sigma_v). \tag{2.1.6}$$

We can use the relation among the number of the faces, edges and vertices of triangular decompositions (which we denote as s_2, s_1 and s_0)

$$\begin{aligned} s_2 - s_1 + s_0 &= 2(1-h) \\ \sum_{vertex} q = 3s_2, \quad s_1 &= \frac{3}{2}s_2, \end{aligned}$$

and RHS of (2.1.6) can be rewritten as

$$\sum_{vertex} 4\pi\big(1 - \frac{q}{6}\big) = 4\pi\big(s_0 - \frac{1}{2}s_2\big) = 8\pi(1-h), \tag{2.1.7}$$

and correctly gives the Euler characteristics. The matter fields can be represented the color structure of triangles:

$$\int d^2\sigma\sqrt{g}g^{\alpha\beta}\partial_\alpha X\partial_\beta X \to \sum_{\langle i,j\rangle}(X_i - X_j)^2, \tag{2.1.8}$$

where i, j is the indices of triangles and $\langle i, j\rangle$ runs over all neighbored triangles. Finally, the Boltzmann weight e^{-S} can be written as

$$Z(g_0, \kappa) = \sum_{conf.} g_0^{2(h-1)}\kappa^{s_2}\prod_{i=1}^{s_2}\int dX_i \prod_{<ij>} G(X_i - X_j) \tag{2.1.9}$$

$$\kappa = \exp\big(-\frac{\sqrt{3}a^2\lambda}{4\pi}\big) \tag{2.1.10}$$

$$G(X) = \exp\big(-\frac{1}{4\pi\alpha'}X^2\big). \tag{2.1.11}$$

This weight can be obtained when we assign the weight on each face, edge and vertex of triangles, and introduce color structures on triangles.[2]

As shown in the below briefly, the partition function (2.1.9) corresponds to the free energy of a matrix model. Let us consider a matrix model:

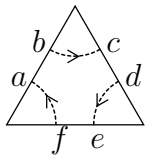
$$Z \equiv \int dM e^{-N\mathrm{tr}V(M)} \tag{2.1.12}$$

$$V(x) = \frac{1}{2}x^2 + \frac{\lambda}{3}x^3, \tag{2.1.13}$$

where M is $N \times N$ Hermitian matrix. If we consider a perturbation with respect to λ, the Feynman rules can be illustrated as

$$\text{propagator}: \quad \begin{matrix} a \dashleftarrow d \\ b \dashrightarrow c \end{matrix} \quad \sim \quad \langle M(x)^a{}_b M(y)^c{}_d\rangle = \frac{1}{N}e^{-|x-y|/4\pi\alpha'}\delta^a{}_d\delta^c{}_b \tag{2.1.14}$$

$$\text{interaction vertex}: \quad \sim \quad N\kappa\delta^c{}_b\delta^e{}_d\delta^a{}_f. \tag{2.1.15}$$



Since the triangles are glued each other so to align the direction of index lines, the obtained Feynman diagrams correspond to triangular decompositions of oriented two-dimensional surface. The power of N in the Boltzmann weight is the Euler characteristics of the triangular decomposition: The number of the index loops, the propagator and the interaction vertex, which give the factor N, $1/N$ and N, correspond to the number of vertices, edges and faces of a diagram. Since the power of λ in the Boltzmann weight is the number of triangles,

[2]Note that we only give a naive correspondence (for example, we do not consider the symmetry factor).

we can compute the free energy as the sum of all connected diagrams with the Boltzmann weight $\lambda^{s_2} N^{2(1-h)}$, which corresponds to that of two-dimensional gravity (2.1.9) with

$$\frac{1}{N} \leftrightarrow g_0 \tag{2.1.16}$$

$$\lambda \leftrightarrow \exp\big(\frac{\sqrt{3}a^2}{4\pi}\Lambda\big). \tag{2.1.17}$$

In this sense, the matrix model (2.1.13) dynamically generates triangular decomposition, and represents two-dimensional (pure) gravity.[3] In the following subsections, we will see the analytic properties of this model.

2.1.2. Large-N limit

The most important property of this model is that this model can be solved analytically in the large-N limit. In this subsection, we review the large-N limit of the model.

In Sect. 2.1.1 we discretized two-dimensional surfaces only by triangles. In general we can decompose the surfaces by polygons, and then we can represent a surface in various ways. In fact, when we consider polygonal decompositions, the model can have the continuum limit which corresponds to two-dimensional gravity coupling to matter fields (although unfortunately the continuum theory becomes non-unitary in general). Then, in the following, we consider the action

$$Z \equiv \int dM e^{-N\beta \mathrm{tr} V(M)} \tag{2.1.18}$$

$$V(x) = \sum_{n=1}^{m} v_n x^{2n}, \tag{2.1.19}$$

instead of (2.1.13). Here we set $v_1 = 1/2$ for the normalization of the propagator. We consider only even potentials for simplicity, but in general we can consider odd terms because the properties of continuum limits only depend on the behavior near critical points.

Let us consider the large-N limit of the model. Since the action (2.1.19) is the trace of a Hermitian matrix M, the action only depends on the eigenvalues of M. We can decompose M to the eigenvalues and unitary matrix, and then the partition function can be rewritten as

$$Z = \int \prod_{i=1}^{N} d\lambda_i |\Delta(\lambda)|^2 e^{-N\beta \sum_{i=1}^{N} V(\lambda_i)}, \tag{2.1.20}$$

by integrating out the unitary matrix. Here, $\Delta(\lambda) \equiv \prod_{i<j}(\lambda_i - \lambda_j)$ is Vandermode determinant caused from the measure dM. Now this is the system of eigenvalues with effective

[3] Matter fields can be introduced by assigning color to triangles, or use polygons not just triangles.

action

$$S_{\text{eff}} \equiv N\beta \sum_{i=1}^{N} V(\lambda_i) - 2\log|\Delta(\lambda)|, \tag{2.1.21}$$

and the equation of motion can be derived as

$$\frac{2}{N}\sum_{i\neq j}\frac{1}{\lambda_i-\lambda_j} = \beta V'(\lambda_i). \tag{2.1.22}$$

[4]

When we discuss the large-N limit of the model, it is convenient to introduce the resolvent

$$\omega(z) \equiv \frac{1}{N}\mathrm{tr}\Big(\frac{1}{z-M}\Big) = \frac{1}{N}\sum_i \frac{1}{z-\lambda_i}. \tag{2.1.23}$$

We can transform by using EOM (2.1.22) that

$$\begin{aligned}
\frac{\beta}{N}\sum_i \frac{V(\lambda_i)'}{z-\lambda_i} &= \frac{2}{N^2}\sum_{i\neq j}\frac{1}{\lambda_i-\lambda_j}\frac{1}{z-\lambda_i} \\
&= \frac{1}{N^2}\sum_{i\neq j}\frac{1}{\lambda_i-\lambda_j}\Big(\frac{1}{z-\lambda_i}-\frac{1}{z-\lambda_j}\Big) \\
&= \frac{1}{N^2}\sum_{i\neq j}\frac{1}{(z-\lambda_i)(z-\lambda_j)} \\
&= \Big(\frac{1}{N}\sum_i \frac{1}{z-\lambda_i}\Big)^2 - \frac{1}{N^2}\sum_i\frac{1}{(z-\lambda_i)^2} \\
&= w(z)^2 + \frac{1}{N}w(z)'.
\end{aligned} \tag{2.1.24}$$

On the other hand, the LHS can be transformed to $\beta V(z)'w(z) - \frac{\beta}{N}\sum_i \frac{V(z)'-V(\lambda_i)'}{z-\lambda_i}$. Then we obtain the equation for $w(z)$

$$w(z)^2 + \frac{1}{N}w(z)' - \beta V(z)'w(z) + \frac{\beta^2}{4}R(z) = 0, \tag{2.1.25}$$

where $\frac{\beta^2}{4}R(z) \equiv \frac{\beta}{N}\sum_i \frac{V(z)'-V(\lambda_i)'}{z-\lambda_i}$ is a polynomial of z. Since the second term of LHS is subleading in the large-N limit, this equation becomes a quadratic equation

$$w(z)^2 - \beta V(z)'w(z) + \frac{\beta^2}{4}R(z) = 0, \tag{2.1.26}$$

[4]Although we can consider the equation of motion for the original action (2.1.19), it is insignificant because the quantum corrections have the same order of the classical solution, due to the high degree of freedom of the matrix M.

and we obtain

$$w(z) = \frac{\beta}{2}\left(V(z)' - \sqrt{V(z)'^2 - R(z)^2}\right), \tag{2.1.27}$$

where the sign is determined by $w(z) \to 1/z(|z| \to \infty)$. In the large-$N$ limit, the eigenvalues λ_i distributes densely, and the resolvent can be written by eigenvalue distribution $\rho(\lambda)$ as

$$w(z) = \int d\lambda \frac{\rho(\lambda)}{z-\lambda} \tag{2.1.28}$$

$$\frac{\beta^2}{4}R(z) = \beta \int d\lambda \rho(\lambda) \frac{V(z)' - V(\lambda)'}{z-\lambda}. \tag{2.1.29}$$

The EOM (2.1.22) now become

$$2\fint d\lambda' \frac{\rho(\lambda')}{\lambda - \lambda'} = \beta V(\lambda)', \tag{2.1.30}$$

and by using the equation $\frac{1}{z \pm i\epsilon} = P\frac{1}{z} \mp i\pi\delta(z)$ we can obtain the properties of the cut of $w(z)$

$$w(x+i0) + w(x-i0) = \beta V(x)' \tag{2.1.31}$$

$$w(x+i0) - w(x-i0) = -2\pi i \rho(x) \tag{2.1.32}$$

These equations and the property $w(z) \to 1/z(|z| \to \infty)$ also lead us to (2.1.27).

We can determine $w(z)$ completely (i.e. determine $R(z)$) by some assumption for the eigenvalue distribution. For example, let us consider the case the eigenvalue distributes in $[-a, a]$. The property (2.1.32) means that $w(z)$ has a single cut on $[-a, a]$ (the resulting solution is thus called "one-cut solution"). $V(z)'^2 - R(z)$ can be written as $M(z)^2(z^2 - a^2)$ since this can be negative for $z \in [-a, a]$. Since $M(z)$ is also a polynomial of z we can determine $M(z)$ and $a(\beta)$ from $w(z) \to 1/z(|z| \to \infty)$, but we here give an explicit form of $w(z)$ as an integral. The simplest one-cut homogeneous solution of (2.1.31) is $\sqrt{z^2 - a^2}$. By dividing (2.1.31) by $\sqrt{z^2 - a^2}$ we obtain

$$\left.\frac{iw(z)}{\sqrt{z^2-a^2}}\right|_{z=x+i0} - \left.\frac{iw(z)}{\sqrt{z^2-a^2}}\right|_{z=x-i0} = \frac{\beta V(x)'}{\sqrt{a^2-x^2}}. \tag{2.1.33}$$

This represents the discontinuity of $\frac{iw(z)}{\sqrt{z^2-a^2}}$ on the cut. When we define $\frac{iw(z)}{\sqrt{z^2-a^2}} = \int_{-a}^{a} dx \frac{\sigma}{z-x}$, we obtain $-2\pi\sigma(x) = \frac{V(x)'}{\sqrt{a^2-x^2}}$ in the same way for (2.1.32). Then we can transform $w(z)$ as

$$\begin{aligned} w(z) &= \frac{1}{2\pi}\sqrt{z^2-a^2}\int_{-a}^{a} \frac{dx}{z-x}\frac{\beta V(x)'}{\sqrt{a^2-x^2}} \\ &= \int_0^a \frac{dx}{2\pi}\Big(\frac{1}{z-x} - \frac{1}{z+x}\Big)\beta V(x)' \frac{\sqrt{z^2-a^2}}{\sqrt{a^2-x^2}} \\ &= \int_0^a \frac{dx}{\pi}\frac{\beta x V(x)'}{z^2-x^2}\frac{\sqrt{z^2-a^2}}{\sqrt{a^2-x^2}} \end{aligned} \tag{2.1.34}$$

By the asymptotic property $w(z) \to 1/z(|z| \to \infty)$ we obtain

$$\beta(a)^{-1} = \int_0^a \frac{dx}{\pi} \frac{xV(x)'}{\sqrt{a^2 - x^2}} = \sum_{n=1}^{m} \frac{v_n a^{2n}}{B(n, 1/2)}, \tag{2.1.35}$$

where $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

$w(z)$ can be written as an integral of the coupling constant β. We define

$$U(A) \equiv \int_-^{\sqrt{A}} \frac{dx}{\pi} \frac{xV(x)'}{\sqrt{a^2 - x^2}}. \tag{2.1.36}$$

Here, $U(A)|_{A=a^2} = \beta(a)^{-1}$ and by using

$$\int_0^{a^2} \frac{dx}{\pi} \frac{xV(x)'}{\sqrt{a^2 - x^2}} = \frac{1}{B(2B(n, 1/2))} \int_0^{a^2} dA \frac{A^{n-1}}{\sqrt{z^2 - A}}, \tag{2.1.37}$$

we can rewrite $w(z)$ as

$$\begin{aligned}
\frac{1}{\beta} w(z) &= \int_0^a \frac{dx}{\pi} \frac{xV(x)'}{z^2 - x^2} \frac{\sqrt{z^2 - a^2}}{\sqrt{a^2 - x^2}} \\
&= \sum_{n=1}^{m} 2nv_n \int_0^a \frac{dx}{\pi} \frac{x^{2n}}{x^2 - x^2} \frac{\sqrt{z^2 - a^2}}{\sqrt{a^2 - x^2}} \\
&= \sum_{n=1}^{m} \frac{nv_n}{B(n, 1/2)} \int_0^{a^2} dA \frac{A^{n-1}}{\sqrt{z^2 - A}} \\
&= \int_0^{a^2} dA \frac{U(A)'}{\sqrt{z^2 - A}} \\
&= \int_0^{\beta^{-1}} \frac{d\beta'^{-1}}{\sqrt{z^2 - a^2(\beta')}}.
\end{aligned} \tag{2.1.38}$$

2.1.3. Orthogonal polynomial method

We close this section by reviewing another analytic method called "orthogonal polynomial method", which we use to calculate the partition function of this model analytically.

In orthogonal polynomial method, we consider a series of orthogonal polynomials that satisfy the following equations:

$$\int_{-\infty}^{\infty} d\lambda e^{-N\beta V(\lambda)} P_n(\lambda) P_m(\lambda) = s_n \delta_{nm}, \tag{2.1.39}$$

where $P_n(\lambda) = \lambda^n + \cdots$ is n-th order polynomial, and s_n is the norm of P_n. The Vandermonde determinant $\Delta(\lambda)$ can be rewritten by these polynomials as

$$\Delta(\lambda) = \det(\lambda_i^{j-1}) = \det\left(P_{j-1}(\lambda_i)\right). \tag{2.1.40}$$

Thus the partition function can be represented as the product of the norm of the polynomials

$$\begin{aligned} Z &= \int \prod_l d\lambda_l e^{-N\beta V(\lambda)} \sum_{\sigma_1,\sigma_2} (-1)^{\sigma_1}(-1)^{\sigma_2} \prod_k P_{i_k-1}(\lambda_k) \prod_j P_{i_j-1}(P_j) \\ &= N! \prod_{i=1}^{N-1} s_i. \end{aligned} \tag{2.1.41}$$

If we define $f_k \equiv s_k/s_{k-1}$ and $\xi \equiv k/N$, and take a large-N limit, ξ becomes continuous in the limit, and f_k becomes a continuous function $f(\xi)$. The free energy can be written in the limit as

$$\begin{aligned} F \equiv \frac{1}{N^2} \log Z &= \frac{1}{N} \sum_k (1 - \frac{k}{N}) \log f_k \\ &\simeq \int_0^1 d\xi (1-\xi) \log f(\xi). \end{aligned} \tag{2.1.42}$$

Let us consider the relation between $f(\xi)$ and β. Since $P_n(\lambda)$ is orthogonal to any $P_i(\lambda)$ $(i < n-1)$, we can deform

$$\lambda P_n = P_{n+1} + r_n P_{n-1}. \tag{2.1.43}$$

Here we used the relation $\int e^{-N^\beta V(\lambda)} d\lambda \lambda P_n(\lambda)^2 = 0$, which is established when $V(x)$ is even. The coefficient r_n can be determined as $r_n = f_n$ by calculating $\int d\lambda e^{-N\beta V(\lambda)} \lambda P_n(\lambda) P_{n-1}(\lambda)$ in two ways. On the other hand, by using the equation $P_n' = nP_{n-1} + \cdots$ we can rewrite

$$\begin{aligned} n s_{n-1} &= \int d\lambda e^{-N\beta V(\lambda)} P_n' P_{n-1} \\ &= \int d\lambda e^{-N\beta V(\lambda)} (NV') P_n P_{n-1} \end{aligned} \tag{2.1.44}$$

where we used partital integral in the last equality. We can also deform by using $V(\lambda)' = \sum_{k=1}^m 2kv_k \lambda^{2k-1}$ and (2.1.43) as

$$\begin{aligned} \int d\lambda e^{-N\beta V(\lambda)} V(\lambda)' P_n(\lambda) P_{n-1}(\lambda) &= \sum_{k=1}^m 2kv_k \int d\lambda e^{-N\beta V(\lambda)} \lambda^{2k-1} P_n(\lambda) P_{n-1}(\lambda) \\ &= \sum_{k=1}^m 2kv_k \underbrace{(\overbrace{f \cdots f}^{k} \cdots \overbrace{f \cdots f}^{k})}_{{}_{2k-1}C_k} s_{n-1}. \end{aligned} \tag{2.1.45}$$

If we disregard the difference among f_ks in the large-N limit, we obtain

$$\begin{aligned} \frac{\xi}{\beta} &= \sum_{k=1}^m 2kv_{k\,2k-1}C_k f(\xi)^k \\ &= \sum_{k=1}^m \frac{v_k f(\xi)}{B(n,1/2)} \\ &= U\big(f(\xi)\big). \end{aligned} \tag{2.1.46}$$

2.2. Natural generalization to three-dimensional gravity: tensor model

In Sect. 2.1 we reviewed the dynamical triangulation, the discretized approach of two-dimensional gravity. This approach succeed to describes two-dimensional quantum gravity (i.e. string theory) nonperturbatively, although this is incomplete due to $c = 1$ barrier. In this section, we consider a natural generalization of this approach to three-dimensional gravity.

In three dimension we can decompose three-dimensional volume by tetrahedra. Let us consider as a starting point Euclidean Einstein-Hilbert action

$$S = -\kappa \int d^3x\Lambda \int d^3x\sqrt{g} + \sqrt{g}R, \tag{2.2.1}$$

where κ is the surface gravity and λ is a cosmological constant. As is the case with two-dimensional surface, we decompose three-dimensional volume by tetrahedra, and regard internal of a tetrahedra to be flat. The first term represents the volume of tetrahedral decompositions, and then we can discretize this term by the number of tetrahedra s_3 as

$$\Lambda \int d^3x\sqrt{g} \to \frac{\sqrt{2}a^3}{12}\Lambda s_3. \tag{2.2.2}$$

The second term represents the curvature of the tetrahedral decomposition, which concentrates on edges of tetrahedra: When we consider tetrahedra around an edge as in Fig. 2.3, the angle θ between two faces of a tetrahedra satisfies $\cos\theta = \frac{1}{3}$, and the deficit angle around the edge with q tetrahedra can be given as $(2\pi - \theta q)$. As is the case of triangles, we consider a parallel displacement of a vector v vertical to the edge along infinitesimal loop as in Fig. 2.3, we obtain

$$\begin{aligned} v'^{\mu} &= v^{\mu} - \epsilon_1^{\rho}\epsilon_2^{\sigma} R_{\rho\sigma}{}^{\mu}{}_{\nu} v^{\nu} \\ &= v^{\mu} - \epsilon_1\epsilon_2 R_{12}{}^{\mu}{}_{\nu} v^{\nu}. \end{aligned} \tag{2.2.3}$$

Thus

$$\begin{cases} v'^1 = v^1 - \epsilon_1\epsilon_2 R_{1212} v^2 \\ v'^2 = v^2 + \epsilon_1\epsilon_2 R_{1212} v^1 \end{cases} \tag{2.2.4}$$

$$v'^1 + iv'^2 = e^{i\epsilon_1\epsilon_2 R_{1212}}(v^1 + iv^2), \tag{2.2.5}$$

and $R_{abcd} = 0$ exept for R_{1212}. Thus we obtain the Ricci scalar

$$R = 2R_{1212} = \frac{2}{\epsilon_1\epsilon_2}(2\pi - \theta q) \sim 4\pi(1 - \frac{\theta q}{2\pi})\delta^2(x_\perp), \tag{2.2.6}$$

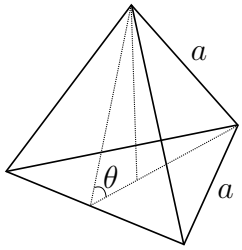


Figure 2.3: A regular tetrahedron with fixed spacing a. The angle θ is a unit of deficit angles.

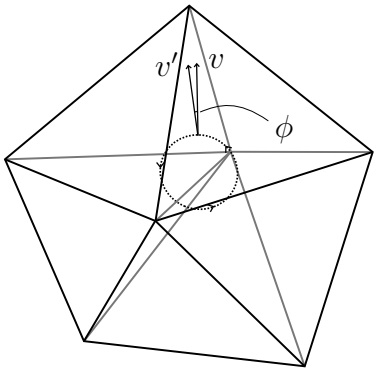


Figure 2.4: A parallel displacement of a vector in a tetrahedral decomposition. The vector v becomes v' after the displacement along the dotted circle. The deficit angle is given as $\phi = (2\pi - 5\theta)$.

and

$$4\pi\kappa \int d^3x\sqrt{g}R \to 4\pi\kappa \sum_i a(1-\frac{\theta q_i}{2\pi}), \tag{2.2.7}$$

where we denote the edges of tetrahedral decomposition as i. Finally, by using the relation among the number of tetrahedra s_3, triangles s_2, edges s_1 and vertices s_0 in a tetrahedral decomposition of three-dimensional manifold

$$\begin{aligned} s_2 = 2s_3, \quad \sum_i q_i = 6s_3 \\ s_3 - s_2 + s_1 - s_0 = 0 \quad , \end{aligned}$$

we can discretize (2.2.1) as

$$S = -4\pi\kappa a s_0 + [\frac{\sqrt{2}a^3}{12}\Lambda - 4\pi\kappa a(1-\frac{3\theta}{\pi})]s_3 \tag{2.2.8}$$

The free energy can be written as

$$\begin{aligned} F &= \sum_{\text{conf.}} \frac{1}{s} e^{-S} \\ &= \sum_{\text{conf.}} \frac{1}{s} (\exp(-4\pi\kappa a))^{s_0} [\exp(\frac{\sqrt{2}a^3}{12}\Lambda - 4\pi\kappa a(1-\frac{3\theta}{\pi}))]^{s_3}, \end{aligned} \tag{2.2.9}$$

where the summation is taken over all possible tetrahedral decompositions of three-dimensional manifolds, and s is the symmetry factor.

This discretized partition function can be represented as the free energy of tensor model [20][19][21]. In this model, we consider the interaction vertex corresponding to a tetrahedra, and the propagator which glue them at their faces. For example, we consider the action [20][19]

$$S = \sum_{A,B,C=1}^{M} \frac{1}{6}\phi_{ABC}\phi^*_{ABC} - \frac{g}{12}\sum_{A\sim I=1}^{M} \phi_{ABC}\phi_{ADE}\phi_{FBE}\phi_{FDC}, \tag{2.2.10}$$

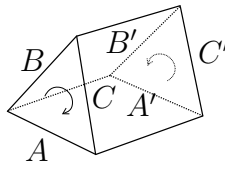
where ϕ is a symmetric rank-3 tensor

$$\phi_{ABC} = \phi_{BCA} = \phi_{CAB} \tag{2.2.11}$$

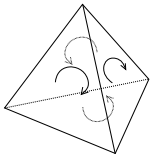
$$\phi^*_{ABC} = \phi_{CBA}. \tag{2.2.12}$$

Here we denote the edges of a tetrahedron as $A, B, C, \cdots$. The Feynman rules of this action

can be given as

propagator : 

$$\sim \frac{1}{3}(\delta_{AA'}\delta_{BB'}\delta_{CC'} + \delta_{AB'}\delta_{BC'}\delta_{CA'} + \delta_{AC'}\delta_{BA'}\delta_{CB'})$$

interactionterm : 
$$\sim g\delta\cdots\delta, \tag{2.2.13}$$

and the Feynman diagrams can be constructed by gluing them with proper orientation. That is, if we assign orientation to four faces of a tetrahedron uniformly, two faces are identified with opposite orientation in the diagram.

In (2.2.11) we consider the model where the indices are assigned to edges. We can also consider the model where the indices are assigned to vertices [20][21]:

$$S = \sum_{a,b,c=1}^{N} \frac{1}{6}\phi_{abc}\phi^*_{abc} - \frac{g}{12}\sum_{a\sim d=1}^{N} \phi_{abc}\phi_{bad}\phi_{cbd}\phi_{acd}, \tag{2.2.14}$$

where ϕ is again a symmetric rank-3 tensor

$$\phi_{abc} = \phi_{bca} = \phi_{cab} \tag{2.2.15}$$

$$\phi^*_{abc} = \phi_{cba}, \tag{2.2.16}$$

but the indices represent the vertices of a tetrahedron. This model also generates the diagrams which correspond to tetrahedral decompositions.

In particular, in the model (2.2.14) the free energy can be given as

$$\begin{aligned} F &= \log(\int D\phi \exp(-S)) \\ &= \sum_{\text{conf.}} \frac{1}{s} N^{s_0} g^{s_3}. \end{aligned} \tag{2.2.17}$$

This corresponds to the discretized Einstein-Hilbert action (2.2.8) with

$$\exp(-4\pi\kappa a) \leftrightarrow N, \quad \exp(\frac{\sqrt{2}a^3}{12}\Lambda - 4\pi\kappa a(1 - \frac{3\theta}{\pi})) \leftrightarrow g. \tag{2.2.18}$$

We thus can regard this model represents discretized three-dimensional gravity.

If we consider the large-N limit of these models, we obtain the nonperturbative definition of three-dimensional gravity. However, the dynamical variables of these models are rank-3 tensor, which do not have a direct analogy of diagonalization of matrix, and these models thus do not have such good analytic properties that the matrix models have. In particular, we cannot apply a steepest descent method to these models because of large quantum corrections.

Chapter 3

Triangle–hinge models

In this chapter, we review our models [1] (and a part of [2, 3]) which generate three-dimensional random volumes. We first show the idea of this model in Sect. 3.1. Next we define our models and show that the models are characterized by semisimple associative algebras $\mathcal{A}$ in Sect. 3.2. We give the Feynman rules of the models in Sect. 3.3, and discuss on the strucure of the Boltzmann weight in Sect. 3.4. We then give a few examples of the Feynman diagrams, and show that some diagrams are not manifolds in Sect. 3.5. we then consider matrix rings as the defining associative algebras of the models in Sect. 3.6. We show that such models can be constructed that generate only manifolds as Feynman diagrams, by introducing a color structure to the models and letting the associative algebras have centers whose dimensions play the role of free parameters (to count the number of vertices) in Sect. 3.7. We can easily see the relation between three-dimensional gravity and our models in Sect. 3.8. We note a novel strong-weak duality in Sect. 3.9. We also give another example of semisimple associative algebra in Sect. 3.10. We close this chapter by noting for the orientability of the diagrams in Sect 3.11.

3.1. Basic idea

In this section, we show the basic idea of our models. We first explain the diagrams we are concerned with and give the rule to assign a Boltzmann weight to each diagram. We then write down the action which generates such diagrams as Feynman diagrams.

We want to give a set of diagrams which corresponds to tetrahedral decompositions of three-dimensional manifolds. Our basic idea is that we consider a set of diagrams, $\{\gamma\}$, consisting of triangles glued together along multiple hinges as in [29], instead of tetrahedra themselves. In other words, we decompose tetrahedral decompositions to a set of triangles and a set of multiple hinges (see Fig. 3.1). In order to assign Boltzmann weight to diagram

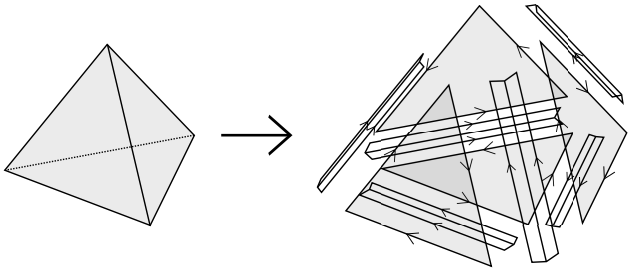


Figure 3.1: Decomposition of a three-dimensional diagram to triangles and hinges [1].

γ, we assign numbers to triangles and hinges, and give the rule to glue them.

For each edge of a triangle, we draw an arrow and assign an index from a finite set $\{I\}$. We repeat the same procedure for the hinges. We then assign the real numbers C^{IJK} and $Y_{I_1 \ldots I_k}$ to the indexed triangles and hinges, respectively, as in Fig. 3.2.[1] We require that

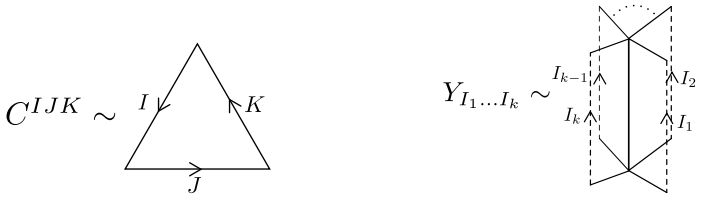


Figure 3.2: Triangles and hinges [1].

C^{IJK} and $Y_{I_1 \ldots I_k}$ be cyclically symmetric, in order to reflect the symmetries that the original triangles and hinges have.

Then, we glue the triangles and hinges to reconstruct the original diagram (tetrahedral decomposition) in such a way that the identified edges have the same index. In doing this, there may appear the case where the arrows of a triangle and a hinge have opposite directions, reflecting the fact that there are two types of gluing edges. To treat such cases, we introduce a tensor ${T_I}^J$ which reverses the direction of an arrow (see Fig. 3.3). Note that

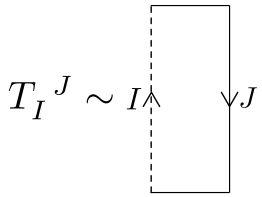


Figure 3.3: Tensor ${T_I}^J$. It changes the direction of arrow [1].

the tensor $T = ({T_I}^J)$ should be involutory because the direction of an arrow comes back to the original one after ${T_I}^J$ is applied twice:

$$ {T_I}^K {T_K}^J = {\delta_I}^J. \tag{3.1.1} $$

[1]The edges of a triangle will be drawn in solid lines while those of a hinge in dotted lines.

Furthermore, we postulate the following relations:

$$T_{I_1}{}^{J_1} \dots T_{I_k}{}^{J_k} Y_{J_1 \dots J_k} = Y_{I_k \dots I_1}, \tag{3.1.2}$$

$$C^{LMN} T_L{}^I T_M{}^J T_N{}^K = C^{KJI}, \tag{3.1.3}$$

since a hinge (or a triangle) whose arrows are all flipped is equivalent to a hinge (or a triangle) with indices in reverse order.

We define the Boltzmann weight $w(\gamma)$ of diagram γ to be the product of C^{IJK} and $Y_{I_1 \dots I_k}$ followed by the summation over the indices on the edges:

$$w(\gamma) = \frac{1}{S(\gamma)} \sum_{\{I_e\}} \prod_{f:\,\text{triangle}} C^{IJK}(f) \prod_{h:\,\text{hinge}} Y_{I_1 \dots I_k}(h). \tag{3.1.4}$$

Here, I_e are the indices on the edges, and $S(\gamma)$ is the symmetry factor of the diagram. The indices in C^{IJK} and $Y_{I_1 \dots I_k}$ are contracted when the corresponding edges are identified (with $T_I{}^J$ inserted appropriately if necessary).

The above diagrams with the prescribed Boltzmann weights can be generated as Feynman diagrams from the action[2]

$$S[A, B] = \frac{1}{2} A_I B^I - \frac{\lambda}{6} C^{IJK} A_I A_J A_K - \sum_{k \geq 2}^{\infty} \frac{\mu_k}{2k} B^{I_1} \cdots B^{I_k} Y_{I_1 \dots I_k}, \tag{3.1.5}$$

where the dynamical variables A_I and B^I satisfy the relations

$$A_I = T_I{}^J A_J, \qquad B^I = B^J T_J{}^I. \tag{3.1.6}$$

We have included the coupling constants λ and μ_k ($k \geq 2$) to count the numbers of triangles and k-hinges, respectively. In order to specify the directions of arrows, the index line will be a double line by setting index I to be double index $I = (i, j)$.

It may be already clear, but we here explain how the action generates the diagrams we are concerned with. There are two kinds of interaction terms, one corresponding to triangles C^{IJK} and the other to k-hinges $Y_{I_1 \dots I_k}$ ($k \geq 2$). The kinetic term $(1/2) A_I B^I$ yields a propagator that glues an edge of a triangle and that of a hinge. Note that two triangles cannot be glued to each other without an intermediate hinge, and two hinges cannot be glued to each other without an intermediate triangle. In order to handle the case where

[2]Note that the 2-hinges (hinges with two edges) have been included as vertices. This means that the set of the resulting diagrams contain the full set of triangular decompositions of two-dimensional surfaces, which consist only of triangles and 2-hinges. However, as we discuss in Sect. 3.7.3, the introduction of color structure can exclude all such diagrams except for a tetrahedral decomposition. The above diagram in fig. 3.1 still survive, but is then interpreted as representing three-sphere S^3 obtained by gluing two tetrahedra face to face.

the tensor $T_I{}^J$ needs to be inserted, we should multiply every leg of interaction terms by the factor $\delta_I{}^J + T_I{}^J$. However, this is equivalent to inserting the projector $\left(\delta_I{}^J + T_I{}^J\right)/2$ in every propagator, which in turn is equivalent to requiring that the dynamical variables be invariant under the action of $T_I{}^J$.

In summary, our model is characterized by the data $(C^{IJK}, Y_{I_1\ldots I_k}, T_I{}^J)$ that satisfy the constraints (3.1.1)–(3.1.3). In the next section we show that most of the constraints can be solved by considering semisimple associative algebras.

3.2. Algebraic construction

In this section, we give an algebraic construction of the model data $(C^{IJK}, Y_{I_1\ldots I_k}, T_I{}^J)$ (see [29] and also [43, 44] for a related idea).

Let $\mathcal{R}$ be a real semisimple associative algebra.[3] That is, $\mathcal{R}$ is a linear space over $\mathbb{R}$ with multiplication (denoted by $\times$) that satisfies the associativity, $(B_1 \times B_2) \times B_3 = B_1 \times (B_2 \times B_3)$. If one introduces a basis $\{E_I\}$ as $\mathcal{R} = \bigoplus_I \mathbb{R}\, E_I$, the multiplication is expressed in the form $E_I \times E_J = Y_{IJ}{}^K E_K$, where the structure constants $Y_{IJ}{}^K$ satisfy the relations $Y_{IJ}{}^L Y_{LK}{}^M = Y_{IL}{}^M Y_{JK}{}^L$ due to associativity. The k-hinge tensor $Y_{I_1\ldots I_k}$ can then be constructed from $Y_{IJ}{}^K$ as

$$Y_{I_1\ldots I_k} \equiv Y_{I_1 J_1}{}^{J_k} Y_{I_2 J_2}{}^{J_1} \ldots Y_{I_k J_k}{}^{J_{k-1}}. \tag{3.2.1}$$

It is easy to see that $Y_{I_1\ldots I_k}$ are cyclically symmetric.[4] The two-hinge tensor Y_{IJ} is called the *metric* of $\mathcal{R}$ and will often be denoted by G_{IJ}; $G_{IJ} = Y_{IJ} = Y_{IK}{}^L Y_{JL}{}^k$. It is known [43] that the associative algebra $\mathcal{R}$ is semisimple (i.e., isomorphic to a direct sum of matrix rings) if and only if $G = (G_{IJ})$ has its inverse $G^{-1} \equiv (G^{IJ})$. The constraints (3.1.1) and (3.1.2) can be solved if there exists an involutory *anti*automorphism $T : \mathcal{R} \to \mathcal{R}$. In fact, the coefficients $T_I{}^J$ in $T(E_I) = T_I{}^J E_J$ satisfies (3.1.1) when T is involutory. Furthermore, when T is an antiautomorphism: $T(E_J \times E_I) = T(E_I) \times T(E_J)$, we have the relations $T_I{}^K T_J{}^L Y_{KL}{}^N = Y_{JI}{}^M T_M{}^N$, which ensure (3.1.2) to hold.

Such an antiautomorphism can be naturally constructed when we set the index I to be a double index $I = (i, j)$ $(i, j = 1, \ldots, N)$ in order to assign arrows to the edges of triangles and hinges. To see this, we let $\mathcal{R}$ take the form

$$\mathcal{R} = \mathcal{A} \otimes \bar{\mathcal{A}}, \tag{3.2.2}$$

[3] The following construction does not involve the operation of complex conjugation and thus can be readily generalized to associative algebras over the complex field.

[4] The k-hinge tensor can also be expressed as $Y_{I_1\ldots I_K} = \mathrm{Tr}_{\mathcal{R}}\left(E^{\mathrm{reg}}_{I_1} \cdots E^{\mathrm{reg}}_{I_k}\right)$, where E^{reg}_I are the representation matrix of the basis $\{E_I\}$ in the regular representation of $\mathcal{R}$; $E^{\mathrm{reg}}_I = \left((E^{\mathrm{reg}}_I)^J{}_K = Y_{IK}{}^J\right)$ [43].

where $\mathcal{A}$ and $\bar{\mathcal{A}}$ are linear spaces of the same dimension N. In orfer for concrete description, we fix an isomorphism from $\mathcal{A}$ to $\bar{\mathcal{A}}$ and denote it by σ.[5] We assume that $\mathcal{A}$ is a semisimple associative algebra with multiplication $\times$. We introduce a multiplication (also denoted by $\times$) to $\bar{\mathcal{A}}$ such that $\sigma : \mathcal{A} \to \bar{\mathcal{A}}$ is an algebra *anti*automorphism, $\sigma(a \times b) = \sigma(b) \times \sigma(a)$ $(\forall a,\, b \in \mathcal{A})$.[6] Then $\mathcal{R} = \mathcal{A} \otimes \bar{\mathcal{A}}$ naturally becomes an associative algebra as the tensor product of two associative algebras. The antiautomorphism $T : \mathcal{R} \to \mathcal{R}$ now can be defined such as to map an element $B = \sum b \otimes \bar{b} \in \mathcal{R}$ to

$$T(B) = \sum T(b \otimes \bar{b}) \equiv \sum \sigma^{-1}(\bar{b}) \otimes \sigma(b). \tag{3.2.3}$$

One can easily show that T is certainly an antiautomorphism. We thus find that the constraints (3.1.1) and (3.1.2) can be solved by giving an associative algebra $\mathcal{A}$.

Note that $\mathcal{R}$ can also be thought of as the set of algebra endomorphisms of $\mathcal{A}$ by regarding $\bar{\mathcal{A}}$ as the dual linear space of $\mathcal{A}$:

$$\mathcal{R} = \mathcal{A} \otimes \bar{\mathcal{A}} = \mathrm{End}\,\mathcal{A}. \tag{3.2.4}$$

One then can also define an antiautomorphism $\bar{T}$ for the dual of $\mathcal{R}$, $\bar{\mathcal{R}} \equiv \bar{\mathcal{A}} \otimes \mathcal{A}$, such that the following relation holds:

$$\langle \bar{T}(A),\, T(B) \rangle = \langle A,\, B \rangle \quad (\forall A \in \bar{\mathcal{R}},\, \forall B \in \mathcal{R}), \tag{3.2.5}$$

where $\langle\ ,\ \rangle$ is the paring between $\bar{\mathcal{R}}$ (the dual of $\mathcal{R}$) and $\mathcal{R}$.

We rephrase the above construction in terms of the bases $\{e_i\}$ and $\{\bar{e}^i\}$ of $\mathcal{A}$ and $\bar{\mathcal{A}}$:

$$\mathcal{A} = \bigoplus_{i=1}^{N} \mathbb{R}\, e_i, \qquad \bar{\mathcal{A}} = \bigoplus_{i=1}^{N} \mathbb{R}\, \bar{e}^i. \tag{3.2.6}$$

We first represent the isomorphism σ as

$$\sigma(e_i) = \sigma_{ij}\, \bar{e}^j, \qquad \sigma^{-1}(\bar{e}^i) = \sigma^{ij}\, e_j, \tag{3.2.7}$$

and write the structure constants of the multiplication on $\mathcal{A}$ as $e_i \times e_j = y_{ij}{}^k e_k$. Then those of $\bar{\mathcal{A}}$ (appearing in $\bar{e}^i \times \bar{e}^j = \bar{y}^{ij}{}_k \bar{e}^k$) are determined from the requirement of antihomomorphism, $\sigma(e_i \times e_j) = \sigma(e_j) \times \sigma(e_i)$, to be

$$\bar{y}^{ij}{}_k = \sigma^{il} \sigma^{jm} y_{ml}{}^n \sigma_{nk}. \tag{3.2.8}$$

If we take the basis of $\mathcal{R}$ to be $E_I = E_i{}^j = e_i \otimes \bar{e}^j$, then the structure constants $Y_{I_1 I_2}{}^{I_3}$ are given by

$$Y_{I_1 I_2}{}^{I_3} = Y_{i_1}{}^{j_1}{}_{i_2}{}^{j_2\, i_3}{}_{j_3} = y_{i_1 i_2}{}^{i_3}\, \bar{y}^{j_1 j_2}{}_{j_3}, \tag{3.2.9}$$

[5] σ can be taken arbitrarily because it can be absorbed into an automorphism of $\mathcal{A}$ or $\bar{\mathcal{A}}$.

[6] Such multiplication exists uniquely for a given σ (see (3.2.8)).

from which the k-hinge tensor is given by

$$Y_{I_1\dots I_k} = Y_{i_1}{}^{j_1}\cdots{}_{i_k}{}^{j_k} = y_{i_1\dots i_k}\,\bar{y}^{j_1\dots j_k} \tag{3.2.10}$$

with

$$y_{i_1\dots i_k} \equiv y_{i_1 j_1}{}^{j_k} y_{i_2 j_2}{}^{j_1}\dots y_{i_k j_k}{}^{j_{k-1}}, \tag{3.2.11}$$

$$\bar{y}^{i_1\dots i_k} \equiv \bar{y}^{i_1 j_1}{}_{j_k}\bar{y}^{i_2 j_2}{}_{j_1}\dots\bar{y}^{i_k j_k}{}_{j_{k-1}} = \sigma^{i_1 j_1}\cdots\sigma^{i_k j_k} y_{j_k\dots j_1}. \tag{3.2.12}$$

In particular, the metric of $\mathcal{R}$ takes the form

$$G_{I_1 I_2} = G_{i_1}{}^{j_1}{}_{i_2}{}^{j_2} = g_{i_1 i_2}\,\bar{g}^{j_1 j_2}, \tag{3.2.13}$$

where $g_{i_1 i_2}$ and $\bar{g}^{j_1 j_2}$ are the metrics of $\mathcal{A}$ and $\bar{\mathcal{A}}$, respectively; $g_{i_1 i_2} \equiv y_{i_1 k}{}^{\ell}\, y_{i_2 \ell}{}^{k}$, $\bar{g}^{j_1 j_2} = \bar{y}^{j_1 k}{}_{\ell}\,\bar{y}^{j_2 \ell}{}_{k}$.[7] We easily see that $\mathcal{R}$ is semisimple if $\mathcal{A}$ is, because $G_{I_1 I_2}$ has its inverse when $g_{i_1 i_2}$ does (and so does $\bar{g}^{j_1 j_2}$).

The antiautomorphism $T_I{}^J$ is now expressed as $T(e_i\otimes\bar{e}^j) \equiv \sigma^{-1}(\bar{e}^j)\otimes\sigma(e_i) = \sigma^{jk}\sigma_{il}\, e_k\otimes\bar{e}^l$, that is,

$$T_{I_1}{}^{I_2} = T_{i_1}{}^{j_1 i_2}{}_{j_2} = \sigma_{i_1 j_2}\,\sigma^{j_1 i_2}. \tag{3.2.14}$$

For the dual algebra $\bar{\mathcal{R}} = \bar{\mathcal{A}}\otimes\mathcal{A}$, regarding $\{\bar{e}^i\}$ as the dual basis of $\{e_i\}$, we set a basis of $\bar{\mathcal{R}}$ to be $\bar{E}^I = \bar{E}^i{}_j = \bar{e}^i\otimes e_j$, which leads to the pairing $\langle\bar{E}^{I_1},\, E_{I_2}\rangle = \delta^{i_1}_{i_2}\,\delta^{j_2}_{j_1}$. Then the antiautomorphism $\bar{T}$ on $\bar{\mathcal{R}}$ is expressed as $\bar{T}(\bar{E}^I) \equiv \bar{E}^J(T^{-1})_J{}^I = \bar{E}^J T_J{}^I$.

The dynamical variables A_I and B^I in (3.1.5) can be regarded as elements of $\bar{\mathcal{R}}$ and $\mathcal{R}$, respectively:

$$A = A_I\bar{E}^I = A_i{}^j\,\bar{e}^i\otimes e_j \in \bar{\mathcal{R}}, \qquad B = B^I E_I = B^i{}_j\, e_i\otimes\bar{e}^j \in \mathcal{R}. \tag{3.2.15}$$

The condition (3.1.6) is then expressed as $\bar{T}(A) = A$, $T(B) = B$. With these double indices, the action (3.1.5) is written as

$$S = \frac{1}{2}A_i{}^j B^i{}_j - \frac{\lambda}{6}C^{i\ k\ m}_{j\ l\ n}A_i{}^j A_k{}^l A_m{}^n - \sum_{k\geq 2}\frac{\mu_k}{2k}B^{i_1}{}_{j_1}\dots B^{i_k}{}_{j_k}y_{i_1\dots i_k}\bar{y}^{j_1\dots j_k}, \tag{3.2.16}$$

where the tensor $C^{i\ k\ m}_{j\ l\ n}$ is arbitrary as long as it satisfies the condition (3.1.3). It is often convenient to use $A_{ij} \equiv \sigma_{jk}A_i{}^k$, $B^{ij} \equiv B^i{}_k\sigma^{kj}$, and $C^{ijklmn} \equiv C^{i\ k\ m}_{j'\ l'\ n'}\sigma^{j'j}\sigma^{l'l}\sigma^{n'n}$. Then the conditions (3.1.3) and (3.1.6) can be rewritten to the form where σ does not appear:

$$C^{ijklmn} = C^{nmlkji}, \tag{3.2.17}$$

$$A_{ij} = A_{ji}, \quad B^{ij} = B^{ji}. \tag{3.2.18}$$

[7] Note that the cyclically symmetric tensor $y_{i_1\dots i_k}$ can also be written as

$$y_{i_1 i_2 i_3} = y_{i_1 i_2}{}^{j_3} g_{j_3 i_3}, \quad y_{i_1\dots i_k} = y_{i_1 j_1 l_1}\, g^{j_1 l_2} y_{i_2 j_2 l_2}\, g^{j_2 l_3}\dots y_{i_k j_k l_k}\, g^{j_k l_1}.$$

The action then becomes

$$S = \frac{1}{2}A_{ij}B^{ij} - \frac{\lambda}{6}C^{ijklmn}A_{ij}A_{kl}A_{mn} - \sum_{k\geq 2}\frac{\mu_k}{2k}B^{i_1j_1}\cdots B^{i_kj_k}y_{i_1\ldots i_k}y_{j_k\ldots j_1}. \tag{3.2.19}$$

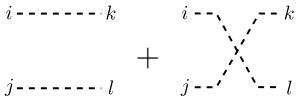
Thus, a set of models can be defined by giving semisimple associative algebras $\mathcal{A}$ and the tensors C^{ijklmn} satisfying (3.2.17). The isomorphism $\sigma : \mathcal{A} \to \bar{\mathcal{A}}$ can be taken arbitrarily and is regarded as a sort of gauge freedom in choosing the basis of $\mathcal{A}$ or $\bar{\mathcal{A}}$.

3.3. Feynman rules

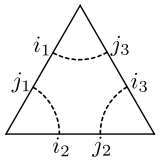
As stated in the last section, the tensor C^{ijklmn} in (3.2.19) can be chosen arbitrarily as long as it satisfies the condition (3.2.17). For a while, we set it to be

$$C^{ijklmn} = g^{jk}g^{lm}g^{ni}, \tag{3.3.1}$$

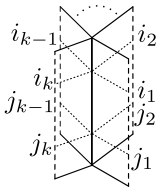
which one can easily show to satisfy (3.2.17).[8] Then the Feynman rules of the action (3.2.19) become

propagator :


$$\sim \langle A_{ij}B^{kl}\rangle = \delta_i{}^k\delta_j{}^l + \delta_i{}^l\delta_j{}^k, \tag{3.3.2}$$

triangle :


$$\sim \lambda\, g^{j_1i_2}\, g^{j_2i_3}\, g^{j_3i_1}, \tag{3.3.3}$$

k-hinge :


$$\sim \mu_k\, y_{i_1\ldots i_k}\, y_{j_k\ldots j_1}. \tag{3.3.4}$$

Recall that the arrows are now expressed with double lines, and thus, when the first (or second) term of the propagator (3.3.2) is used the edges are glued in the same (or opposite) direction.

The free energy of this model takes the form

$$\log Z = \sum_{\gamma}\frac{1}{S(\gamma)}\,\lambda^{s_2(\gamma)}\Big(\prod_{k\geq 2}\mu_k^{s_1^k(\gamma)}\Big)\,\mathcal{F}(\gamma). \tag{3.3.5}$$

Here, the sum $\sum_\gamma$ is taken over all possible connected Feynman diagrams $\{\gamma\}$, and $S(\gamma)$ is the symmetry factor of diagram γ. $s_2(\gamma)$ is the number of triangles, and $s_1^k(\gamma)$ the number of

[8] This choice (3.3.1) will be slightly modified when we introduce a color structure to our models. Other choices will be discussed in Sect. 4.1.

k-hinges. $\mathcal{F}(\gamma)$ denotes the product of $y_{i_1 \cdots i_k}$ and g^{ij} with the indices contracted according to a given Wick contraction, and we call $\mathcal{F}(\gamma)$ the *index function of diagram* γ. We here regard two diagrams as being the same if the indices are contracted in the same manner. The numerical coefficients in the action (3.2.19) are chosen such that independent diagrams give only the symmetry factors to the free energy, by taking into account the symmetry (rotation and flip) of triangles and hinges. We stress that the free energy will have a different form from (3.3.5) if we make a different choice for C^{ijklmn} other than (3.3.1).

We now show that the index function $\mathcal{F}(\gamma)$ can be expressed as the product of the contributions from *two-dimensional* surfaces, each surface enclosing a vertex of diagram γ. To see this, we first note from the Feynman rules (3.3.2)–(3.3.4) that even a connected Feynman diagram generally gives disconnected networks of index lines. This is because each hinge has a pair of junction points as for index lines and two index lines out of the same edge of a hinge can enter two different hinges after passing through an adjacent triangle (see Fig. 3.4). We further note that the index lines on two different hinges can be connected

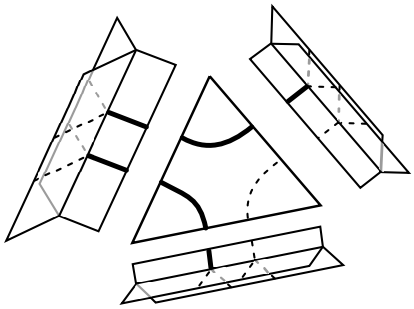


Figure 3.4: A part of index networks [1]. Two index lines (depicted in bold lines) come out of the same edge of the left hinge and enter the right adjacent triangle. The upper index line then leaves the triangle and enters the upper hinge, while the lower index line enters the lower hinge.

(through an intermediate triangle) if and only if the hinges share the same vertex of γ. This means that the *connected* index networks have a one-to-one correspondence to the vertices of γ. We thus find that the index function $\mathcal{F}(\gamma)$ of diagram γ is the product of the contributions from connected index networks (each assigned to a vertex of γ) and has the form

$$\mathcal{F}(\gamma) = \prod_{v:\,\text{vertex of }\gamma} \zeta(v). \tag{3.3.6}$$

We also call $\zeta(v)$ the index function (more precisely, the *index function of vertex* v). Note that every connected index network takes the form of a polygonal decomposition of a closed surface (not necessarily a sphere and may include monogons or digons), where a k-valent junction (or k-junction) corresponds to a k-hinge where k index lines meet (see Fig. 3.5).[9]

[9]In order to avoid possible confusions between the terms for Feynman diagrams and those for index

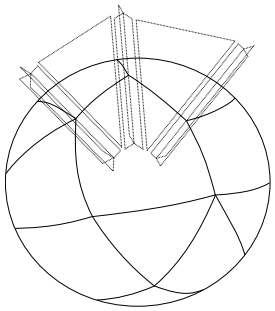


Figure 3.5: Index network around a vertex [1]. It represents a polygonal decomposition of a closed surface (not necessarily a sphere) around a vertex. A k-valent junction in the index network corresponds to a k-hinge in the original diagram, where k index lines meet. A segment connecting two junctions in the index network corresponds to an intermediate triangle between two hinges. A closed index loop with ℓ segments (ℓ-gon) corresponds to a three-dimensional region surrounded by ℓ triangles (ex. a corner of a tetrahedron).

We here make a few comments. The first comment is on the uniqueness in interpreting an index network as a polygonal decomposition of a closed surface. In fact, if one regards an index network simply as a wire frame (i.e., as a collection of segments), then it is not a unique procedure to assign polygonal faces in the frame such that the resulting configuration forms a closed surface.[10] However, our index network is not simply a wire frame, and has the information on how the indices are contracted (i.e. the order of indices appearing in $y_{i_1 \cdots i_k}$). We thus can uniquely assign faces to the holes of the index network by carefully following the contraction of indices. We will see in Sect. 3.6 that the assignment is straightforward when models are given by matrix rings as the defining associative algebras.

The second comment is on the manifoldness of a diagram γ. Since there is a two-dimensional surface around each vertex of γ, we can say that there is a three-dimensional (solid) cone at each vertex, the base and apex of a cone being the connected index network around a vertex and the vertex itself, respectively. For example, if an index network has the topology of two-sphere S^2, then the corresponding cone is a 3-dimensional ball B^3. These cones characterize the neighborhoods of the vertices of the diagram γ.[11] Note that γ represents a three-dimensional (combinatorial) manifold if γ gives a tetrahedral decomposition and the neighborhood of every vertex is homeomorphic to B^3. In Sect. 3.6, by taking $\mathcal{A}$ to be a matrix ring and introducing a color structure to the models in Sect. 3.7, we show that

networks, we call the vertices and edges in the index network the junctions and segments, respectively.

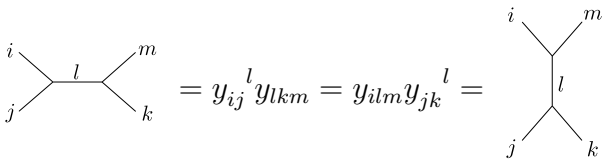
[10]For example, there arises such ambiguity if a diagram includes a triangle shared by more than three tetrahedra, as in n-simplex ($n \geq 5$) which can be constructed from triangles and multiple hinges.

[11]Some diagram (as the one in footnote 10) may be better regarded as being a higher dimensional object. In this case, the above three-dimensional cone will be treated as a part of the neighborhood of a vertex in the higher dimensional object.

the set of possible Feynman diagrams can be drastically reduced such that only (and all of the) manifolds are generated.

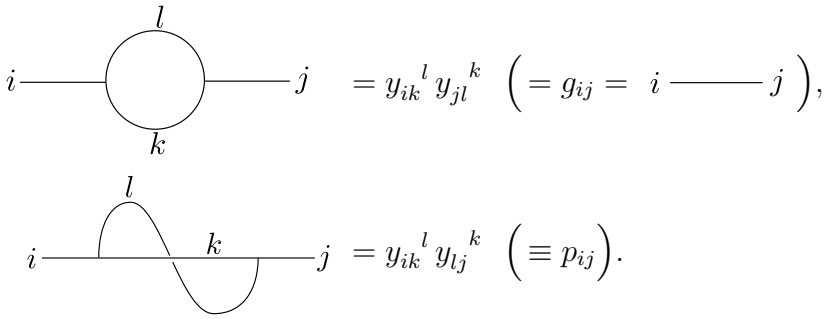
3.4. Evaluation of diagrams

The index function $\mathcal{F}(\gamma)$ can be easily evaluated by deforming each connected index network with the use of the associativity of $y_{ij}{}^{k} = y_{ijl}\, g^{lk}$ [43]:



(3.4.1)

In deformation there may appear two kinds of index loops:



(3.4.2)

(3.4.3)

The former index loop diagram can be replaced by a single solid line, while the loop in the latter index diagram cannot be removed. Actually p_{ij} (or more precisely, $p_i^j \equiv p_{ik}g^{kj}$) is a projector to the center of algebra $\mathcal{A}$, $Z(\mathcal{A})$, as can be checked easily [43]. If a given connected index network does not produce a projector p_{ij} in the process of deformation, the index network can always be deformed to a single circle after repeatedly using (3.4.1) and (3.4.2) and gives the value $g_{ij}g^{ij} = N$, where N is the dimension of $\mathcal{A}$. On the other hand, if a given connected index network admits the appearance of a projector p_{ij}, the value of the index network is generally less than N.[12]

Note that the two deformations (3.4.1) and (3.4.2) are actually the local moves of two-dimensional surfaces.[13] Therefore, the index function $\zeta(v)$ of vertex v gives a two-dimensional topological invariant defined by the associative algebra $\mathcal{A}$ [43], and thus has the form $\zeta(v) = \mathcal{I}_{g(v)}$, where $g(v)$ is the genus of the network.[14] Thus the index function of

[12]For example, $p_{ij}g^{ij} = p_i^i$ gives the linear dimension of $Z(\mathcal{A})$.

[13]Namely, any two index networks can be obtained from each other by a repetitive use of (3.4.1) and (3.4.2) if and only if the two index networks represent two-dimensional surfaces of the same topology [43].

[14]We already know some of the general results, $\mathcal{I}_{g=0} = \mathrm{dim}\mathcal{A} = N$, $\mathcal{I}_{g=1} = \mathrm{dim}Z(\mathcal{A})$.

diagram γ is expressed as

$$\mathcal{F}(\gamma) = \prod_{v:\,\text{vertex}} \zeta(v) = \prod_{v:\,\text{vertex}} \mathcal{I}_{g(v)}. \tag{3.4.4}$$

3.5. Examples

In this section we give a few examples of the diagrams generated in our models. If our aim is to apply the models to three-dimensional gravity, we should be able to assign three-dimensional volume to each diagram, and it is thus preferable that the diagrams can be regarded as collections only of tetrahedra. However, as we see in the examples below, there arise a lot of undesired diagrams (i.e. the diagrams which cannot be regarded as collections of tetrahedra). We will show in the Sect. 3.7 that such undesired diagrams can be automatically excluded by taking specific associative algebras and modifying the form (3.3.1), with an appropriate limit of parameters.

3.5.1. Diagrams representing tetrahedral decompositions of manifolds

First, we consider a diagram which represents a tetrahedral decomposition of three-dimensional sphere S^3 (see Fig. 3.6). This is the boundary of the so-called 5-cell or a 4-simplex and can be constructed from five tetrahedra. Note that the diagram has ten triangles, ten 3-hinges

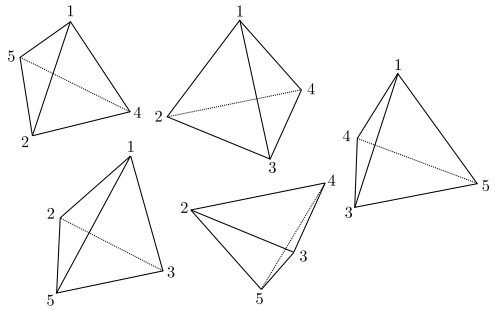
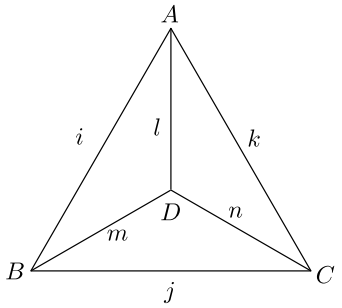


Figure 3.6: A decomposition of S^3 with five tetrahedra [1]. The tetrahedra are glued together at their faces so that each of the points $1, \ldots, 5$ represents a single vertex.

and five vertices. All the index networks around vertices have the same topology and give triangular decompositions of S^2 as in Fig. 3.7. Thus, the neighborhood of each vertex is homeomorphic to B^3. Since every index network can be deformed to a single circle, the index function of each vertex takes the value N; $\zeta(v) = N = \mathcal{I}_{g(v)=0}$. Thus, the contribution



$$= y_{i_A l_A k_A}\, y_{j_B m_B i_B}\, y_{k_C n_C j_C}\, y_{l_D m_D n_D}$$
$$g^{i_A i_B} g^{j_B j_C} g^{k_C k_A} g^{l_A l_D} g^{m_B m_D} g^{n_C n_D}$$

Figure 3.7: A contraction of indices around a vertex [1]. This represents a triangular decomposition of S^2 and gives the value N.

from this diagram to the free energy is given by

$$\frac{1}{S}\,\lambda^{10}\mu_3^{10}\,(\mathcal{I}_{g=0})^5 = \frac{1}{S}\,\lambda^{10}\mu_3^{10}\,N^5, \tag{3.5.1}$$

where S is the symmetry factor of the diagram.

The next example is a diagram which represents a tetrahedral decomposition of three-dimensional torus T^3 (see Fig. 3.8). The diagram has twelve triangles, four 4-hinges and

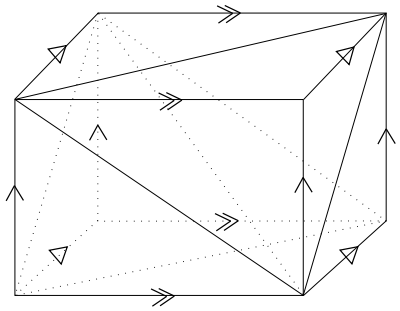


Figure 3.8: A tetrahedral decomposition of T^3 [1]. This is made by identifying the boundaries of a cuboid which consists of six tetrahedra.

three 6-hinges. It has only a single vertex due to the identification in the diagram. The index network around the vertex also represents S^2 as in the previous example for S^3. Thus, the contribution from this diagram is given by

$$\frac{1}{S}\,\lambda^{12}\mu_4^4\,\mu_6^3\,\mathcal{I}_{g=0} = \frac{1}{S}\,\lambda^{12}\mu_4^4\,\mu_6^3\,N. \tag{3.5.2}$$

We can easily generalize the above results to such diagrams that represent tetrahedral decompositions of three-dimensional closed manifolds. Since the neighborhood of every vertex is homeomorphic to B^3, the contribution from such a diagram to the free energy is given by

$$\frac{1}{S}\,\lambda^{s_2}\Big(\prod_{k\geq 2}\mu_k^{s_1^k}\Big)\,(\mathcal{I}_{g=0})^{s_0} = \frac{1}{S}\,\lambda^{s_2}\Big(\prod_{k\geq 2}\mu_k^{s_1^k}\Big)\,N^{s_0}, \tag{3.5.3}$$

where s_2, s_1^k and s_0 represent, respectively, the number of triangles, k-hinges and vertices of the diagram. Note that since the topology of three-dimensional manifolds cannot be distinguished by s_2, s_1^k and s_0 alone,[15] it can happen that topologically different manifolds give contributions of the same form. However, we in principle can distinguish the topology by carefully looking at the way of tetrahedral decompositions, although this is usually a tedious task. Another way to examine the topology of diagrams is to evaluate a set of topological invariants of each diagram as in [29].

3.5.2. Diagrams corresponding to pseudomanifolds

Our models also generate diagrams that have vertices whose neighborhoods are not three-dimensional ball B^3. One of such diagrams is depicted in Fig. 3.9, which consists of four tetrahedra, eight triangles, five edges and three vertices. The neighborhood of vertex 3 is homeomorphic to B^3, but that of vertex 1 (and also that of vertex 2) has the topology of cone over T^2. In fact, the index network around vertex 1 gives a polygonal decomposition of two-dimensional torus T^2.

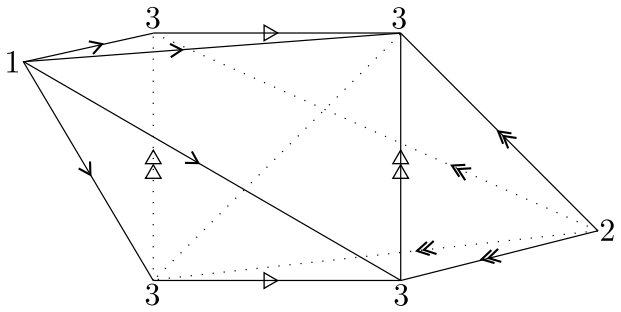


Figure 3.9: A diagram which does not give a manifold [1]. The neighborhood of of vertex 3 is B^3 but that of vertex 1 (and also 2) is a cone over T^2.

One can check that the Euler characteristic of the diagram is not zero. Thus, this diagram should not give a manifold (but still gives a pseudomanifold). The contribution from this diagram to the free energy can be evaluated to be

$$\frac{1}{S}\,\lambda^8\,\mu_4^3\,\mu_6^2\,\mathcal{I}_{g=0}\,(\mathcal{I}_{g=1})^2, \tag{3.5.4}$$

where $\mathcal{I}_{g=0}$ comes from the index network around vertex 3 and equals $g_{ij}g^{ij} = N$. By contrast, two of $\mathcal{I}_{g=1}$ come from vertices 1 and 2, and have the value $p_{ij}g^{ij} = p_i^i$, which is the linear dimension of the center of $\mathcal{A}$.

[15] We can read the number of tetrahedra, s_3, since the Euler characteristic of three-dimensional closed manifold is zero, $s_0 - \sum_k s_1^k + s_2 - s_3 = 0$.

3.5.3. Diagrams including singular cells

There also arise diagrams which do not give tetrahedral decompositions. A few simple diagrams are depicted in Fig. 3.10. Although they have the topology of S^3, it is not suitable to assign three-dimensional volume.

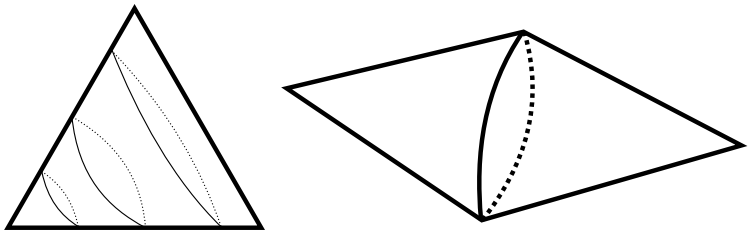


Figure 3.10: Diagrams with singular cells [1].

3.6. Example of associative algebra: matrix ring

In this section, we consider matrix rings as the defining associative algebras of the models. In next section we show that such models can be constructed that generate only manifolds as Feynman diagrams, by introducing a color structure to the models and letting the associative algebras have centers whose dimensions play the role of free parameters (to count the number of vertices).

Matrix ring $M_n(\mathbb{R})$ is the set of real-valued matrices of size n. This is an associative algebra with the same rules of addition, scalar product and multiplication as those of matrices. Note that $\mathcal{A}$ has the linear dimension $N = n^2$. Matrix ring is one of the simplest semisimple associative algebras because any semisimple associative algebra is isomorphic to a direct sum of matrix rings. In this section we analyze a model where $\mathcal{A}$ is set to be a matrix ring $M_n(\mathbb{R})$. We take its basis to be $\{e_{ab}\}$ $(a, b = 1, \ldots, n)$, where e_{ab} is a matrix unit whose (c, d) element is given by $(e_{ab})_{cd} = \delta_{ac}\,\delta_{bd}$. The structure constants can be read from the multiplication rule of matrices:

$$e_{ab} \times e_{cd} = \delta_{bc}\, e_{ad} = \delta^e_a\, \delta_{bc}\, \delta^f_d\, e_{ef} \equiv y_{abcd}{}^{ef} e_{ef}. \tag{3.6.1}$$

We stress that the double index (a, b) corresponds to the single index i $(i = 1, \ldots, N)$ in Sect. 3.2.[16] One can compute $y_{i_1 i_2 \ldots i_k} = y_{a_1 b_1 a_2 b_2 \ldots a_k b_k}$ and $g^{ij} = g^{abcd}$ as

$$y_{a_1 b_1 a_2 b_2 \ldots a_k b_k} = n\, \delta_{b_1 a_2} \delta_{b_2 a_3} \cdots \delta_{b_k a_1}, \qquad g^{abcd} = \frac{1}{n} \delta_{ad} \delta_{bc}. \tag{3.6.2}$$

[16]The index I in Sect. 3.1 thus becomes a quadruple index as $I = (i, j) = (a, b, c, d)$.

By setting the tensor C^{ijklmn} as in (3.3.1):

$$C^{a_1b_1c_1d_1a_2b_2c_2d_2a_3b_3c_3d_3} = \frac{1}{n^3}\,\delta^{d_1a_2}\delta^{c_1b_2}\delta^{d_2a_3}\delta^{c_2b_3}\delta^{d_3a_1}\delta^{c_3b_1}, \tag{3.6.3}$$

the action (3.2.19) has the form

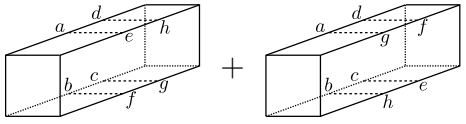
$$S = \frac{1}{2}A_{abcd}B^{abcd} - \frac{\lambda}{6n^3}A_{bacd}A_{dcef}A_{feab} - \sum_{k\geq 2}\frac{n^2\mu_k}{2k}B^{a_1a_2b_2b_1}B^{a_2a_3b_3b_2}\cdots B^{a_ka_1b_1b_k}, \tag{3.6.4}$$

where A and B satisfy the following relations because of the symmetry property (3.2.18):

$$A_{abcd} = A_{cdab}, \quad B^{abcd} = B^{cdab}. \tag{3.6.5}$$

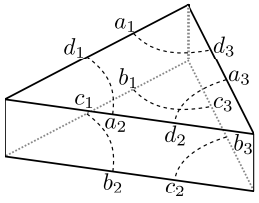
The Feynman rules for the action (3.6.4) can be expressed with quadruple lines as follows:

propagator :



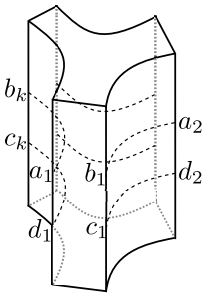
$$\sim \langle A_{abcd}B^{efgh}\rangle = \delta_a{}^e\delta_b{}^f\delta_c{}^g\delta_d{}^h + \delta_a{}^g\delta_b{}^h\delta_c{}^e\delta_d{}^f, \tag{3.6.6}$$

triangle :



$$\sim \frac{\lambda}{n^3}\,\delta^{d_1a_2}\delta^{c_1b_2}\ldots\delta^{d_3a_1}\delta^{c_3b_1}, \tag{3.6.7}$$

k-hinge :



$$\sim n^2\mu_k\,\delta_{b_1a_2}\delta_{c_1d_2}\ldots\delta_{b_ka_1}\delta_{c_kd_1}. \tag{3.6.8}$$

Note that each of the index lines in (3.3.2)–(3.3.4) becomes a double line. Moreover, the index line dose not have branch points in this case due to the index structure of hinges [see (3.6.8)]. Thus, as depicted in Fig. 3.11, the identification of the index network with a polygonal decomposition of two-dimensional surface can be done automatically (and uniquely)[17] [although this identification can also be carried out uniquely even when $\mathcal{A}$ is not a matrix ring, as argued in the first comment following (3.3.6)].

The contribution from each index network to the free energy can be calculated just as in the standard matrix model. To see this, we first note that each polygon gives a factor of n

[17]Each polygonal face is specified as the region bounded by a closed loop for index a.

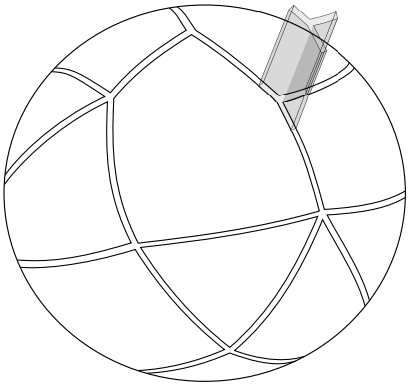


Figure 3.11: Index network around a vertex v when $\mathcal{A}$ is a matrix ring $M_n(\mathbb{R})$ [1]. Everything is the same as Fig. 3.5 except that the index lines are now double lines. The index network represents a closed oriented surface (not necessarily a sphere).

because each index loop (i.e. the index contraction with respect to one of the double index) gives $\delta_a^a = n$. We also see from the coefficients in (3.6.7) and (3.6.8) that each segment in the polygonal decomposition gives n^{-1} (one-third contribution from a triangle) and each junction gives n (one-half contribution from a hinge). In total, the contribution from the index network around vertex v in the original diagram is given by $n^{\#(\text{polygon})-\#(\text{segment})+\#(\text{junction})} = n^{2-2g(v)}$, where $g(v)$ is the genus of the index network around v. One can easily see that an insertion of the projector p_{ij}, (3.4.3), into the diagram corresponds to attaching a handle to the index network (as in [43]) and decreases the power of n by two.

3.7. restriction to tetrahedral decompositions

In this section we give a prescription to choose the parameters in our models such that only tetrahedral decompositions of three-dimensional manifolds are generated as Feynman diagrams. The strategy is given as follows: As we will show in the proof of the theorem in Sect. 3.7.3, one can ensure a diagram to be a tetrahedral decomposition if the index network around every vertex is a *triangular* decomposition of two-dimensional surface. This condition can be realized by introducing a color structure to the models, as we will carry out in Sect. 3.7.1.

Furthermore, the manifoldness of the resulting diagrams can be ensured by appropriately choosing the defining associative algebra $\mathcal{A}$ such that the following two conditions are realized: (i) the number of vertices can be fixed by using free parameters in $\mathcal{A}$, and (ii) $\mathcal{I}_0(v) \gg \mathcal{I}_g(v)$ for $g \geq 1$. In fact, due to the expression $\mathcal{F}(\gamma) = \prod_v \mathcal{I}_{g(v)}$ [see (3.4.4)], the dominant contributions come from the diagrams whose index networks all have the topology of two-sphere (and thus the neighborhood of every vertex has the topology of three-ball),

namely, from the diagrams that represent (combinatorial) manifolds. If $\mathcal{A}$ does not have free parameters to fix the number of vertices, we extend $\mathcal{A}$ as needed. This extension will be carried out for matrix rings in Sect. 3.7.2.

3.7.1. Color structure

In Sect. 3.5.3 we argued that undesired diagrams appear in our models. In this subsection we show that they can be excluded by introducing a "color structure" to our models.

Let the size n of matrices be a multiple of three, $n = 3m$. We then modify the tensor (3.6.3) to

$$C^{a_1b_1c_1d_1a_2b_2c_2d_2a_3b_3c_3d_3} = \frac{1}{n^3}\,\omega^{d_1a_2}\omega^{b_2c_1}\omega^{d_2a_3}\omega^{b_3c_2}\omega^{d_3a_1}\omega^{b_1c_3}, \tag{3.7.1}$$

where ω is a permutation matrix of the form

$$\omega \equiv \begin{pmatrix} 0 & 1_m & 0 \\ 0 & 0 & 1_m \\ 1_m & 0 & 0 \end{pmatrix}, \qquad 1_m : m \times m \,\text{unit matrix}. \tag{3.7.2}$$

This modification[18] corresponds to inserting ω and $\omega^{-1} = \omega^T$ in a pair into two index lines on every segment in each index network (see Fig. 3.12). Note that only ω (not ω^{-1}) are accumulated when following the arrows in each index line. Thus, the value of a closed index loop forming ℓ-gon changes from $\operatorname{tr} 1_n = n = 3m$ to

$$\operatorname{tr}(\omega^\ell) = \begin{cases} n & (\ell = 0 \mod 3) \\ 0 & (\ell \neq 0 \mod 3). \end{cases} \tag{3.7.3}$$

We thus see that the index function of a diagram gives a nonvanishing value only when the index network around every vertex has a polygonal decomposition where the number of segments of each polygon is a multiple of three.

[18] Although we only discuss the case $\mathcal{A} = M_{3m}(\mathbb{R})$, we can also introduce the color structure to other algebras by taking the tensor product of the form $\mathcal{R} = (\mathcal{A} \otimes M_3(\mathbb{R})) \otimes (\bar{\mathcal{A}} \otimes \overline{M_3(\mathbb{R})})$. Note that $M_m(\mathbb{R}) \otimes M_3(\mathbb{R}) = M_{3m}(\mathbb{R})$. Then, the variables A and B are expressed as $A_{ij(abcd)} = A_{ji(cdab)}$ and $B^{ij(abcd)} = B^{ji(cdab)}$, and the action has the form

$$\begin{aligned} S = {} & \frac{1}{2}\,A_{ij(abcd)}\,B^{ij(abcd)} \\ & - \frac{\lambda}{6\cdot 3^3}\,A_{ij(a_1b_1c_1d_1)}\,g^{jk}A_{kl(a_2b_2c_2d_2)}\,g^{lm}A_{mn(a_3b_3c_3d_3)}\,g^{ni}\,\omega^{d_1a_2}\omega^{b_2c_1}\omega^{d_2a_3}\omega^{b_3c_2}\omega^{d_3a_1}\omega^{b_1c_3} \\ & - \sum_{k\geq 2}\frac{3^2\mu_k}{2k}\,B^{i_1j_1(a_1a_2b_2b_1)}\cdots B^{i_kj_k(a_ka_1b_1b_k)}\,y_{i_1\ldots i_k}\,y_{j_k\ldots j_1}. \end{aligned}$$

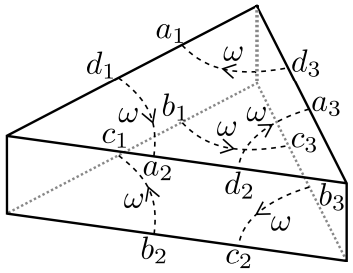


Figure 3.12: Triangles with the color structure [1]. ω has the value ω^{da} when it is inserted into the index line from d to a. Note that $\omega^{bc} = (\omega^{-1})^{cb}$.

Note that such polygonal decompositions with nonvanishing index functions have the following dependence on the coupling constants. Suppose that the index network around vertex v has $t_2(v)$ polygons, $t_1(v)$ segments and $t_0(v)$ junctions. Here, $t_2(v) = \sum_\ell t_2^\ell(v)$ with $t_2^\ell(v)$ the number of ℓ-gons, and $t_0(v) = \sum_k t_0^k(v)$ with $t_0^k(v)$ the number of k-junctions. It is easy to see that the function $d(v) \equiv 2t_1(v) - 3t_2(v)$ can be expressed as $d(v) = \sum_\ell (\ell - 3)\, t_2^\ell(v)$. Thus $d(v)$ is nonnegative for the C's in (3.7.1) because monogons and digons are excluded due to the color structure [i.e., $t_2^{\ell=1}(v) = t_2^{\ell=2}(v) = 0$]. Recalling that the contribution from each diagram is given by

$$\frac{1}{S}\,\lambda^{s_2}\Big(\prod_{k\geq 2}\mu_k^{s_1^k}\Big)\prod_{v:\,\text{vertex}} n^{2-2g(v)}, \tag{3.7.4}$$

and noting that the identification rule of the polygonal decompositions gives the relations

$$s_2 = \frac{1}{3}\sum_v t_1(v), \quad s_1^k = \frac{1}{2}\sum_v t_0^k(v), \tag{3.7.5}$$

we find another expression of (3.7.4):

$$\frac{1}{S}\,\lambda^{s_2}\Big(\prod_{k\geq 2}\mu_k^{s_1^k}\Big)\prod_{v:\,\text{vertex}} n^{2-2g(v)} = \frac{1}{S}\prod_{v:\,\text{vertex}}\Big[\Big[\prod_{k\geq 2}(\lambda^2\mu_k)^{\frac{1}{2}t_0^k(v)}\Big]\Big(\frac{n}{\lambda}\Big)^{2-2g(v)}\Big(\frac{1}{\lambda}\Big)^{\frac{1}{3}d(v)}\Big]. \tag{3.7.6}$$

Therefore, if we expand the free energy around $\lambda = \infty$ with $\lambda^2\,\mu_k$ and n/λ being fixed, the leading contribution comes from such diagrams that satisfy $d(v) = 0$ for every vertex v, namely, from the diagrams where every index network forms a *triangular* decomposition.

3.7.2. Counting the number of vertices

One may think from (3.7.6) that it would be possible by taking a limit $n/\lambda \to \infty$ to single out the diagrams where the index networks are all homeomorphic to two-sphere S^2. However, this is not the case. For example, let us consider a diagram which includes an index network forming a two-torus T^2. Since the index network gives the contribution of $(n/\lambda)^0 = 1$,

we cannot distinguish a diagram whose vertices all give index networks homeomorphic to S^2 from a diagram which has the same number of such vertices whose index networks are homeomorphic to S^2 but also has extra vertices whose index networks are homeomorphic to T^2, because the contributions from the two diagrams to the free energy have the same form.

This problem comes from the fact that we cannot control the number of vertices only with the coupling constants existing in the model with $\mathcal{A} = M_n(\mathbb{R})$. However, this can be remedied by setting the algebra $\mathcal{A}$ to be the direct sum of K copies of matrix ring $\mathcal{A}_0 = M_n(\mathbb{R})$,[19]

$$\mathcal{A} = \underbrace{\mathcal{A}_0 \oplus \cdots \oplus \mathcal{A}_0}_{K \text{ copies}} = K\mathcal{A}_0. \tag{3.7.7}$$

In fact, the index function of a diagram with s_0 vertices becomes proportional to K^{s_0} since the index network around each vertex gives a factor of K independently, and thus (3.7.6) changes to[20]

$$\frac{1}{S}\prod_{v:\,\text{vertex}}\Big[K\Big[\prod_{k\geq 2}(\lambda^2\mu_k)^{\frac{1}{2}t_0^k(v)}\Big]\Big(\frac{n}{\lambda}\Big)^{2-2g(v)}\Big(\frac{1}{\lambda}\Big)^{\frac{1}{3}d(v)}\Big]. \tag{3.7.8}$$

Therefore, we can single out the diagrams where every index network is homeomorphic to S^2, by picking out only the diagrams whose index function gives the values with the same power of K as that of n^2.

3.7.3. Reduction to manifolds

Combining the results in Sect. 3.7.1 and Sect. 3.7.2, we can reduce the set of possible diagrams to those whose index networks all give triangular decompositions of two-sphere S^2. We then can apply the following theorem to conclude that these diagrams represent tetrahedral decompositions of three-dimensional manifolds:

Theorem. *Assume that the index network around every vertex in diagram γ gives a triangular decomposition of two-sphere. Then, γ represents a tetrahedral decomposition of a three-dimensional manifold.*

Proof. We label the vertices, triangles and hinges of diagram γ as v $(= 1, 2, 3, \ldots)$, f $(= i, j, k, \ldots)$ and h $(= A, B, C, \ldots)$, respectively. Let $\mathcal{T}_v$ denote the index network around vertex v, which we assume to have a form of triangular decomposition of two-sphere. Note that every corner of a triangle in γ corresponds to a segment of the index network around some vertex (see Fig. 3.13). We denote by f_v the segment which is lying on triangle f and

[19] The following prescription to count the number of vertices can be directly applied to any associative algebras $\mathcal{A}_0$.

[20] Note that K equals the linear dimension of $Z(\mathcal{A})$.

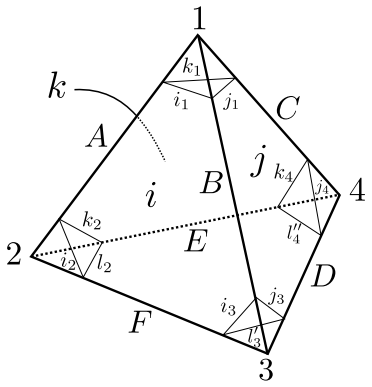


Figure 3.13: A part of diagram γ [1]. The triangle (i_1, j_1, k_1) is a part of the index network around vertex 1, which has a form of triangular decomposition.

is placed in the corner at vertex v.

We choose a vertex (say $v = 1$) and focus on an "index triangle" formed by three segments i_1, j_1, k_1 in $\mathcal{T}_1$. Here, i, j, k are the triangles on which the three segments live. Since all the edges of each triangle are attached to hinges, there are hinges $A = (12)$, $B = (13)$, $C = (14)$, $D = (34)$, $E = (42)$, $F = (23)$ as in Fig. 3.13.[21] As is depicted there, the three index triangles (i_2, k_2, l_2), (i_3, j_3, l_3') and (j_4, k_4, l_4'') ensure the existence of the corresponding triangles l, l' and l'', respectively. We are now going to give a detailed description of these triangles and show that they all coincide, $l = l' = l''$.

We first take a look at hinge $A = (12)$. We assume that $2 \to 1$ is the positive direction of hinge A and label the triangles such that triangle k is to the immediate left of i when seen from vertex 1 (see Fig. 3.13). This means that triangle k is to the immediate right of i when seen from vertex 2, so that i_2 and k_2 are two segments of an index triangle around vertex 2, which will be complemented by the third segment l_2 as in Fig. 3.13. The triangle l on which the segment l_2 lives is glued to triangle i along hinge $F = (23)$, and must be to the immediate left of i when seen from vertex 2 in the direction of F.

We repeat the same argument for hinge $B = (13)$. There, triangle j is to the immediate right of i when seen from vertex 1. This means that triangle j is to the immediate left of i when seen from vertex 3, so that i_3 and j_3 are two segments of an index triangle around vertex 3, which will be complemented by the third segment l_3' as in Fig. 3.13. The triangle l' on which the segment l_3' lives is glued to triangle i along hinge $F = (23)$, and must be to the immediate right of i when seen from vertex 3 in the direction of F. However, this means that l' is to the immediate left of i when seen from vertex 2, and thus two triangles l and l' must be the same.

[21] Note that some of vertices $1, 2, 3, 4$ may represent the same vertex because the index triangles around them may belong to the same connected component of an index network.

The same argument can also be made for hinge $C = (14)$, and we obtain $l = l' = l''$, from which we see that there exists a tetrahedron surrounded by four triangles i, j, k, l. By repeating the same arguments for all the index triangles around every vertex, we conclude that diagram γ gives a tetrahedral decomposition. Furthermore, since the index network around every vertex represents a triangular decomposition of S^2, the neighborhood of every vertex is homeomorphic to B^3. Therefore, the diagram γ gives a tetrahedral decomposition of a three-dimensional manifold. □

3.8. Relation to three-dimensional gravity

We have shown that a class of our models allow us to single out the diagrams which represent tetrahedral decompositions of three-dimensional manifolds. Such models can be used to define discretized three-dimensional Euclidean gravity. In fact, we only need to follow the arguments given in [45, 19, 20].

The action of three-dimensional Euclidean gravity is given by

$$S_0 = -\kappa_0 \int d^3x \sqrt{g}\, R + \Lambda_0 \int d^3x \sqrt{g}, \tag{3.8.1}$$

where κ_0 corresponds to the bare gravitational coupling and Λ_0 to the bare cosmological constant. This can be discretized by using regular tetrahedra with fixed spacing a as

$$S_{\rm EH} = -4\pi\kappa_0 a\, s_0 + \Big[\frac{\sqrt{2}a^3}{12}\Lambda_0 - 4\pi\kappa_0 a\Big(1 - \frac{3\theta}{\pi}\Big)\Big] s_3. \tag{3.8.2}$$

Here, s_0 and s_3 denote the number of vertices and tetrahedra, respectively, and $\theta \equiv \arccos(1/3)$ is the angle between two neighboring triangles in a regular tetrahedron. The free energy of this action is then given by

$$\begin{aligned} \log Z_{\rm EH} &= \sum_{\rm config.} \frac{1}{S}\, e^{-S_{\rm EH}} \\ &= \sum_{\rm config.} \frac{1}{S}\, (e^{4\pi\kappa_0 a})^{s_0} \left(e^{-\frac{\sqrt{2}a^3}{12}\Lambda_0 + 4\pi\kappa_0 a\left(1-\frac{3\theta}{\pi}\right)}\right)^{s_3}, \end{aligned} \tag{3.8.3}$$

where S is the symmetry factor.

In our models, on the other hand, each diagram representing a tetrahedral decomposition contributes to the free energy as

$$\frac{1}{S}\, \lambda^{s_2} \mu^{s_1} N^{s_0}, \tag{3.8.4}$$

Here we have set $\mu_k \equiv \mu$ $(\forall k \geq 2)$. Since the relations $s_2 = 2s_3$ and $s_1 = s_0 + s_3$ hold for tetrahedral decompositions of a three-dimensional manifold, the contribution takes the form

$$\frac{1}{S}\, (\mu N)^{s_0} (\lambda^2 \mu)^{s_3}. \tag{3.8.5}$$

Comparing (3.8.3) and (3.8.5), we obtain the relations between the coupling constants of the two models,

$$\mu N = e^{4\pi\kappa_0 a}, \qquad \lambda^2\mu = e^{-\frac{\sqrt{2}a^3}{12}\Lambda_0 + 4\pi\kappa_0 a\left(1-\frac{3\theta}{\pi}\right)}. \tag{3.8.6}$$

3.9. Duality

We comment that there exists a novel strong-weak duality which interchanges the roles of triangles and hinges when $\mathcal{A}$ is a matrix ring. We expect this duality to play an important role when we further study the analytic properties of the models in the future.

We first recall that one has two choices when introducing a structure of associative algebra to the tensor product of linear spaces, $\mathcal{R} = \mathcal{A} \otimes \bar{\mathcal{A}}$ [see (3.2.4)]. The first is the algebra structure as the tensor product of two associative algebras $\mathcal{A}$ and $\bar{\mathcal{A}}$. This is the structure we have used exclusively so far, and gives the multiplication (3.2.9) (denoted by $\times$), which can also be written as

$$(B_1 \times B_2)^{ij} \equiv (B_1 \times B_2)^i{}_k\, \sigma^{kj} = B_1^{kl} B_2^{mn} y_{km}{}^i y_{nl}{}^j \quad \text{for} \quad B_1, B_2 \in \mathcal{R}. \tag{3.9.1}$$

The second is the algebra structure as the set of endomorphisms of $\mathcal{A}$; $\mathcal{R} = \mathrm{End}\,\mathcal{A}$. The multiplication is defined as the composition of two linear operators acting on $\mathcal{A}$ and will be denoted by dot "$\cdot$":

$$B_1 \cdot B_2 = (B_1 \cdot B_2)^i{}_j e_i \otimes \bar{e}^j \equiv (B_1)^i{}_k (B_2)^k{}_j e_i \otimes \bar{e}^j \quad \text{for} \quad B_1, B_2 \in \mathcal{R}, \tag{3.9.2}$$

which can also be written as

$$(B_1 \cdot B_2)^{ij} \equiv (B_1 \cdot B_2)^i{}_k\, \sigma^{kj} = B_1^{ik} \sigma_{kl} B_2^{lj}. \tag{3.9.3}$$

We will show that there is a duality between the two algebra structures when $\mathcal{A}$ is a matrix ring.

We first set $\sigma_{ij} = g_{ij}$. This is possible because σ can be chosen in an arbitrary way [see a comment following (3.2.19)]. Then, when $\mathcal{A} = M_n(\mathbb{R})$, the multiplications are represented as

$$(B_1 \times B_2)^{abcd} = B_1^{aefd} B_2^{ebcf}, \tag{3.9.4}$$

$$(B_1 \cdot B_2)^{abcd} = B_1^{abef} g_{efgh} B_2^{ghcd} = n\, B_1^{abef} B_2^{fecd}. \tag{3.9.5}$$

We now introduce the dual variables $\tilde{B}$ to B as

$$\tilde{B}^{abcd} \equiv B^{bcda}, \tag{3.9.6}$$

which satisfy the symmetry property $\tilde{B}^{abcd} = \tilde{B}^{cdab}$ due to (3.6.5). Then one can easily show from (3.9.4) and (3.9.5) that the two multiplications are interchanged for the dual variables:

$$(B_1 \times B_2)^{abcd} = \frac{1}{n}(\tilde{B}_2 \cdot \tilde{B}_1)^{bcda}, \qquad (B_1 \cdot B_2)^{abcd} = n(\tilde{B}_1 \times \tilde{B}_2)^{bcda}. \tag{3.9.7}$$

We further introduce the variables $\tilde{A}$ dual to A as

$$\tilde{A}_{abcd} \equiv A_{bcda} \left(= \tilde{A}_{cdab}\right). \tag{3.9.8}$$

Then the action (3.6.4) can be rewritten in terms of the dual variables $\tilde{A}$ and $\tilde{B}$ to the form

$$S = \frac{1}{2}\tilde{A}_{abcd}\tilde{B}^{abcd} - \frac{\lambda}{6n^3}\tilde{A}_{abcd}\tilde{A}_{befc}\tilde{A}_{eadf} - \sum_{k\geq 2}\frac{n^2\mu_k}{2k}\tilde{B}^{a_1b_1b_2a_2}\tilde{B}^{a_2b_2b_3a_3}\dots\tilde{B}^{a_kb_kb_1a_1}. \tag{3.9.9}$$

Note that the way to contract the indices of $\tilde{A}$ (or $\tilde{B}$) in the dual action (3.9.9) is the same as that of B (or A) in the original action (3.6.4). This means that a triangle for the original variables, (3.6.7), now plays the role of a 3-hinge for the dual variables, and a k-hinge for the original variables, (3.6.8), plays the role of a k-gon for the dual variables. We thus find that the action (3.9.9) for the dual variables generates the dual diagrams to the original ones, consisting of 3-hinges (dual to original triangles) and polygons (dual to original hinges).[22] Note that the large N limit in (3.7.6) ($n \to \infty$ with $\lambda^2\mu_k$ and n/λ being fixed) gives $\lambda \to \infty$ and $\mu_k = \mu \to 0$. Since λ and μ are interchanged in the duality transformation, one sees that this duality is actually a strong-weak duality.

3.10. Another example: group ring

In this section, we investigate the models where $\mathcal{A}$ is set to be a group ring $\mathbb{R}[G]$, and demonstrate how the models depend on details of the group structure of G. We assume that G is a finite group with order $|G|$ in order to avoid introducing regularizations, although most of the relations below can be applied to continuous compact groups.

3.10.1. Action for a group ring

Group ring $\mathbb{R}[G]$ is an associative algebra linearly spanned by the elements of G, $\mathbb{R}[G] = \bigoplus_{x\in G}\mathbb{R}e_x$, with multiplication rule determined by that of group G,

$$e_x \times e_y = e_{xy}. \tag{3.10.1}$$

[22] The duality between the two actions will become more symmetric if one allows k-gons to appear in the original action for all $k \geq 2$.

The structure constants ${y_{x,y}}^z$ are then given by

$$ {y_{x,y}}^z = \delta(xy, z). \tag{3.10.2} $$

Here, the contraction of repeated indices is understood to represent the integration with the normalized Haar measure $\int dx \equiv \frac{1}{|G|}\sum_x$:

$$ {y_{x,y}}^z e_z \equiv \int dz\, {y_{x,y}}^z e_z = \frac{1}{|G|}\sum_z {y_{x,y}}^z e_z, \tag{3.10.3} $$

and $\delta(x, y)$ is the delta function with respect to this measure:

$$ \delta(x, y) \equiv |G|\, \delta_{x,y}, \qquad \int dx\, f(x)\, \delta(x, y) = f(y). \tag{3.10.4} $$

From the definition we obtain

$$ y_{x_1,x_2,\ldots,x_k} = \delta(x_1 x_2 \cdots x_k, 1), \tag{3.10.5} $$

$$ g^{x,y} = \delta(xy, 1), \tag{3.10.6} $$

where 1 is the identity of G. Therefore, the action (3.2.19) can be written with the symmetric dynamical variables $A_{x,y} = A_{y,x}$ and $B^{x,y} = B^{y,x}$ as

$$ \begin{aligned} S[A, B] = {} & \frac{1}{2} A_{x,y}\, B^{x,y} - \frac{\lambda}{6} A_{x^{-1},y}\, A_{y^{-1},z}\, A_{z^{-1},x} \\ & - \sum_{k\geq 2} \frac{\mu_k}{2k}\, B^{x_1 y_1} \cdots B^{x_k y_k}\, \delta(x_1 \cdots x_k, 1)\, \delta(y_k \cdots y_1, 1). \end{aligned} \tag{3.10.7} $$

3.10.2. The Feynman rules and the free energy for a group ring

The action (3.10.7) can be rewritten to a form similar to that of matrix ring, by expressing everything in terms of the irreducible representations of G. To show this, we first write the delta function as

$$ \delta(x_1 \cdots x_k, 1) = \sum_R d_R \operatorname{tr}(D_R(x_1) \cdots D_R(x_k)), \tag{3.10.8} $$

where the sum is taken over all the irreducible representations R of G with the representation matrix $D_R(x) = (D^R_{ab}(x))$ $(x \in G)$, and d_R is the dimension of representation R, $d_R = \operatorname{tr} D_R(1)$. Then, the action (3.10.7) can be rewritten to the form

$$ \begin{aligned} S = {} & \frac{1}{2} \sum_{R,S} d_R\, d_S A^{RS}_{abcd} B^{RS}_{abcd} - \frac{\lambda}{6} \sum_{R_1,R_2,R_3} d_{R_1} d_{R_2} d_{R_3} A^{R_1 R_2}_{a_1 b_1 b_2 a_2} A^{R_2 R_3}_{a_2 b_2 b_3 a_3} A^{R_3 R_1}_{a_3 b_3 b_1 a_1} \\ & - \sum_{k\geq 2} \frac{\mu_k}{2k} \sum_{R,S} d_R\, d_S\, B^{RS}_{a_1 a_2 b_2 b_1} \cdots B^{RS}_{a_k a_1 b_1 b_k}, \end{aligned} \tag{3.10.9} $$

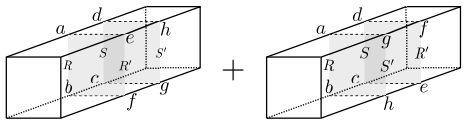
where

$$A^{RS}_{abcd} \equiv \int dxdy\, A_{x,y} D^R_{ab}(x) D^S_{cd}(y) = A^{SR}_{cdab}, \tag{3.10.10}$$

$$B^{RS}_{abcd} \equiv \int dxdy\, B^{x,y} D^R_{ba}(x^{-1}) D^S_{dc}(y^{-1}) = B^{SR}_{cdab}. \tag{3.10.11}$$

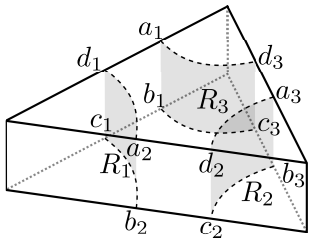
This action gives the following Feynman rules:

propagator :



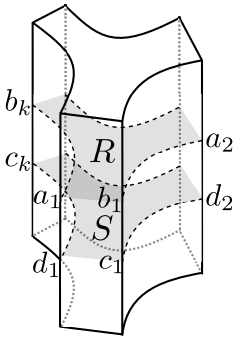
$$\sim \langle A^{RS}_{abcd} B^{R'S'}_{efgh} \rangle = \frac{1}{d_R d_S} (\delta_{ae}\delta_{bf}\delta_{cg}\delta_{dh}\delta^{RR'}\delta^{SS'} + \delta_{ag}\delta_{bh}\delta_{ce}\delta_{df}\delta^{RS'}\delta^{SR'}), \tag{3.10.12}$$

triangle :



$$\sim \lambda\, d_{R_1} d_{R_2} d_{R_3} \delta_{a_1 d_2} \delta_{c_1 b_2} \ldots \delta_{a_k d_1} \delta_{c_k d_1}, \tag{3.10.13}$$

k-hinge :



$$\sim \mu_k\, d_R\, d_S\, \delta_{a_1 b_2} \delta_{d_1 c_2} \ldots \delta_{a_k b_1} \delta_{d_k c_1}. \tag{3.10.14}$$

We thus see that the index network around every vertex is again expressed as a closed surface with double index lines, and its index function is determined only by the Euler characteristics of the polygonal decomposition:[23]

$$\mathcal{F}(\gamma) = \prod_{v:\,\text{vertex}} \mathcal{I}_{g(v)} = \prod_{v:\,\text{vertex}} \Big[\sum_R (d_R)^{2-2g(v)} \Big]. \tag{3.10.15}$$

Here, elementary group theory shows that $\sum_R d_R^2 = |G|$, and $\sum_R d_R^0$ gives the number of irreducible representations which equals that of conjugate classes. For example, when G is the symmetric group S_n, we have

$$\sum_R d_R^2 = |G| = n!\,, \qquad \sum_R d_R^0 = p_n, \tag{3.10.16}$$

[23]This expression can be naturally understood if $\mathcal{F}(\gamma)$ is regarded as the real sector of the index function for the complexified algebra $\mathcal{A}^{\mathbb{C}} = \mathbb{C}[G]$, because the group ring $\mathbb{C}[G]$ can be expressed as the direct sum of $M_{d_R}(\mathbb{C})$ over R, $\mathbb{C}[G] = \bigoplus_R M_{d_R}(\mathbb{C})$.

where p_n denotes the number of partitions of n. Therefore, if G admits the relations $\sum_R d_R^2 \gg \sum_R d_R^{2-2g}$ $(g \geq 1)$, the index networks of spherical topology have a large value of index function compared to those of higher genera.

We also can introduce a color structure as in Sect. 3.7.1 and can control the number of vertices by considering the direct sum of K copies of group ring as in Sect. 3.7.2. Therefore, we can again single out the diagrams which give tetrahedral decompositions of three-dimensional manifolds.

3.11. Orientability

We conclude this chapter by giving a detailed proof of the statement that all the tetrahedral decompositions generated by the action (3.6.4) are orientable, by clarifying the definition of orientation for Feynman diagrams in the model.

We first recall that a thickened triangle has two triangular sides, on each of which directed index lines can be drawn (see Fig. 3.12). Given a tetrahedron T formed by four triangular sides (each coming from a thickened triangle), we embed it to a three-dimensional Euclidean space E^3 as a regular tetrahedron of unit volume. Note that there can be two embeddings f^+ and f^- (up to rotations and translations in E^3), depending on whether the directions of index lines are counterclockwise or clockwise when seen from the center of the embedded tetrahedron (see Fig. 3.14).[24] We say the former embedding to be *positive* and the latter

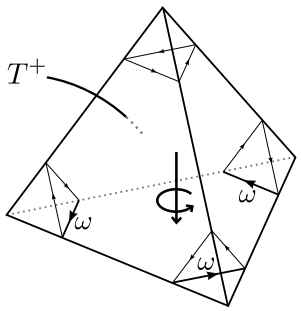


Figure 3.14: A positively oriented tetrahedra T^+, corresponding to a positive embedding of T in E^3 [3].

negative. We then define an *oriented tetrahedron* $T^\pm$ to be the pair of tetrahedron and embedding, $T^\pm \equiv (T, f^\pm)$.

When two positively oriented tetrahedra T_1^+ and T_2^+ are glued at a triangle Δ, we say that the orientation is preserved if the two positive embeddings f_1^+ and f_2^+ can be extended

[24]Note that if the directions of index lines are counterclockwise for one side, they are also counterclockwise for the other three sides because index lines are connected in such a way that the direction is preserved.

(with the use of rotations and translations) to a common embedding f of $T_1^+ \cup T_2^+$ such that the images of two tetrahedra are in opposite positions with respect to the intermediate triangle Δ (see Fig. 3.15). We then say that a tetrahedral decomposition Γ is *orientable* if

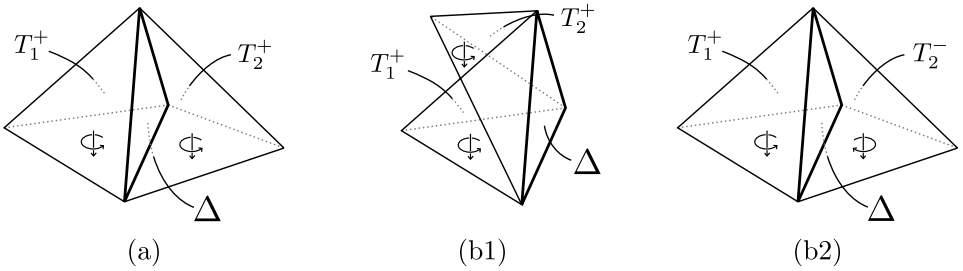


Figure 3.15: Two tetrahedra T_1 and T_2 glued at triangle Δ [3]. (a): both tetrahedra are positively oriented, and the orientation is preserved because they are in opposite positions with respect to Δ. (b1): both are positively oriented, but the orientation is not preserved because they are in the same position with respect to Δ. (b2): they are in opposite positions but are differently oriented.

the orientation is preserved for any two adjacent tetrahedra of positive orientation.

The above orientability condition actually holds for Feynman diagrams discussed in Sect. 3.3. In fact, the index lines on the two sides of a thickened triangle can be drawn in opposite directions as in Fig. 3.12, and thus, for any two adjacent tetrahedra there always exists a natural extension of their positive embeddings such that the images of two tetrahedra are in opposite positions with respect to the triangle. Since it holds for every two adjacent tetrahedra, we conclude that all the tetrahedral decompositions are orientable.

Chapter 4

Extentions of triangle–hinge models

In this chapter, we review our paper [2, 3] where we discuss the extension of triangle–hinge models so that the models correspond to many physics, such as three-dimensional gravity coupling to matter fields. Although these extension should play an important role when we consider the continuum theory, we leave it for future work and we do not consider these extension in Chap. 5 or Chap. 8. The reader interested in these chapters can skip this chapter.

4.1. Matter fields

The prescription to reduce the configurations to tetrahedral decompositions which we discuss in Sect. 3.7 also works when $\mathcal{A}_{\rm grav}$ is extended to a tensor product of the form $\mathcal{A} = \mathcal{A}_{\rm grav} \otimes \mathcal{A}_{\rm mat}$. Here, $\mathcal{A}_{\rm grav}$ is again $M_{n=3m}(\mathbb{R})$, and $\mathcal{A}_{\rm mat}$ is another semisimple associative algebra to be characterizing matter degrees of freedom. In fact, since the structure constants of $\mathcal{A}$ are given by the product of the structure constants of $\mathcal{A}_{\rm grav}$ and those of $\mathcal{A}_{\rm mat}$, the index function $\mathcal{F}(\gamma)$ of each diagram γ is factorized to the product of the contributions from $\mathcal{A}_{\rm grav}$ and $\mathcal{A}_{\rm mat}$ if we set the tensor C to take a factorized form $C = C_{\rm grav} C_{\rm mat}$:

$$\mathcal{F}(\gamma) \equiv \mathcal{F}(\gamma; \mathcal{A}) = \mathcal{F}(\gamma; \mathcal{A}_{\rm grav})\, \mathcal{F}(\gamma; \mathcal{A}_{\rm mat}) \equiv \mathcal{F}_{\rm grav}(\gamma)\, \mathcal{F}_{\rm mat}(\gamma). \tag{4.1.1}$$

Then, by setting $C_{\rm grav}$ to the form (3.7.1) and by taking the limit $n \to \infty$ with $n^2\, \mu_k$ and n/λ being fixed as in Sect. 3.7, the index function $\mathcal{F}(\gamma)$ vanishes unless γ represents a tetrahedral decomposition, and thus we can reduce the set of possible diagrams to tetrahedral decompositions independently of the choice of $\mathcal{A}_{\rm mat}$.[1] In this section, assuming that this reduction is already made, we show that a set of colors (representing matter degrees of

[1] Note that the introduction of matter degrees of freedom may further reduce the set of possible diagrams because $\mathcal{F}_{\rm mat}(\gamma)$ may vanish for a subset of simplicial decompositions.

freedom) can be assigned to simplices of arbitrary dimensions (tetrahedra, triangles, edges and vertices) by choosing $\mathcal{A}_{\rm mat}$ and interaction terms appropriately.

Note that for $\mathcal{A} = \mathcal{A}_{\rm grav} \otimes \mathcal{A}_{\rm mat}$ the dynamical variables take the form $A_{abcd,ij}$ and $B^{abcd,ij}$, where (ab, cd) are matrix indices for $\mathcal{A}_{\rm grav}$ and (i, j) are those for $\mathcal{A}_{\rm mat}$. In this section, we omit the indices $a, b, \ldots$ with respect to $\mathcal{A}_{\rm grav}$ in order to simplify expressions. We will denote the set of colors by $\mathcal{J}$ and the number of elements by $|\mathcal{J}|$.

4.1.1. Coloring tetrahedra

We show that tetrahedra can be colored despite the fact that the action (3.6.4) does not have interaction terms corresponding to tetrahedra. We first set $\mathcal{A}_{\rm mat} = M_{|\mathcal{J}|}(\mathbb{R}) = \bigoplus_{\alpha,\beta\in\mathcal{J}} \mathbb{R}\, e_{\alpha\beta}$ and let the interaction terms take the form[2]

$$- \sum_{\alpha,\beta\in\mathcal{J}} \frac{\lambda_{\alpha\beta}}{6|\mathcal{J}|^3} \sum_{\alpha_1,\ldots,\delta_3\in\mathcal{J}} A_{\alpha_1\beta_1\gamma_1\delta_1} A_{\alpha_2\beta_2\gamma_2\delta_2} A_{\alpha_3\beta_3\gamma_3\delta_3} \times p_\alpha^{\delta_1\alpha_2} p_\alpha^{\delta_2\alpha_3} p_\alpha^{\delta_3\alpha_1} p_\beta^{\beta_3\gamma_2} p_\beta^{\beta_2\gamma_1} p_\beta^{\beta_1\gamma_3}, \tag{4.1.2}$$

where $\lambda_{\alpha\beta} = \lambda_{\beta\alpha}$, and p_α is the projection matrix to the α-th component:

$$p_\alpha^{\alpha_1\alpha_2} = \delta_\alpha^{\alpha_1}\, \delta_\alpha^{\alpha_2}. \tag{4.1.3}$$

The interaction terms can be expressed by thickened triangles as in Fig. 4.1, where the projection matrices p_α and p_β are inserted to the index lines such that each side of the triangle has its own color. Thus, each index triangle at a corner of a tetrahedron gives a factor of the form $\mathrm{tr}(p_{\alpha_1} p_{\alpha_2} p_{\alpha_3})$ if there meet three triangles with colors α_1, α_2, α_3 at the corner (see Fig. 4.2). Since there are four corners in a tetrahedron, the tetrahedron

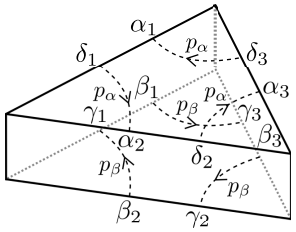


Figure 4.1: A thickened triangle [2]. The upper (lower) side has color α (β).

[2]Recall that we are only looking at the matter part. Actually, the variable A has extra indices of $\mathcal{A}_{\rm grav} = M_{n=3m}(\mathbb{R})$ as $A_{abcd,\alpha\beta\gamma\delta}$, and the interaction terms (4.1.2) have extra factors (3.7.1).

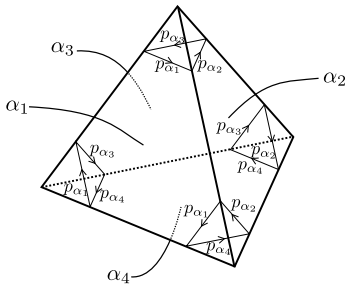


Figure 4.2: Index triangles inside a tetrahedron with triangles colored as in (4.1.2) [2].

illustrated in Fig. 4.2 gives the factor

$$\begin{aligned}&\mathrm{tr}(p_{\alpha_1}p_{\alpha_2}p_{\alpha_3})\,\mathrm{tr}(p_{\alpha_2}p_{\alpha_1}p_{\alpha_4})\,\mathrm{tr}(p_{\alpha_1}p_{\alpha_3}p_{\alpha_4})\,\mathrm{tr}(p_{\alpha_3}p_{\alpha_2}p_{\alpha_4})\\ &\qquad = \begin{cases} 1 & (\alpha_1=\alpha_2=\alpha_3=\alpha_4)\\ 0 & (\text{otherwise})\end{cases}. \end{aligned} \tag{4.1.4}$$

This means that the index function $\mathcal{F}(\gamma)$ can take nonvanishing values only when four index triangles of each tetrahedron have the same color (say, α), which enables us to say that the tetrahedron has a definite color α. We thus succeed in coloring tetrahedra in γ. The parameters $\lambda_{\alpha\beta}$ in (4.1.2) represent the coupling constants of local interactions among matter degrees of freedom on tetrahedra, because $\lambda_{\alpha\beta}$ appears in $\mathcal{F}(\gamma)$ when the corresponding triangle is shared by neighboring tetrahedra of colors α and β.

If we take the set of colors to be $\mathcal{J}=\mathbb{R}^D=\{\mathbf{x}\}$ and let the coupling constants $\lambda_{\mathbf{x},\mathbf{y}}$ ($\mathbf{x},\mathbf{y}\in\mathbb{R}^D$) take nonvanishing values only around $\mathbf{y}$ as a function of $\mathbf{x}$, then $\mathbf{x}$ can be interpreted as the target space coordinates of a tetrahedron in $\mathbb{R}^D$. Since neighboring tetrahedra are locally connected in $\mathbb{R}^D$, the model can describe the dynamics of membranes in $\mathbb{R}^D$. Instead, if we take $\mathcal{J}$ to be a finite set with $|\mathcal{J}|=q$, then the model can describe a q-state spin system on random volumes. In particular, if we consider the case $q=2$ (with colors $\alpha=\pm$), then the model represents three-dimensional quantum gravity coupled to the Ising model. The system is ferromagnetic when $\lambda_{++}\geq\lambda_{+-}$ and $\lambda_{--}\geq\lambda_{+-}$. If the global $\mathbb{Z}_2$ symmetry ($+\leftrightarrow-$) is explicitly broken by setting $\lambda_{++}\neq\lambda_{--}$, then the model describes a system in the presence of an external magnetic field. With generic q, we can construct the q-state Potts models or the RSOS models [46] on random volumes by appropriately choosing $\lambda_{\alpha\beta}$.

4.1.2. Coloring triangles

Triangles can be colored by making an argument similar to the one in Sect, 4.1.1. We set $\mathcal{A}_{\rm mat} = M_s(\mathbb{R}) = \bigoplus_{\alpha,\beta=1}^{s} \mathbb{R}\, e_{\alpha\beta}$,[3] and let the interaction terms take the form

$$-\sum_{\mu\in\mathcal{J}} \frac{\lambda_\mu}{6s^2} A_{\alpha_1\beta_1\gamma_1\delta_1} A_{\alpha_2\beta_2\gamma_2\delta_2} A_{\alpha_3\beta_3\gamma_3\delta_3}\, u_\mu^{\delta_1\alpha_2} u_\mu^{\delta_2\alpha_3} u_\mu^{\delta_3\alpha_1} u_\mu^{\beta_3\gamma_2} u_\mu^{\beta_2\gamma_1} u_\mu^{\beta_1\gamma_3}. \tag{4.1.5}$$

The model with (4.1.5) generates diagrams where a color μ ($\mu \in \mathcal{J}$) is assigned to each triangle. If three triangles (with colors μ, ν, ρ) meet at a corner of a tetrahedron to construct an index triangle, the index function gets the factor $\mathrm{tr}(u_\mu u_\nu u_\rho)$. Then, a tetrahedron illustrated in Fig. 4.3 gives a factor of the form

$$\mathrm{tr}(u_\mu u_\nu u_\rho)\,\mathrm{tr}(u_\nu u_\mu u_\sigma)\,\mathrm{tr}(u_\mu u_\rho u_\sigma)\,\mathrm{tr}(u_\rho u_\nu u_\sigma). \tag{4.1.6}$$

Such factors behave as the coupling constants of local interactions among matter degrees

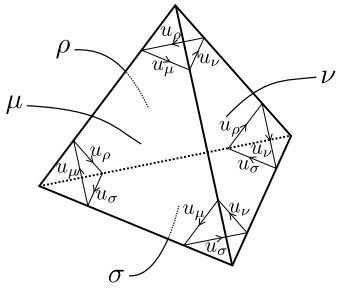


Figure 4.3: Index triangles inside a tetrahedron formed by colored triangles [2].

of freedom located on triangles.

There is another prescription to assign colors to triangles. We introduce $|\mathcal{J}|$ copies of variables A and B [denoted by $A^{(r)}$ and $B^{(r)}$ ($r = 1, \ldots, |\mathcal{J}|$)], and let the action take the form

$$\begin{aligned} S =& \sum_{r\in\mathcal{J}} \frac{1}{2} A^{(r)}_{ij} B^{ij}_{(r)} - \sum_{r\in\mathcal{J}} \frac{\lambda_r}{6} A^{(r)}_{ij} A^{(r)}_{kl} A^{(r)}_{mn}\, g^{jk} g^{lm} g^{ni} \\ & - \sum_{r_1\ldots r_k\in\mathcal{J}} \sum_{k\geq 2} \frac{\mu_k^{r_1\ldots r_k}}{2k}\, B^{i_1 j_1}_{(r_1)} \cdots B^{i_k j_k}_{(r_k)}\, y_{i_1\ldots i_k} y_{j_k\ldots j_1}. \end{aligned} \tag{4.1.7}$$

Then, this model also generates tetrahedral decompositions with colored triangles. The index function $\mathcal{F}(\gamma)$ of the model (4.1.7) gets the factor $\mu_k^{r_1\ldots r_k}$ from a k-hinge shared by k triangles with colors $r_1, \ldots, r_k$, and thus has a form different from (4.1.6). Therefore, although matter degrees of freedom are assigned to triangles in both (4.1.5) and (4.1.7), they give different local interactions (at least apparently).

[3]The linear dimension s can be set to any value as long as the coupling constants (4.1.6) take desired forms.

4.1.3. Coloring edges

There are two prescriptions to assign colors to edges as is the case in coloring triangles.

As in the first prescription in Sect. 4.1.2, we take $\mathcal{A}_{\rm mat} = M_s(\mathbb{R}) = \bigoplus_{\alpha,\beta=1}^{s} \mathbb{R}\, e_{\alpha\beta}$. We now let the interaction terms corresponding to hinges take the form

$$-\sum_{m\in\mathcal{J}}\sum_{k\geq 2}\frac{s^2\mu_k^m}{2k}B^{\alpha_1\beta_1\gamma_1\delta_1}\cdots B^{\alpha_k\beta_k\gamma_k\delta_k}\,u^m_{\beta_1\alpha_2}\ldots u^m_{\beta_k\alpha_1}u^m_{\gamma_1\delta_2}\ldots u^m_{\gamma_k\delta_1}. \tag{4.1.8}$$

This generates diagrams where each edge has a color m ($m \in \mathcal{J}$), and each index triangle gives the factor $\mathrm{tr}(u^{m_1}u^{m_2}u^{m_3})$ depending on the colors of the edges. They give the coupling constants of local interactions among matter degrees of freedom located on edges.

Another prescription to assign colors to edges can be given by modifying the action (3.6.4) to the form

$$\begin{aligned} S =& \sum_{r\in\mathcal{J}}\frac{1}{2}A^{(r)}_{ij}B^{ij}_{(r)} - \sum_{r_1,r_2,r_3\in\mathcal{J}}\frac{\lambda_{r_1r_2r_3}}{6}A^{(r_1)}_{ij}A^{(r_2)}_{kl}A^{(r_3)}_{mn}g^{jk}g^{lm}g^{ni} \\ & -\sum_{r\in\mathcal{J}}\sum_{k\geq 2}\frac{\mu_k}{2k}B^{i_1j_1}_{(r)}\cdots B^{i_kj_k}_{(r)}y_{i_1\ldots i_k}y_{j_k\ldots j_1}. \end{aligned} \tag{4.1.9}$$

Each triangle gives the factor $\lambda_{r_1r_2r_3}$ if three hinges (with colors r_1, r_2, r_3) meet there.

4.1.4. Coloring vertices

Vertices can also be colored despite the fact that the action (3.6.4) does not have interaction terms corresponding to vertices.

We first set the matter associative algebra to be $\mathcal{A}_{\rm mat} = \mathcal{A}^{(1)}_{\rm mat} \oplus \ldots \oplus \mathcal{A}^{(|\mathcal{J}|)}_{\rm mat}$, and let the interaction terms corresponding to hinges take the form

$$-\sum_{\alpha,\beta\in\mathcal{J}}\sum_{k\geq 2}\frac{\mu_k^{\alpha\beta}}{2k}\,B^{i_1j_1}\cdots B^{i_kj_k}y^{(\alpha)}_{i_1\ldots i_k}\,y^{(\beta)}_{j_k\ldots j_1}. \tag{4.1.10}$$

Here $y^{(\alpha)}_{i_1\ldots i_k}$ are the coupling constants constructed from the structure constants $y^{(\alpha)\,k}_{ij}$ of $A^{(\alpha)}_{\rm mat}$ and take nonvanishing values only when all the indices $i_1, \ldots, i_k$ belong to $\mathcal{A}^{(\alpha)}_{\rm mat}$. Accordingly, all the junctions in the same connected index network should have the same color α in order for the index function $\mathcal{F}(\gamma)$ to take nonvanishing values. Thus, we can assign a color to the index network of each vertex in diagram γ, and can say that the model generates diagrams with colored vertices. The matter degrees of freedom located on vertices have local interactions, and two neighboring vertices with colors α and β (connected by a hinge) gives the factor $\mu_k^{\alpha\beta}$ to $\mathcal{F}(\gamma)$.

The above coloring of vertices can also be realized by setting $\mathcal{A}_{\rm mat} = M_{|\mathcal{J}|}(\mathbb{R})$ and letting the interaction terms corresponding to hinges take the form

$$-\sum_{\alpha,\beta\in\mathcal{J}}\sum_{k\geq 2}\frac{|\mathcal{J}|^2\mu_k^{\alpha\beta}}{2k}\,B^{\alpha_1\beta_1\gamma_1\delta_1}\cdots B^{\alpha_k\beta_k\gamma_k\delta_k}p^{\alpha}_{\beta_1\alpha_2}\cdots p^{\alpha}_{\beta_k\alpha_1}p^{\beta}_{\gamma_1\delta_2}\cdots p^{\beta}_{\gamma_k\delta_1}, \tag{4.1.11}$$

where p^α is the projection matrix to the α-th component [the same as the one given in (4.1.3)]. It is easy to see that this model is dual to the model with (4.1.2) through the duality transformations (3.9.8) and (3.9.6). That is, the action with the interaction term (4.1.11) can be regarded as a q-state system on the dual lattice of γ ($q = |\mathcal{J}|$).

We thus conclude that matter degrees of freedom can be introduced to triangle–hinge models such that they live on simplices of any dimensions and interact with themselves locally.

4.1.5. Relations to colored tensor models

We can further construct various kinds of models by combining several prescriptions explained in the previous subsections. For example, we show in this subsection that three-dimensional colored tensor models [24] can be realized as triangle–hinge models by coloring tetrahedra, triangles and edges *at a time*.

4.1.6. Feynman rules of colored tensor models

We first review the Feynman rules of three-dimensional colored tensor models (see, e.g., [24] for a review). The dynamical variables of colored tensor models are given by a pair of rank-three tensors ϕ^μ_{IJK} and $\bar\phi^\mu_{IJK}$ with no symmetry properties under permutations of the subscripts I, J, K. The tensors represent two kinds of colored triangles, where $\{I\}$ is the set of indices assigned to edges, and $\{\mu\} = \{1, 2, 3, 4\}$ are the colors assigned to triangles.[4] The action takes the form

$$S = \sum_{\mu=1}^{4}\phi^\mu_{IJK}\bar\phi^\mu_{IJK} + \kappa\,\phi^1_{IJK}\phi^2_{KML}\phi^3_{MJN}\phi^4_{LNI} + \bar\kappa\,\bar\phi^1_{IJK}\bar\phi^2_{KML}\bar\phi^3_{MJN}\bar\phi^4_{LNI}. \tag{4.1.12}$$

Looking at the way of contraction of indices I, one easily sees that this action generates the Feynman diagrams where the interaction vertices can be identified with tetrahedra which are glued at their faces through the propagator. Since there are two types of interaction terms $\kappa\,\phi^4$ and $\bar\kappa\,\bar\phi^4$, the set of tetrahedra can be decomposed to two different classes, which we label with $\alpha = \pm$, respectively. We assign four different colors $\mu = 1, \ldots, 4$ to four triangles of each tetrahedron. This coloring of triangles naturally introduces the coloring of

[4] In the original Boulatov model [22] the index I runs over the elements of group manifold $SU(2)$.

six edges in a tetrahedron, and we assign color $(\mu\nu) = (\nu\mu)$ to an edge if the edge is shared by two triangles with colors μ and ν ($\mu \neq \nu$). Since the tensors ϕ^{μ}_{IJK} and $\bar{\phi}^{\mu}_{IJK}$ have no permutation symmetry with respect to the subscripts, two tetrahedra can be glued at their faces only when two triangles to be identified have the same color μ and two edges to be identified have the same color $(\mu\nu)$ as in Fig. 4.4. We say that the tetrahedron has positive (or negative) orientation if triangles $1, 2, 3$ are located clockwise (or counterclockwise) when seen from triangle 4 (see Fig. 4.4). Since the kinetic term has the form $\phi\,\bar{\phi}$ (not including ϕ^2 or $\bar{\phi}^2$), two adjacent tetrahedra must have different orientations.

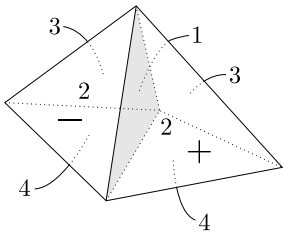


Figure 4.4: A part of a Feynman diagram in colored tensor models [2]. There are two tetrahedra, one corresponding to an interaction vertex proportional to κ and the other to $\bar{\kappa}$. The two adjacent tetrahedra have opposite orientations.

The Feynman rules for colored tensor models (4.1.12) thus can be summarized as follows:

1. Interaction vertices are represented by two types (orientations) of tetrahedra, $\alpha = \pm$, and any two adjacent tetrahedra have different types.

2. Four different colors $\mu = 1, \ldots, 4$ are assigned to four triangles of each tetrahedron, such that the assignment agrees with the orientation of the tetrahedron when $\alpha = +$, while it is opposite when $\alpha = -$.

3. Two tetrahedra are glued at their faces in such a way that two triangles to be identified have the same color μ and two edges to be identified have the same color $(\mu\nu)$.

4.1.7. Realization of colored tensor models as triangle–hinge models

The above Feynman rules for three-dimensional colored tensor models can be reproduced from triangle–hinge models by coloring tetrahedra, triangles and edges at a time. To see this, we set the matter associative algebra $\mathcal{A}_{\mathrm{mat}}$ to be a matrix ring $M_{2s}(\mathbb{R})$ and let the

action take the form

$$
\begin{aligned}
S = &\sum_{(\mu\nu)} \frac{1}{2} A^{(\mu\nu)}_{\alpha\beta\gamma\delta} B^{\alpha\beta\gamma\delta}_{(\mu\nu)} - \frac{\lambda}{6(2s)^3} \sum_{\mu=1}^{4} \frac{1}{6} \sum_{\substack{\nu,\rho,\sigma=1 \\ (\mu\nu\rho\sigma)\text{: all different}}}^{4} A^{(\mu\nu)}_{\alpha_1\beta_1\gamma_1\delta_1} A^{(\mu\rho)}_{\alpha_2\beta_2\gamma_2\delta_2} A^{(\mu\sigma)}_{\alpha_3\beta_3\gamma_3\delta_3} \\
&\times \left(u^{\delta_1\alpha_2}_{+\mu} u^{\delta_2\alpha_3}_{+\mu} u^{\delta_3\alpha_1}_{+\mu} u^{\beta_3\gamma_2}_{-\mu} u^{\beta_2\gamma_1}_{-\mu} u^{\beta_1\gamma_3}_{-\mu} + u^{\delta_1\alpha_2}_{-\mu} u^{\delta_2\alpha_3}_{-\mu} u^{\delta_3\alpha_1}_{-\mu} u^{\beta_3\gamma_2}_{+\mu} u^{\beta_2\gamma_1}_{+\mu} u^{\beta_1\gamma_3}_{+\mu} \right) \\
&- \sum_{k\geq 2} \frac{n^2 \mu_k}{2k} \sum_{(\mu\nu)} B^{\alpha_1\alpha_2\beta_2\beta_1}_{(\mu\nu)} B^{\alpha_2\alpha_3\beta_3\beta_2}_{(\mu\nu)} \cdots B^{\alpha_k\alpha_1\beta_1\beta_k}_{(\mu\nu)}. \qquad (4.1.13)
\end{aligned}
$$

Here, the indices $(\mu\nu) = (\nu\mu)$ $(\mu, \nu = 1, \ldots, 4;\ \mu \neq \nu)$ stand for the colors assigned to edges, the sum $\sum_{(\mu\nu)}$ is taken over all different colors of edges, and we again have neglected the gravity part which ensures the resulting Feynman diagrams to form a set of tetrahedra (see footnote 2). We further assume the matrices $u_{\pm\mu}$ to have the form

$$
u_{+\mu} = \begin{pmatrix} u_\mu & 0 \\ 0 & 0 \end{pmatrix}, \quad u_{-\mu} = \begin{pmatrix} 0 & 0 \\ 0 & (u_\mu)^{\mathrm{T}} \end{pmatrix}. \qquad (4.1.14)
$$

Here $s \times s$ matrices u_μ are chosen such that they satisfy[5]

$$
\mathrm{tr}(u_\mu u_\nu u_\rho)\,\mathrm{tr}(u_\nu u_\mu u_\sigma)\,\mathrm{tr}(u_\mu u_\rho u_\sigma)\,\mathrm{tr}(u_\rho u_\nu u_\sigma) = \begin{cases} 1 & (\epsilon_{\mu\nu\rho\sigma} = +1) \\ 0 & (\text{otherwise}) \end{cases}, \qquad (4.1.15)
$$

where $\epsilon_{\mu\nu\rho\sigma}$ is the totally antisymmetric tensor with $\epsilon_{1234} = 1$. The interaction vertices corresponding to triangles can be expressed by thickened triangles as in Fig. 4.5. Note

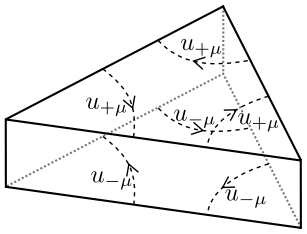


Figure 4.5: A thickened triangle vertex coming from the action (4.1.13), which realizes colored tensor models [2].

[5]For example, one can take the following 6×6 matrices:

$$
u_1 = 2^{-\frac{2}{3}} \begin{pmatrix} 0 & \sigma_1 & 0 \\ 0 & 0 & \sigma_1 \\ 0 & 0 & 0 \end{pmatrix}, \quad u_2 = 2^{-\frac{2}{3}} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i\sigma_2 \\ -i\sigma_2 & 0 & 0 \end{pmatrix},
$$

$$
u_3 = 2^{-\frac{2}{3}} \begin{pmatrix} 0 & \sigma_3 & 0 \\ 0 & 0 & 0 \\ \sigma_3 & 0 & 0 \end{pmatrix}, \quad u_4 = 2^{\frac{1}{3}} \begin{pmatrix} 1 & i\sigma_2 & 0 \\ 0 & 1 & -\sigma_3 \\ -\sigma_1 & 0 & 1 \end{pmatrix},
$$

where σ_i $(i = 1, 2, 3)$ are the Pauli matrices.

that we make colorings for simplices of three different dimensions (tetrahedra, triangles and edges), which are described in Sect. 4.1.1, Sect. 4.1.2 and Sect. 4.1.3, respectively. In fact, each tetrahedron has a type (orientation) $\alpha = \pm$, each triangle has a color $\mu = 1, \ldots, 4$, and each edge has a color $(\mu\nu) = (\nu\mu)$ $(\mu \neq \nu)$. The interaction terms corresponding to triangles indicate that the three edges of a triangle of color μ have different colors $(\mu\nu)$, $(\mu\rho)$, $(\mu\sigma)$. Note that we particularly set $\lambda_{\alpha\beta}$ as $\lambda_{++} = \lambda_{--} = 0$ (and $\lambda_{+-} = \lambda_{-+} = \lambda$), so that any two adjacent tetrahedra have different types. As can be seen from (4.1.6), a tetrahedron of type $\alpha = +$ (or $\alpha = -$) gives the factor

$$\mathrm{tr}\big(u_{\alpha\mu}u_{\alpha\nu}u_{\alpha\rho}\big)\mathrm{tr}\big(u_{\alpha\nu}u_{\alpha\mu}u_{\alpha\sigma}\big)\mathrm{tr}\big(u_{\alpha\mu}u_{\alpha\rho}u_{\alpha\sigma}\big)\mathrm{tr}\big(u_{\alpha\rho}u_{\alpha\nu}u_{\alpha\sigma}\big) \quad (\alpha = \pm), \tag{4.1.16}$$

which takes a nonvanishing value $(= 1)$ only when the four colors μ, ν, ρ, σ are all different and correspond to the positive (or negative) orientation. Thus, a tetrahedron in a nonvanishing Feynman diagram has a positive orientation when its color is $\alpha = +$ and has a negative orientation when $\alpha = -$. Furthermore, if two triangles sharing an edge of color $(\mu\nu)$ $(\mu \neq \nu)$ belong to the same tetrahedron, then one of the two triangles has color μ and the other has color ν. In fact, any triangle connected to a hinge of color $(\mu\nu)$ must have color μ or ν, but two triangles sharing the edge $(\mu\nu)$ must have different colors if they belong to the same tetrahedron.

Therefore, the Feynman diagrams generated by the action (4.1.13) consist of tetrahedra where two adjacent tetrahedra have different orientations $\alpha = \pm$, and four triangles in each tetrahedron have different colors $\mu = 1, \ldots, 4$ such as to be consistent with the orientation of the tetrahedron. Furthermore, the coloring of each edge does not depend on the choice of a tetrahedron including the edge, which leads us to the interpretation that two tetrahedra are glued at their faces such that the edges to be identified have the same color. We thus conclude that the Feynman diagrams obtained from the action (4.1.13) obey the same Feynman rules obtained from the action (4.1.12) of three-dimensional colored tensor models.

4.2. Unoriented theory

In this section, we define unoriented membrane theories in terms of tetrahedral decompositions. A realization of unoriented membrane theories within the framework of triangle–hinge models will be given in the next section.

4.2.1. Matrix models for unoriented strings

As a warm-up before discussing unoriented membrane theories, we review the definition of unoriented string theories and how some of them are realized in terms of real symmetric matrix models.

We first recall that an *oriented* open string is an oriented one-dimensional object with two ends. If we forget about the target-space degrees of freedom, the scattering processes of oriented open strings are represented by Feynman diagrams of Hermitian matrix models:

$$S[M] = \mathrm{tr}\Big(\frac{1}{2}\,M^2 - \frac{\lambda}{3}\,M^3\Big), \tag{4.2.1}$$

where $M = (M_{ij}) = M^\dagger$ is an $N \times N$ Hermitian matrix. In fact, the propagator and the interaction vertex are expressed as[6]

$$\text{propagator}: \quad \begin{matrix} i \text{ -------- } l \\ j \text{ -------- } k \end{matrix} \sim \delta_{il}\delta_{jk} \ \big(= \overbrace{M_{ij}M_{kl}}\big), \tag{4.2.2}$$

$$\text{interaction}: \quad \sim \lambda\, \delta^{jk}\, \delta^{lm}\, \delta^{ni}\,. \tag{4.2.3}$$

Each Feynman diagram can also be thought of as a triangular decomposition of an orientable two-dimensional surface by representing it with the dual diagram.

We now introduce a transformation Ω which acts on one-string states and inverts the worldsheet parity (the orientation of string). *Unoriented* open string theories are then defined as theories where the transformation Ω is gauged (see, e.g., [47]). Namely, we demand that every propagator in the open-string channel be invariant under the action of Ω. This is realized by inserting the projector $(1+\Omega)/2$ to every propagator. If we do not change the form of interaction, the Feynman rules are then expressed as follows (we have rescaled the projector for later convenience):

$$\text{propagator}: \quad \begin{matrix} i \text{ -------- } l \\ j \text{ -------- } k \end{matrix} \ + \ \begin{matrix} i \text{ ---\textbackslash/--- } l \\ j \text{ ---/\textbackslash--- } k \end{matrix} \sim \delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl} \ \big(= \overbrace{X_{ij}X_{kl}}\big), \tag{4.2.4}$$

$$\text{interaction}: \quad \sim \lambda\, \delta^{jk}\, \delta^{lm}\, \delta^{ni}. \tag{4.2.5}$$

It is easy to see that the above Feynman rules are obtained from a real symmetric matrix model:

$$S = \mathrm{tr}\Big(\frac{1}{4}\,X^2 - \frac{\lambda}{6}\,X^3\Big), \tag{4.2.6}$$

[6]Note that in $\mathrm{tr}(M^3) = C^{ijklmn}M_{ij}M_{kl}M_{mn}$, only such components of C^{ijklmn} survive that are totally symmetric under the permutation of three pairs of indices, (ij), (kl), (mn). However, when we write $C^{ijklmn} = \delta^{jk}\delta^{lm}\delta^{ni}$, we intentionally think that C^{ijklmn} are only cyclically symmetric for three pairs of indices, and distinguish two diagrams, one coming from C^{ijklmn} and the other from C^{klijmn}. Of course, by summing two ways of Wick contractions in calculating the free energy, the two diagrams appear in a combined way and C^{ijklmn} will be automatically symmetrized. This "trick" enables us to identify a Feynman diagram with a triangulated surface, and is widely and implicitly adopted in the study of matrix models.

where $X = (X_{ij}) = X^T$ is an $N \times N$ real symmetric matrix.

A Feynman diagram for the above unoriented string theory can also be represented as a collection of triangles glued together along two-hinges. In fact, if we express the vertex C^{ijklmn} (only with a cyclic symmetry) by an oriented triangle, the two contractions in (4.2.4) can be illustrated as in Fig. 4.6. The first contraction leads to a gluing of two oriented

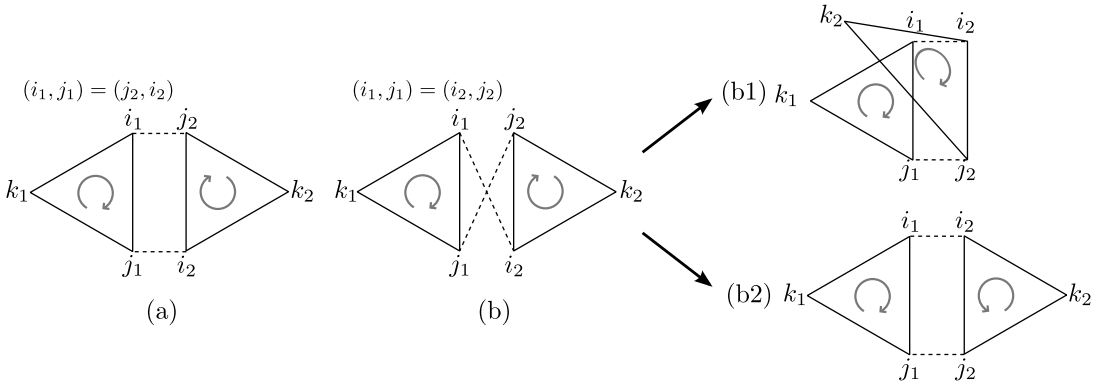


Figure 4.6: Two ways to identify edges of triangles of positive orientation. The orientation is preserved for (a): $(i_1, j_1) = (j_2, i_2)$, but is not for (b): $(i_1, j_1) = (i_2, j_2)$. The local two-dimensional orientations of triangles induce one-dimensional orientations of the edges to be identified [3]. The identification (b) can also be expressed as (b1) or (b2). The expression (b2) is necessarily accompanied by the flip of the right triangle, which means that the local two-dimensional orientation is not preserved when one moves from the left triangle to the right triangle across the identified edge.

triangles with the orientation being preserved, while the second contraction to a gluing for which the orientation is not preserved.

4.2.2. Unoriented membrane theories

We have seen above that unoriented string theories are obtained from oriented theories by gauging the worldsheet parity transformation Ω. We now apply the same prescription to membrane theories in order to define *unoriented membrane theories*; We first prepare oriented models and introduce the worldvolume parity transformation Ω that inverts the orientation of open membrane, and then gauge the transformation Ω by inserting $(1 + \Omega)/2$ to every propagator of open membrane in the original oriented models. In the rest of this section, we assume that worldvolumes in oriented models are already represented as tetrahedral decompositions.

4.2.3. Open membranes of disk topology as fundamental objects

We first argue that the worldvolume dynamics of oriented *closed* membranes of various topologies can also be regarded as that of oriented *open membranes of disk topology*. In fact, tetrahedra in a tetrahedral decomposition can be thought of as interaction vertices that are connected with propagators of membrane of disk topology (i.e. triangles). One thus may say that a worldvolume theory of closed membranes of arbitrary topologies has a dual picture where open membranes of disk topology play fundamental roles, despite the fact that open membranes can have topologies other than disk (such as disks with handles).

4.2.4. Fundamental triplets for oriented membranes

Given an oriented model, we focus on two adjacent, positively oriented tetrahedra T_1^+ and T_2^+ in a tetrahedral decomposition Γ, where T_1^+ and T_2^+ are glued by identifying a triangle Δ_1 in T_1^+ with a triangle Δ_2 in T_2^+ (the resulting identified triangle will be denoted by Δ). Note that the orientation of a tetrahedron naturally induces the positive orientation for four triangles belonging to the tetrahedron, and we represent them by arrows as in Fig. 4.7 (a). We express the identification of edges at Δ as

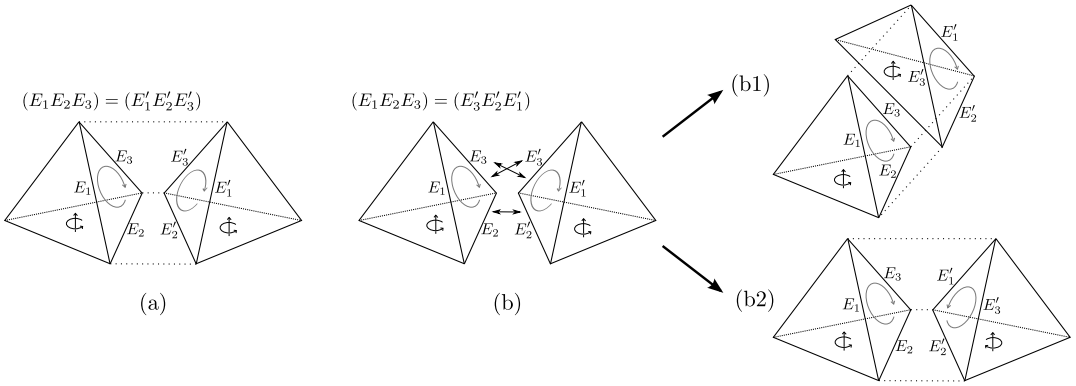


Figure 4.7: Two ways to identify triangles of tetrahedra of positive orientation [3]. The orientation is preserved for (a): $(E_1E_2E_3) = (E_1'E_2'E_3')$, but is not for (b): $(E_1E_2E_3) = (E_3'E_2'E_1')$. The local three-dimensional orientations of tetrahedra induce two-dimensional orientations of the triangles to be identified. The identification (b) can also be expressed as (b1) or (b2). The expression (b2) is necessarily accompanied by the orientation change of the right tetrahedron, which means that the local three-dimensional orientation is not preserved when one moves from the left tetrahedron to the right tetrahedron across the identified triangle.

$$(E_1E_2E_3) = (E_1'E_2'E_3')\,, \tag{4.2.7}$$

where E_1, E_2, E_3 (or E'_1, E'_2, E'_3) are the edges of Δ_1 (or Δ_2). Note that the orientations of Δ_1 and Δ_2 must be opposite in order to form an oriented tetrahedral decomposition, and thus the three-dimensional orientation is preserved when one moves from the inside of T_1 to that of T_2 through the identified triangle Δ [Fig. 4.7 (a)]. Three-dimensional orientation is also preserved for the two other tetrahedral decompositions that are obtained from (4.2.7) by cyclically permuting the edges (E'_1, E'_2, E'_3). We denote by $\Gamma_1 (= \Gamma)$, Γ_2, Γ_3, respectively, the tetrahedral decompositions corresponding to the three edge-identifications that preserve the orientation,

$$\Gamma_1 : (E_1E_2E_3) = (E'_1E'_2E'_3), \quad \Gamma_2 : (E_1E_2E_3) = (E'_2E'_3E'_1), \quad \Gamma_3 : (E_1E_2E_3) = (E'_3E'_1E'_2) . \tag{4.2.8}$$

We will call $(\Gamma_1, \Gamma_2, \Gamma_3)$ the *fundamental triplet associated with triangle* Δ.

4.2.5. Definition of unoriented membrane theories

In addition to the edge-identifications (4.2.8) [leading to the fundamental triplet $(\Gamma_1, \Gamma_2, \Gamma_3)$] , we introduce another triplet $(\tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$ that are obtained, respectively, by the following edge-identifications at the same triangle Δ :

$$\tilde{\Gamma}_1 : (E_1E_2E_3) = (E'_3E'_2E'_1), \quad \tilde{\Gamma}_2 : (E_1E_2E_3) = (E'_1E'_3E'_2), \quad \tilde{\Gamma}_3 : (E_1E_2E_3) = (E'_2E'_1E'_3) . \tag{4.2.9}$$

Note that, in contrast to (4.2.8), three-dimensional orientation is not preserved across Δ [see Fig. 4.7 (b)]. We introduce a transformation Ω that interchanges two triplets $(\Gamma_1, \Gamma_2, \Gamma_3)$ and $(\tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$, and define *unoriented membrane theories* to be those that are obtained from the oriented theories by acting the projection operator $(1+\Omega)/2$ on every triangle. We will call the set $(\Gamma_1, \Gamma_2, \Gamma_3, \tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$ the *fundamental sextet associated with triangle* Δ . So far we have assumed that the tetrahedral decomposition Γ_1 is orientable, but one can easily see that Γ_1 is not necessarily orientable for the above definition of a sextet to make sense because we focus only on local configurations around triangle Δ. In the rest of paper, we understand that the domain of definition for Ω is extended so as to include nonorientable tetrahedral decompositions.

Note that each sextet $(\Gamma_1, \ldots, \tilde{\Gamma}_3)$ consists of both manifolds and nonmanifolds, unlike the two-dimensional cases where Ω always relates a manifold to another manifold. In fact, suppose that a tetrahedral decomposition Γ_1 represents a three-dimensional manifold. Then, the change of the edge-identification at Δ from $(E_1E_2E_3) = (E'_1E'_2E'_3)$ to $(E_1E_2E_3) = (E'_3E'_2E'_1)$ gives rise to a singularity at the midpoint of edge $E_2 = E'_2$ in $\tilde{\Gamma}_1$ around which we cannot define a local orientation. The appearance of singularity will be demonstrated explicitly when we consider an example in Sect. 4.3.3.

4.3. Triangle–hinge models for unoriented membranes

4.3.1. Action and Feynman rules

In this subsection, we realize unoriented membrane theories as triangle–hinge models. We show that they are obtained simply by replacing $C = C_+$ in the original oriented models (3.6.4) with $C = C_+ + C_-$:

$$\begin{aligned} S &= \frac{1}{2}\,[AB] - \frac{\lambda}{6}\left([C_+AAA] + [C_-AAA]\right) - \sum_k \frac{\mu_k}{2k}\,[Y_k \underbrace{B\cdots B}_{k}] \\ &\equiv \frac{1}{2} A_{abcd} B_{abcd} \\ &\quad - \frac{\lambda}{6n^3}\left(\omega^{d_1a_2}\omega^{d_2a_3}\omega^{d_3a_1} + \omega^{d_3a_2}\omega^{d_2a_1}\omega^{d_1a_3}\right)\omega^{b_3c_2}\omega^{b_2c_1}\omega^{b_1c_3}\, A_{a_1b_1c_1d_1} A_{a_2b_2c_2d_2} A_{a_3b_3c_3d_3} \\ &\quad - \sum_k \frac{n^2\mu_k}{2k} B_{a_1a_2b_2b_1}\cdots B_{a_{k-1}a_kb_kb_{k-1}} B_{a_ka_1b_1b_k}\,. \end{aligned} \tag{4.3.1}$$

Here, C_+ is again given by eq. (3.7.1) (where we denote C_+ simply as C), and C_- by

$$C_-^{a_1b_1c_1d_1a_2b_2c_2d_2a_3b_3c_3d_3} \equiv \frac{1}{n^3}\,\omega^{d_3a_2}\omega^{d_2a_1}\omega^{d_1a_3}\omega^{b_3c_2}\omega^{b_2c_1}\omega^{b_1c_3}\,. \tag{4.3.2}$$

We first note that the interaction vertices corresponding to $[C_+AAA]$ and $[C_-AAA]$ can be expressed by thickened triangles with directed index lines as in Fig. 4.8. Contractions

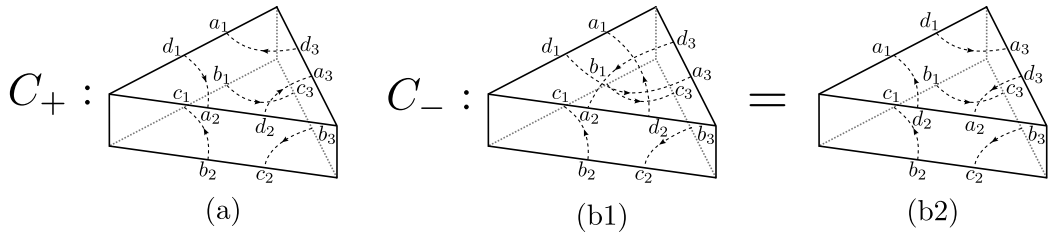


Figure 4.8: Interaction vertices corresponding to (a): $[C_+AAA]$ and (b): $[C_-AAA]$ [3]. If C_+ represents triangle-identifications which preserve three-dimensional local orientation, C_- represents triangle-identifications which do not preserve orientation.

using $[C_+AAA]$ yield the identification of a triangle belonging to a tetrahedron with another triangle belonging to an adjacent tetrahedron so that the orientation is preserved [see Fig. 4.8 (a)]. The two positively oriented tetrahedra thus reside in opposite positions with respect to the thickened triangle, and the edges (a_1, d_1), (a_2, d_2), (a_3, d_3) will be identified with the edges (b_1, c_1), (b_2, c_2), (b_3, c_3), respectively, when we deflate the triangle to get a tetrahedral decomposition. On the contrary, contractions using $[C_-AAA]$ yield an identification of triangles where the orientation is not preserved [see Fig. 4.8 (b1,b2)]. In fact, the indices

of C_- [see (4.3.2)] can be expressed as Fig. 4.8 (b1) or (b2). We use the expression (b1) in a Feynman diagram where the triangle is connected to hinges, but we exploit the other expression (b2) when the thickened triangle is interpreted as representing two triangles to be identified in gluing two tetrahedra of positive orientation. Then, the edges (a_1, d_1), (a_2, d_2), (a_3, d_3) will be identified with the edges (c_1, b_1), (c_2, b_2), (c_3, b_3), respectively, when we deflate triangles to get a tetrahedral decomposition. It is easy to see that the two positively oriented tetrahedra are now in the same position with respect to the triangle and thus will take a configuration of Fig. 3.15 (b1) after the triangle is deflated. This means that the orientation is not preserved for this gluing of tetrahedra.

Note that the direction of arrows on index lines is still preserved for diagrams using C_-. Thus, taking the same large n limit as in the oriented models, we can reduce the set of diagrams such that all their index polygons are triangles,[7] and can conclude that they represent tetrahedral decompositions.

4.3.2. Wick contractions corresponding to the fundamental sextet

Recall that for each triangle Δ in a tetrahedral decomposition $\Gamma = \Gamma_1$, we have the fundamental sextet of tetrahedral decompositions, $(\Gamma_1, \Gamma_2, \Gamma_3, \tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$, which close among themselves under the action of Ω. In this subsection, we write down the corresponding sextet $(\gamma_1, \gamma_2, \gamma_3, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3)$ in unoriented triangle–hinge models.

We first note that, while the *total* number of triangles (as well as that of tetrahedra) is the same among the sextet $(\Gamma_1, \ldots, \tilde{\Gamma}_3)$, this is not the case for those numbers around each edge of triangle Δ. For example, let us consider the case where the three edges of Δ in Γ_1 [denoted by $E_1(= E'_1), E_2(= E'_2), E_3(= E'_3)$] are connected to three different hinges. If we change the identification at Δ from $(E_1E_2E_3) = (E'_1E'_2E'_3)$ to $(E_1E_2E_3) = (E'_2E'_3E'_1)$ to obtain Γ_2, all the three edges E_1, E_2, E_3 must be the same due to triangle-identifications at other triangles.[8] Therefore, Feynman diagrams γ_1 and γ_2 in a triangle–hinge model must have different numbers and different types of hinges if they correspond to Γ_1 and Γ_2, respectively. This means that the constructions of sextets are not so straightforward in triangle–hinge models compared to other models (such as tensor models).

Let us make the above consideration to a more concrete form, considering a triangle Δ in a tetrahedral decomposition Γ_1, at which two positively oriented tetrahedra glued with the orientation being preserved. We first note that there are the following three cases for

[7] As in the original models, a tetrahedron has one index triangle at each corner (see Fig. 4.1.15).

[8] Since we assume that Γ is a tetrahedral decomposition without boundaries, other triangle-identifications in Γ ensure the edge-identifications $E_1 = E'_1$, $E_2 = E'_2$ and $E_3 = E'_3$. See discussions below (4.3.12) for more details.

the three edges I, J, K of triangle Δ:

(1) Three edges I, J, K are connected to three different hinges.

(2) Two and only two of them are connected to the same hinge.

(3) All of them are connected to the same hinge. (4.3.3)

We suppose that Γ_1 is of the type (1) at Δ, and that edges I, J, K are connected to $(p+1)$-, $(q+1)$-, $(r+1)$-hinges, respectively. Including p other edges connected to the $(p+1)$-hinge, we label the edges around the $(p+1)$-hinge as $[I, I_1, \ldots, I_p]$ in a cyclic order. Here, we define the cyclic ordering of edges around a k-hinge as follows (see Fig. 4.9): We first pick

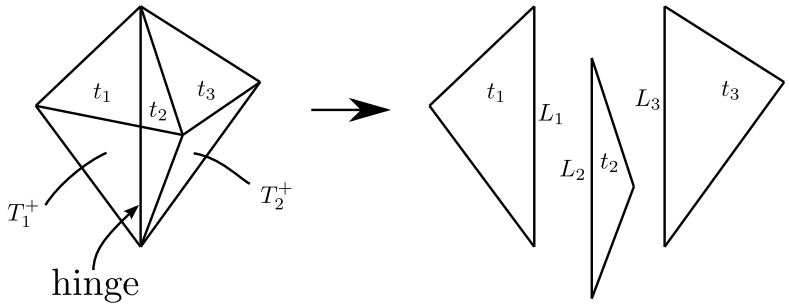


Figure 4.9: Labeling of edges around a k-hinge [3].

up two neighboring triangles t_1 and t_2 belonging to the same tetrahedron T_1^+ of positive orientation, and label their edges connected to the hinge as L_1 and L_2, respectively. Next to T_1^+ there is another positively oriented tetrahedron T_2^+ determined by triangle t_2 and another triangle t_3 sharing the same hinge, and we label as L_3 the edge of t_3 that is connected to the hinge. Repeating this procedure, we obtain a sequence $[L_1, L_2, \ldots, L_k]$ around the k-hinge. Since another choice (t_2, t_3) is possible as the initial pair for the same configuration of edges around the hinge, we should regard the above sequence as being cyclically symmetric, $[L_1, L_2, \ldots, L_k] = [L_2, L_3, \ldots, L_k, L_1]$. Note that, if we take (t_2, t_1) as the initial pair, the edges around the k-hinge will be represented as a sequence in reverse order, $[L_k, \ldots, L_2, L_1]$.

Labeling similarly the edges around the $(q+1)$- and $(r+1)$-hinges by $[J, J_1, \ldots, J_q]$ and $[K, K_1, \ldots, K_r]$, respectively, we have

$$\begin{aligned} \Gamma_1 : \ & \text{edge } I \text{ is connected to a } (p+1)\text{-hinge as } [I, I_1, \ldots, I_p], \\ & \text{edge } J \text{ is connected to a } (q+1)\text{-hinge as } [J, J_1, \ldots, J_q], \\ & \text{edge } K \text{ is connected to a } (r+1)\text{-hinge as } [K, K_1, \ldots, K_r]. \end{aligned} \tag{4.3.4}$$

Then, the remaining tetrahedral decompositions in the sextet have the following configura-

tions:

Γ_2 : three edges I, J, K are connected to a single $(p+q+r+3)$-hinge as $[I, I_1, \ldots, I_p, K, K_1, \ldots, K_r, J, J_1, \ldots, J_q]$. (4.3.5)

Γ_3 : three edges I, J, K are connected to a single $(p+q+r+3)$-hinge as $[I, I_1, \ldots, I_p, J, J_1, \ldots, J_q, K, K_1, \ldots, K_r]$. (4.3.6)

$\tilde{\Gamma}_1$: two edges I, K are connected to a $(p+r+2)$-hinge as $[I, I_1, \ldots, I_p, K, K_1, \ldots, K_r]$, and edge J is connected to a $(q+1)$-hinge as $[J, J_1, \ldots, J_q]$. (4.3.7)

$\tilde{\Gamma}_2$: two edges J, K are connected to a $(q+r+2)$-hinge as $[J, J_1, \ldots, J_q, K, K_1, \ldots, K_r]$, and edge I is connected to a $(p+1)$-hinge as $[I, I_1, \ldots, I_p]$. (4.3.8)

$\tilde{\Gamma}_3$: two edges I, J are connected to a $(p+q+2)$-hinge as $[I, I_1, \ldots, I_p, J, J_1, \ldots, J_q]$, and edge K is connected to a $(r+1)$-hinge as $[K, K_1, \ldots, K_r]$. (4.3.9)

Equations (4.3.4)–(4.3.9) can be understood in the following way. We begin with (4.3.4), which is simplest and obvious. We first split the triangle Δ in Γ_1 to two triangles as in Fig. 4.10 in order to realize the configuration before the edge-identification $(E_1E_2E_3) = (E_1'E_2'E_3')$ is made. This splitting is accompanied by that of edge I to two edges E_1 and

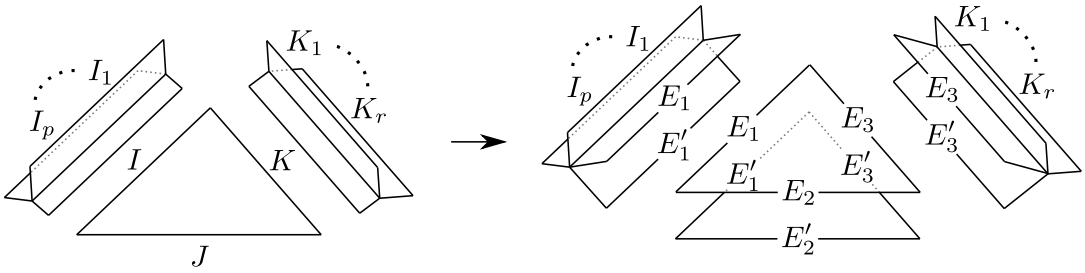


Figure 4.10: The splitting of Δ corresponding to Γ_1 [3]. Edge I becomes two edges E_1 and E_1', and edge J (or K) becomes E_2 and E_2' (or to E_3 and E_3'). The $(p+1)$-, $(q+1)$-, $(r+1)$-hinges accordingly become $(p+2)$-, $(q+2)$-, $(r+2)$-hinges, respectively. Γ_1 is restored by the edge-identification $(E_1E_2E_3) = (E_1'E_2'E_3')$ for the split triangles.

E_1', and that of edge J (or K) to E_2 and E_2' (or to E_3 and E_3'). Accordingly, the $(p+1)$-, $(q+1)$- $(r+1)$-hinges are transformed to $(p+2)$-, $(q+2)$- $(r+2)$-hinges, respectively. Now we follow the sequence of the edges connected to each hinge in the other way around. If we start from the edge E_1 connected to the $(p+2)$-hinge, we then pass through the edges $I_1, \ldots, I_p$ following the original sequence $[I, I_1, \ldots, I_p]$, and reach the edge E_1', which will be identified with the starting edge E_1 (i.e., $E_1' = E_1 = I$) under the edge-identification for Γ_1, $(E_1E_2E_3) = (E_1'E_2'E_3')$. Let us write the total path schematically as a cycle,

$$\bullet\ E_1 \ \to\ I_1 \ \to\ \cdots \ \to\ I_p \ \to\ E_1' = E_1. \tag{4.3.10}$$

Similarly, if we start from the edge E_2 connected to the $(q+2)$-hinge or from the edge E_3 connected to the $(r+2)$-hinge, we then have the following paths:

$$\bullet\ E_2 \to J_1 \to \cdots \to J_q \to E_2' = E_2, \tag{4.3.11}$$

$$\bullet\ E_3 \to K_1 \to \cdots \to K_r \to E_3' = E_3. \tag{4.3.12}$$

Equations (4.3.10)–(4.3.12) are exactly what is expressed in (4.3.4).

Now we consider the tetrahedral decomposition Γ_2, which was obtained from Γ_1 by changing the edge-identification from $(E_1E_2E_3) = (E_1'E_2'E_3')$ to $(E_1E_2E_3) = (E_2'E_3'E_1')$ [see Fig. 4.11 (a)]. If we start from the edge E_1 connected to the $(p+2)$-hinge, we then again

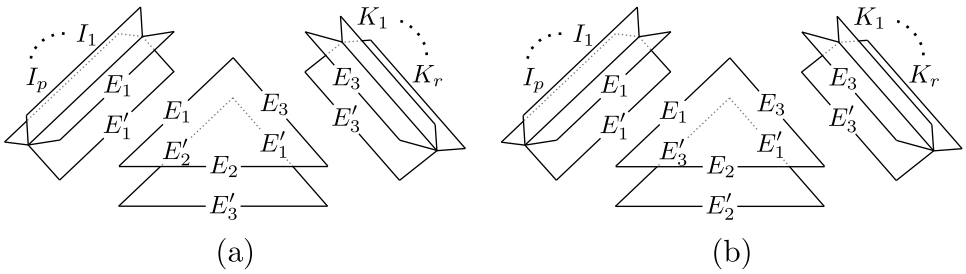


Figure 4.11: The splittings of Δ corresponding to (a): Γ_2 and (b): $\tilde{\Gamma}_1$ [3]. Γ_2 is obtained by the edge-identification $(E_1E_2E_3) = (E_2'E_3'E_1')$ for the split triangles, and $\tilde{\Gamma}_1$ by $(E_1E_2E_3) = (E_3'E_2'E_1')$ for the split triangles.

pass through the edges $I_1, \ldots, I_p$ and reach the edge E_1'. However, this is not the end of journey because E_1' will be identified with E_3 in the edge-identification, and we need to follow another sequence of edges, $K_1, \ldots, K_r$, to reach E_3'. Since E_3' will be identified with E_2, we need to continue the journey; we pass through the edges $J_1, \ldots, J_q$ to reach E_2', which finally will agree with the starting edge E_1. The total path thus can be written as the following cycle:

$$\begin{aligned}\bullet\ E_1 \to I_1 \to \cdots \to I_p \to E_1' = E_3 \to K_1 \to \cdots \to K_r \\ \to E_3' = E_2 \to J_1 \to \cdots \to J_q \to E_2' = E_1.\end{aligned} \tag{4.3.13}$$

Since $E_1' = E_3 = K$, $E_3' = E_2 = J$ and $E_2' = E_1 = I$ under the edge-identification, the path (4.3.13) can be written as (4.3.5). Similarly, (4.3.6) can be understood from the path:

$$\begin{aligned}\bullet\ E_1 \to I_1 \to \cdots \to I_p \to E_1' = E_2 \to J_1 \to \cdots \to J_q \\ \to E_2' = E_3 \to K_1 \to \cdots \to K_r \to E_3' = E_1.\end{aligned} \tag{4.3.14}$$

Equation (4.3.7) can be understood in a similar way, by recalling that $\tilde{\Gamma}_1$ is obtained from Γ_1 by changing the edge-identification from $(E_1E_2E_3) = (E_1'E_2'E_3')$ to $(E_1E_2E_3) = (E_3'E_2'E_1')$

[see Fig. 4.11 (b)], which gives the following two disconnected cycles:

$$\begin{aligned}&\bullet\ E_1 \to I_1 \to \cdots \to I_p \to E'_1 = E_3 \to K_1 \to \cdots \to K_r \to E'_3 = E_1,\\ &\bullet\ E_2 \to J_1 \to \cdots \to J_q \to E'_2 = E_2.\end{aligned} \tag{4.3.15}$$

Similarly, the cycles for $\tilde{\Gamma}_2$ [edge-identification $(E_1E_2E_3) = (E'_1E'_3E'_2)$] are given by

$$\begin{aligned}&\bullet\ E_2 \to J_1 \to \cdots \to J_q \to E'_2 = E_3 \to K_1 \to \cdots \to K_r \to E'_3 = E_2,\\ &\bullet\ E_1 \to I_1 \to \cdots \to I_p \to E'_1 = E_1,\end{aligned} \tag{4.3.16}$$

and those for $\tilde{\Gamma}_3$ [edge-identification $(E_1E_2E_3) = (E'_2E'_1E'_3)$] are given by

$$\begin{aligned}&\bullet\ E_1 \to I_1 \to \cdots \to I_p \to E'_1 = E_2 \to J_1 \to \cdots \to J_q \to E'_2 = E_1,\\ &\bullet\ E_3 \to K_1 \to \cdots \to K_r \to E'_3 = E_3.\end{aligned} \tag{4.3.17}$$

Now that we understand in detail the configurations of the fundamental sextet $(\Gamma_1, \ldots, \tilde{\Gamma}_3)$ associated with triangle Δ, it is easy to translate (4.3.4)–(4.3.9) in terms of unoriented triangle–hinge models, and we obtain the sextet of Feynman diagrams $(\gamma_1, \ldots, \tilde{\gamma}_3)$ as the following groups of Wick contractions:[9]

$$\begin{aligned}\gamma_1:\quad &[C_+A_IA_JA_K]\,X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}\\ &\times [Y_{p+1}B_I^{(+)}B_{I_1}^{\sigma_{I_1}}\cdots B_{I_p}^{\sigma_{I_p}}]\,[Y_{q+1}B_J^{(+)}B_{J_1}^{\sigma_{J_1}}\cdots B_{J_q}^{\sigma_{J_q}}]\,[Y_{r+1}B_K^{(+)}B_{K_1}^{\sigma_{K_1}}\cdots B_{K_r}^{\sigma_{K_r}}]\,,\end{aligned} \tag{4.3.18}$$

$$\begin{aligned}\gamma_2:\quad &[C_+A_IA_JA_K]\,X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}\\ &\times [Y_{p+q+r+3}B_I^{(+)}B_{I_1}^{\sigma_{I_1}}\cdots B_{I_p}^{\sigma_{I_p}}B_K^{(+)}B_{K_1}^{\sigma_{K_1}}\cdots B_{K_r}^{\sigma_{K_r}}B_J^{(+)}B_{J_1}^{\sigma_{J_1}}\cdots B_{J_q}^{\sigma_{J_q}}]\,,\end{aligned} \tag{4.3.19}$$

$$\begin{aligned}\gamma_3:\quad &[C_+A_IA_JA_K]\,X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}\\ &\times [Y_{p+q+r+3}B_I^{(+)}B_{I_1}^{\sigma_{I_1}}\cdots B_{I_p}^{\sigma_{I_p}}B_J^{(+)}B_{J_1}^{\sigma_{J_1}}\cdots B_{J_q}^{\sigma_{J_q}}B_K^{(+)}B_{K_1}^{\sigma_{K_1}}\cdots B_{K_r}^{\sigma_{K_r}}]\,,\end{aligned} \tag{4.3.20}$$

$$\begin{aligned}\tilde{\gamma}_1:\quad &[C_-A_IA_JA_K]\,X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}\\ &\times [Y_{p+r+2}B_I^{(+)}B_{I_1}^{\sigma_{I_1}}\cdots B_{I_p}^{\sigma_{I_p}}B_K^{(+)}B_{K_1}^{\sigma_{K_1}}\cdots B_{K_r}^{\sigma_{K_r}}]\,[Y_{q+1}B_J^{(+)}B_{J_1}^{\sigma_{J_1}}\cdots B_{J_q}^{\sigma_{J_q}}]\,,\end{aligned} \tag{4.3.21}$$

$$\begin{aligned}\tilde{\gamma}_2:\quad &[C_-A_IA_JA_K]\,X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}\\ &\times [Y_{q+r+2}B_J^{(+)}B_{J_1}^{\sigma_{J_1}}\cdots B_{J_q}^{\sigma_{J_q}}B_K^{(+)}B_{K_1}^{\sigma_{K_1}}\cdots B_{K_r}^{\sigma_{K_r}}]\,[Y_{p+1}B_I^{(+)}B_{I_1}^{\sigma_{I_1}}\cdots B_{I_p}^{\sigma_{I_p}}]\,,\end{aligned} \tag{4.3.22}$$

$$\begin{aligned}\tilde{\gamma}_3:\quad &[C_-A_IA_JA_K]\,X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}\\ &\times [Y_{p+q+2}B_I^{(+)}B_{I_1}^{\sigma_{I_1}}\cdots B_{I_p}^{\sigma_{I_p}}B_J^{(+)}B_{J_1}^{\sigma_{J_1}}\cdots B_{J_q}^{\sigma_{J_q}}]\,[Y_{r+1}B_K^{(+)}B_{K_1}^{\sigma_{K_1}}\cdots B_{K_r}^{\sigma_{K_r}}]\,,\end{aligned} \tag{4.3.23}$$

where the superscript σ takes $(+)$ or $(-)$, which represent two types of independent Wick contractions, and we write the first type as $\overbrace{AB}^{(+)}$ or simply as $AB^{(+)}$ and the second type as $\overbrace{AB}^{(-)} = AB^{(-)}$, where we have omitted the indices $i, j, \ldots$.[10] We have written explicitly

[9] Due to the cyclic symmetry of C_+, a group of wick contractions containing $[C_+A_IA_JA_K]$ and that containing $[C_+A_JA_KA_I]$ represent the same Feynman diagram.

[10] Of course, the two pairs appear in a combined way as $\overbrace{AB}^{(+)} + \overbrace{AB}^{(-)}$.

only for the part of the interaction vertices corresponding to Δ (expressed by $[C_{\pm}A_IA_JA_K]$) and the hinges connected to Δ. The remaining part (denoted by $X_{I_1\ldots I_p,J_1\ldots J_q,K_1\ldots K_r}$) is common among the sextet $(\gamma_1,\ldots,\tilde{\gamma}_3)$ and represents the other interaction vertices and their contractions.[11] As for diagrams $\gamma_1,\gamma_2,\gamma_3$, the identification at Δ preserves the orientation as in Fig. 4.7 (a), and thus we have used the vertex $[C_+A_IA_JA_K]$. On the other hand, as for diagrams $\tilde{\gamma}_1,\tilde{\gamma}_2,\tilde{\gamma}_3$, the identification at Δ does not preserve the orientation as in Fig. 4.7 (b), and thus we have used the vertex $[C_-A_IA_JA_K]$.

So far we have assumed that the orientation is preserved at Δ in Γ_1 and also that Γ_1 is of the type (1) in (4.3.3). For other cases, one can also obtain the corresponding sextets $(\gamma_1,\ldots,\tilde{\gamma}_3)$ in a similar way.

4.3.3. Example

To understand the meaning of the above sextet (4.3.18)–(4.3.23), let us consider a simple example. We take a tetrahedral decomposition Γ_1 of a three-sphere, consisting of two tetrahedra glued together at their faces as shown in Fig. 4.12. Diagram γ_1 representing Γ_1 is

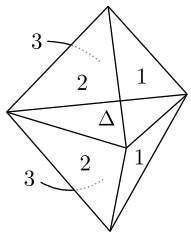


Figure 4.12: Tetrahedral decomposition Γ_1 of three-sphere, where triangles with the same label are identified [3].

realized by the following group of Wick contractions [eq. (4.3.18)]:

$$[C_+A_IA_JA_K]_\Delta[Y_2B_I^{(+)}B_{I_1}^{(-)}][Y_2B_J^{(+)}B_{J_1}^{(-)}][Y_2B_K^{(+)}B_{K_1}^{(-)}]\times X_{I_1J_1K_1}\,, \tag{4.3.24}$$

$$\begin{aligned} X_{I_1J_1K_1} &= [C_+A_{I_1}A_{M_1}A_{N_2}]_1[C_+A_{J_1}A_{N_1}A_{L_2}]_2[C_+A_{K_1}A_{L_1}A_{M_2}]_3 \\ &\quad\times[Y_2B_{L_1}^{(+)}B_{L_2}^{(-)}][Y_2B_{M_1}^{(+)}B_{M_2}^{(-)}][Y_2B_{N_1}^{(+)}B_{N_2}^{(-)}]\,. \end{aligned} \tag{4.3.25}$$

Here, the subscripts $\Delta,1,2,3$ specify the triangles corresponding to the interaction vertices; Δ specifies the triangle at which we change the edge-identification to obtain $\gamma_2,\ldots,\tilde{\gamma}_3$, and $1,2,3$ specify the triangles belonging to the rest part X, which consists of three triangles $(1,2,3)$ and three 2-hinges (see Fig. 4.13). Fig. 4.13 (a1) depicts the part corresponding to X, while Fig. 4.13 (a2) depicts the part corresponding to Δ, which consists of a triangle and three 2-hinges. In Fig. 4.13 (a1) and (a2), edges with the same indices are connected

[11] $X_{I_1\ldots K_r}$ includes A_{I_1}, …, A_{K_r}, which are the partners of B_{I_1}, …, B_{K_r} in the Wick contractions.

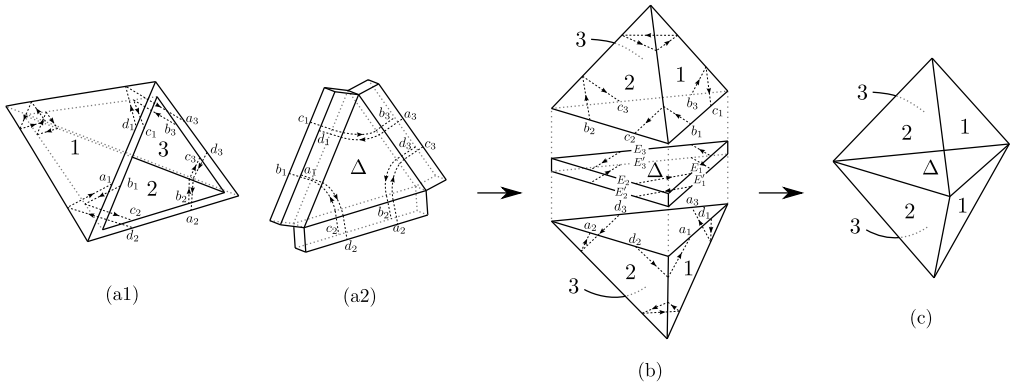


Figure 4.13: Feynman diagram γ_1 representing Γ_1. (a1): the part corresponding to X. (a2): the part corresponding to Δ. The interaction vertices corresponding to triangles are specified by the labels, $\Delta, 1, 2, 3$ [3]. Three 2-hinges exist in (a1) but are not displayed explicitly there. (b): edge-identification at Δ, which is realized by contracting (a1) and (a2). (c): the obtained tetrahedral decomposition Γ_1.

by contractions. Recalling that the edge-identification of Γ_1 is expressed by $(E_1E_2E_3) = (E'_1E'_2E'_3)$, we label the ordered indices (b_i, c_i) and (a_i, d_i) $(i = 1, 2, 3)$ in Fig. 4.13 (a1) as $E_i = E_i(b_i, c_i)$ and $E'_i = E'_i(a_i, d_i)$, respectively. Since edges E_i and E'_i are expressed as in Fig. 4.13 (b) when (a1) and (a2) are combined, we see that the edge-identification $(E_1E_2E_3) = (E'_1E'_2E'_3)$ will certainly be realized after triangle Δ is deflated.

Now we consider diagram γ_2 representing the tetrahedral decomposition Γ_2 that is obtained from Γ_1 by changing the edge-identification from $(E_1E_2E_3) = (E'_1E'_2E'_3)$ to $(E_1E_2E_3) = (E'_2E'_3E'_1)$ (see Fig. 4.14). Note that Γ_2 has the topology of a three-dimensional lens space

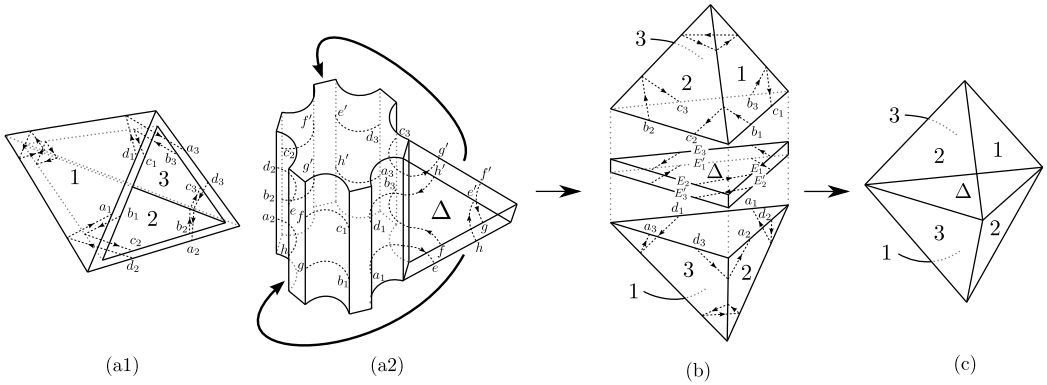


Figure 4.14: Feynman diagram γ_2 representing Γ_2 [3]. (a1): the part corresponding to X. (a2): the part corresponding to Δ. (b): the edge-identification at Δ. (c): the obtained tetrahedral decomposition Γ_2.

$L(3,1)$, as can be seen from Fig. 4.14 (c). Diagram γ_2 is given by (4.3.19), that is,

$$[C_+ A_I A_J A_K][Y_6 B_I^{(+)} B_{I_1}^{(-)} B_K^{(+)} B_{K_1}^{(-)} B_J^{(+)} B_{J_1}^{(-)}] \times X_{I_1 J_1 K_1} \,. \tag{4.3.26}$$

Since the part given by X is common among the sextet, we have the same labeling of edges, $E_i(b_i, c_i)$ and $E'_i(a_i, d_i)$ $(i = 1, 2, 3)$, and diagram γ_2 takes the form shown in Fig. 4.14 (a1) and (a2). In Fig. 4.14 (a2), the ordered indices (b_1, c_1) are connected to (g, f) by index lines, and (a_2, d_2) are connected to (h, e). Thus, as can be seen from Fig. 4.14 (b), edge $E_1(b_1, c_1) = (g, f)$ will be identified with edge $E'_2(a_2, d_2) = (h, e)$ after triangle Δ is deflated. Similarly, edges $E_2(b_2, c_2)$ and $E_3(b_3, c_3)$ will be identified with edges $E'_3(a_3, d_3)$ and $E'_1(a_1, d_1)$, respectively. We thus obtain the edge-identification $(E_1 E_2 E_3) = (E'_2 E'_3 E'_1)$ of Γ_2. In a similar way, we can realize Γ_3 [resulting from the edge-identification $(E_1 E_2 E_3) = (E'_3 E'_1 E'_2)$ at Δ] as a diagram γ_3 [eq. (4.3.20)] of a triangle–hinge model. Γ_3 has the topology of a lens space $L(3,2) = L(3,1)$. Thus, in this simple example, two diagrams γ_2 and γ_3 represent the same tetrahedral decomposition, $\Gamma_2 = \Gamma_3$.

As for the tetrahedral decomposition $\tilde{\Gamma}_1$ [resulting from the edge-identification $(E_1 E_2 E_3) = (E'_3 E'_2 E'_1)$ at Δ], the corresponding diagram $\tilde{\gamma}_1$ is obtained from the following group of Wick contractions [eq. (4.3.21)]:

$$[C_- A_I A_J A_K][Y_4 B_I^{(+)} B_{I_1}^{(-)} B_K^{(+)} B_{K_1}^{(-)}][Y_2 B_J^{(+)} B_{J_1}^{(-)}] \times X_{I_1 J_1 K_1} \,. \tag{4.3.27}$$

The diagram is depicted in Fig. 4.15. In Fig. 4.15 (a2), the ordered indices (b_1, c_1) are

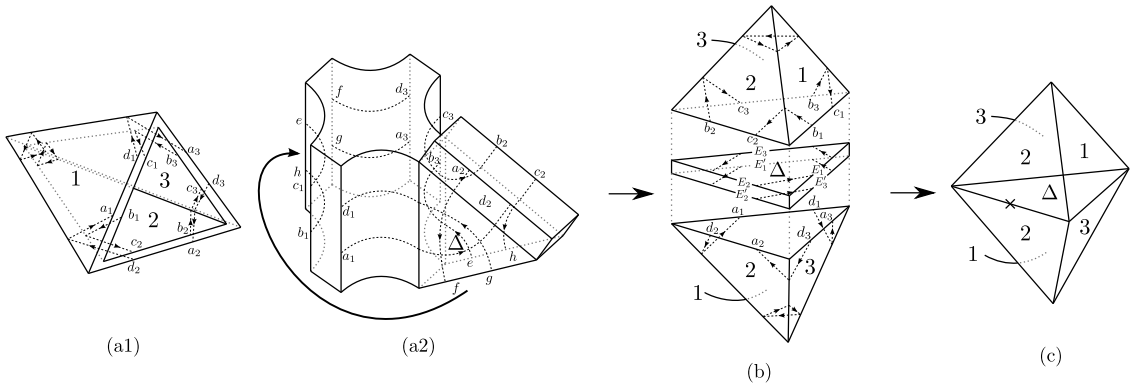


Figure 4.15: Feynman diagram $\tilde{\gamma}_1$ representing $\tilde{\Gamma}_1$ [3]. (a1): the part corresponding to X. (a2): the part corresponding to Δ. (b): the edge-identification at Δ. (c): the obtained tetrahedral decomposition $\tilde{\Gamma}_1$, where the midpoint of the edge shared by triangles 2 and Δ is singular.

connected to (h, e) by index lines, and (d_3, a_3) are connected to (f, g). Thus, due to the edge-identification for C_- explained below (4.3.2), edge $E_1(b_1, c_1) = (h, e)$ will be identified with edge $E'_3(d_3, a_3) = (f, g)$ after triangle Δ is deflated. Similarly, edges $E_2(b_2, c_2)$ and $E_3(b_3, c_3)$ will be identified with edges $E'_2(d_2, a_2)$ and $E'_1(d_1, a_1)$, respectively. We thus obtain

the edge-identification $(E_1E_2E_3) = (E'_3E'_2E'_1)$ of $\tilde{\Gamma}_1$. Although $\tilde{\Gamma}_1$ consists of two tetrahedra, it is not a three-dimensional manifold. In fact, there is a singularity at the midpoint of edge $E_2 = E'_2$ of Δ, around which we cannot define a local orientation. It is easy to see that the other diagrams $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ are realized by (4.3.22) and (4.3.23), respectively, and represent the same tetrahedral decomposition as $\tilde{\Gamma}_1$.

4.3.4. Note on the weights of diagrams

We comment that a sextet $(\gamma_1, \gamma_2, \gamma_3, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3)$ appears in the free energy with the same coefficients. By "the same coefficients" we mean that the numerical factors of these diagrams are the same except for powers of n, λ, μ_k if we treat the common part X as a set of distinguished external vertices and sum over all diagrams representing the same tetrahedral decomposition.

We note that, if a group x of Wick contractions represents a tetrahedral decomposition and if all the interaction vertices are distinguished, then x contributes to the free energy as[12]

$$\frac{1}{s_2!}\Big(\frac{\lambda}{6n}\Big)^{s_2}\prod_{k=1}\Big[\frac{1}{s_1^k!}\Big(\frac{n^2\mu_k}{2k}\Big)^{s_1^k}\Big]. \tag{4.3.28}$$

Here, s_2 and s_1^k denote the numbers of triangles and k-hinges, respectively, in diagram $\gamma = [x]$. Thus, there arise $1/(s_2!\,6^{s_2})$ and $1/(s_1^k!\,(2k)^{s_1^k})$ in the free energy as numerical factors.

If there are n_k internal k-hinges in a diagram,[13] there are $n_k!$ different contractions corresponding to the permutation of these hinges, since the external vertices are distinguished. For each k-hinge, there are $2k$ ways to give the same diagram due to the symmetry of k-hinge vertices. Thus, the numerical factor of each contraction, $1/(n_k!(2k)^{n_k})$, is compensated if we sum these contributions. The numerical factor $1/6$ of triangle Δ is also canceled. Actually, since C_+ has the symmetry (3.2.17) there are six ways to give the same diagram. The above computation ensures that three diagrams $\gamma_1, \gamma_2, \gamma_3$ are generated with unit coefficient in the original triangle–hinge model. Furthermore, since C_- also has the symmetry

$$C_-^{i_1j_1i_2j_2i_3j_3} = C_-^{i_2j_2i_3j_3i_1j_1} = C_-^{j_1i_1j_2i_2j_3i_3}, \tag{4.3.29}$$

$\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3$ are also generated with unit coefficient in an unoriented model.

[12] The n dependence comes from the assumption that all the index polygons are triangles.

[13] Internal hinges mean the parts not in X but connected to Δ.

4.4. Matter fields in unoriented triangle–hinge models

In this section we show that matter fields can be introduced to unoriented triangle–hinge models in the same way as the original triangle–hinge models (3.6.4). We here focus on assigning matter degrees of freedom only to tetrahedra, but the assignment can be done to simplices of any dimensions as in Sect. 4.1.

Introducing matter degrees of freedom is realized by coloring each tetrahedron in tetrahedral decompositions. Actually, we only need to repeat the steps given in Sect. 4.1. We first extend the algebra $\mathcal{A}$ to a tensor product of the form $\mathcal{A} = \mathcal{A}_{\rm grav} \otimes \mathcal{A}_{\rm mat}$. Here, $\mathcal{A}_{\rm grav}$ is again $M_{n=3m}(\mathbb{R})$, and we take $\mathcal{A}_{\rm mat}$ to be $M_{|\mathcal{J}|}(\mathbb{R})$, where $\mathcal{J}$ is the set of colors. Now the dynamical variables A, B have eight indices $A = (A_{ab\alpha\beta,cd\gamma\delta})$, $B = (B_{ab\alpha\beta,cd\gamma\delta})$, where the indices a, b, c, d correspond to $\mathcal{A}_{\rm grav}$, and $\alpha, \beta, \gamma, \delta$ to $\mathcal{A}_{\rm mat}$.[14] Next we set the tensor C to take the form $C = C_+ + C_-$ ($C_\pm$ represents two ways to glue tetrahedra depicted in Fig. 4.8), and assume that each has a factorized form $C_\pm = C_{\pm\rm grav} C_{\pm\rm mat}$. Here, we set $C_{\pm\rm grav}$ to the form (4.3.1), and let $C_{\pm\rm mat}$ take the following form:

$$C_{+\rm mat}^{\alpha_1\beta_1\gamma_1\delta_1\alpha_2\beta_2\gamma_2\delta_2\alpha_3\beta_3\gamma_3\delta_3} = \sum_{\alpha,\beta\in\mathcal{J}} \lambda_{\alpha\beta}\, p_\alpha^{\delta_1\alpha_2} p_\alpha^{\delta_2\alpha_3} p_\alpha^{\delta_3\alpha_1} p_\beta^{\beta_3\gamma_2} p_\beta^{\beta_2\gamma_1} p_\beta^{\beta_1\gamma_3}\,, \tag{4.4.1}$$

$$C_{-\rm mat}^{\alpha_1\beta_1\gamma_1\delta_1\alpha_2\beta_2\gamma_2\delta_2\alpha_3\beta_3\gamma_3\delta_3} = \sum_{\alpha,\beta\in\mathcal{J}} \lambda_{\alpha\beta}\, p_\alpha^{\delta_3\alpha_2} p_\alpha^{\delta_2\alpha_1} p_\alpha^{\delta_1\alpha_3} p_\beta^{\beta_3\gamma_2} p_\beta^{\beta_2\gamma_1} p_\beta^{\beta_1\gamma_3}\,, \tag{4.4.2}$$

where $p_\alpha = (p_\alpha^{\beta\gamma} = \delta_\alpha^\beta \delta_\alpha^\gamma)$ is the projector to the α-th component and $\lambda_{\alpha\beta}$ is a real constant. Equations (4.4.1) and (4.4.2) mean that we insert p_α to each index lines (as ω was inserted for $\mathcal{A}_{\rm grav}$) and take a summation over α and β with weight $\lambda_{\alpha\beta}$. The $\mathcal{A}_{\rm mat}$ part of the interaction vertex $[C_+A^3] + [C_-A^3]$ can be illustrated as in Fig.4.16. Note that p_α is common among

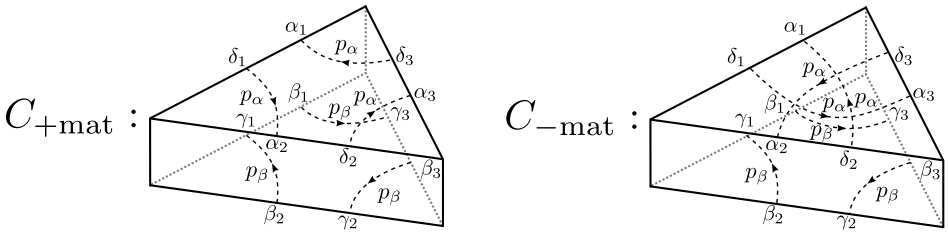


Figure 4.16: Interaction vertices corresponding to triangle ($\mathcal{A}_{\rm mat}$ part) [3]. The upper (lower) side of each triangle has color α (β).

three index lines on each side of a triangle. Thus, one can say that each side of triangle has a color.

In this construction, the index function $\mathcal{F}(\gamma)$ of diagram γ is factorized to the form

$$\mathcal{F}(\gamma) \equiv \mathcal{F}(\gamma;\mathcal{A}) = \mathcal{F}(\gamma;\mathcal{A}_{\rm grav})\mathcal{F}(\gamma;\mathcal{A}_{\rm mat}) \equiv \mathcal{F}_{\rm grav}(\gamma)\mathcal{F}_{\rm mat}(\gamma), \tag{4.4.3}$$

[14] A and B are real-valued matrices symmetric with respect to the pair of indices, $A_{ab\alpha\beta,cd\gamma\delta} = A_{cd\gamma\delta,ab\alpha\beta}$, $B_{ab\alpha\beta,cd\gamma\delta} = B_{cd\gamma\delta,ab\alpha\beta}$.

and the factor $\mathcal{F}_{\rm grav}(\gamma)$ ensures that diagram γ represents a tetrahedral decomposition. The index lines corresponding to $\mathcal{A}_{\rm mat}$ also form index triangles (see Fig.4.17). A tetrahedron

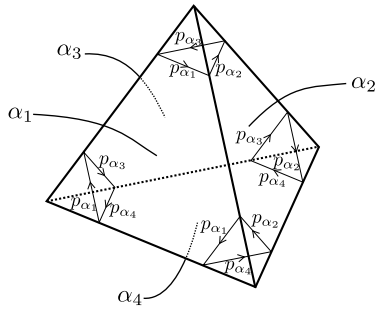


Figure 4.17: Index triangles inside a tetrahedron with triangles colored as in (4.4.1),(4.4.2) [3].

surrounded by four side of triangles with color $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ gives the factor

$$\begin{aligned} &\mathrm{tr}(p_{\alpha_1}p_{\alpha_2}p_{\alpha_3})\mathrm{tr}(p_{\alpha_2}p_{\alpha_1}p_{\alpha_4})\mathrm{tr}(p_{\alpha_1}p_{\alpha_3}p_{\alpha_4})\mathrm{tr}(p_{\alpha_3}p_{\alpha_2}p_{\alpha_4}) \\ &= \begin{cases} 1 & (\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4) \\ 0 & (\text{otherwise}). \end{cases} \end{aligned} \tag{4.4.4}$$

This means that $\mathcal{F}_{\rm mat}(\gamma)$ can take nonvanishing values only when four side of triangles of each tetrahedron have the same color (say α), which enables us to say that the tetrahedron has the color α. We thus succeed in coloring tetrahedra in γ.

If two tetrahedra with color α and β are glued at their faces, the face (corresponding to $C_{\pm{\rm mat}}$) gives the factor $\lambda_{\alpha\beta}$. In this sense the coupling constants $\lambda_{\alpha\beta}$ define a local interaction between the color α and β. If we take the set of colors, $\mathcal{J}$, to be $\mathbb{R}^D = \{\mathbf{x}\}$ and let the coupling constants $\lambda_{\mathbf{x},\mathbf{y}}$ $(\mathbf{x}, \mathbf{y} \in \mathbb{R}^D)$ take nonvanishing values only around $\mathbf{y}$ as a function of $\mathbf{x}$, then $\mathbf{x}$ can be interpreted as the target space coordinates of a tetrahedron in $\mathbb{R}^D$. Since neighboring tetrahedra are locally connected in $\mathbb{R}^D$, the model can describe the dynamics of unoriented membranes embedded in $\mathbb{R}^D$.

Part II

Numerical analysis of triangle–hinge models and the sign problem

Chapter 5

Numerical analysis of triangle–hinge models

In Part II, we make a numerical study of triangle–hinge models. We first make a numerical study of a simplified models, and claim that we need to study the models with the restriction to tetrahedral decompositions as we did in Sect. 3.7, which causes the sign problem. We next review our work [4], where we propose a versatile prescription for the generalized Lefschetz thimble method to solve both the sign problem and the multimodal problem simultaneously. We then perform the Monte Carlo simulation to triangle–hinge models with the restriction.

In this chapter, we perform the Monte Carlo simulation to our models, and show the numerical result. Although we have considered the models generating tetrahedral decompositions, in this chapter we consider the models generating general polygonal decompositions

$$\begin{aligned} S[A,B] =& \frac{1}{2}A_{abcd}B_{abcd} + \sum_{k=1}^{l}\frac{\lambda_k}{2kn^k}A_{a_1b_1b_2a_2}\cdots A_{a_{k-1}b_{k-1}b_ka_k}A_{a_kb_kb_1a_1} \\ &+ \sum_{k=1}^{l}\frac{n^2\mu_k}{2k}B_{a_1a_2b_2b_1}\cdots B_{a_{k-1}a_kb_kb_{k-1}}B_{a_ka_1b_1b_k}. \end{aligned} \tag{5.0.1}$$

We can again restrict the obtained diagrams to polygonal decompositions by inserting matrix ω as in Sect. 3.7. The way to restrict the diagrams to that represents three-dimensional manifold is also completely same for that in Sect. 3.7.

5.1. Redefinition of the models

We consider the partition function with the above action,

$$Z = \int dAdB\, e^{-S[A,B]}. \tag{5.1.1}$$

If the domain of integration is spaces of real symmetric matrices, the integral does not convergent because the action (5.0.1) is not bounded below. This is not suitable to discuss analyticity of the model. We thus redefine the model (5.0.1) by changing the domain of integration. A new domain is given as follows:

$$\begin{cases} A_{abcd} = A_{cdab} = A^*_{badc} = A^*_{dcba} \in \mathbb{C} \\ B_{abcd} = B_{cdab} = B^*_{badc} = B^*_{dcba} \in \mathbb{C}. \end{cases} \tag{5.1.2}$$

This modification does not change the Feynman rules, since the propagator for the new measure is the same as the original one,

$$\langle A_{abcd} B_{efgh} \rangle = \delta_{ae}\delta_{bf}\delta_{cg}\delta_{dh} + \delta_{ag}\delta_{bh}\delta_{ce}\delta_{df} \tag{5.1.3}$$

If the domain of integration is given by (5.1.2), the integral (5.1.1) is finite. it is convenient to change the variables as follows:

$$A_{abcd} = A'_{ij}\tau^i_{ab}\tau^j_{cd}\,, \qquad B_{abcd} = B'_{ij}\sigma^i_{da}\sigma^j_{bc}\,, \tag{5.1.4}$$

where $\{\tau^i\}$ and $\{\sigma^i\}$ $(i = 1, \ldots, n^2)$ are bases of space over $n \times n$ Hermitian matrices. We assume that they are normalized as

$$\mathrm{tr}(\tau^i\tau^j) = \mathrm{tr}(\sigma^i\sigma^j) = \delta^{ij}. \tag{5.1.5}$$

If A and B satisfy (5.1.2), new variables A' and B' are real symmetric matrices ($A'_{ij} = A'_{ji}$, $B'_{ij} = B'_{ji}$). Thus, the dimensions of the domain (5.1.2) is the same as those of space over two real symmetric matrices. Using new variables A' and B', the interaction terms in the action (5.0.1) are written in terms of traces of matrices,

$$A_{a_1b_1b_2a_2}\cdots A_{a_{k-1}b_{k-1}b_ka_k}A_{a_kb_kb_1a_1} = \mathrm{tr}A'^k, \tag{5.1.6}$$

$$B_{a_1a_2b_2b_1}\cdots B_{a_{k-1}a_kb_kb_{k-1}}B_{a_ka_1b_1b_k} = \mathrm{tr}B'^k, \tag{5.1.7}$$

although the quadratic term becomes slightly complicated,

$$A_{abcd}B_{abcd} = A'_{ij}\,\mathrm{tr}(\tau^i\sigma^l\tau^j\sigma^k)\,B'_{kl} \equiv K(A', B') = A'_{ij}K_{ijkl}B'_{kl}\,, \tag{5.1.8}$$

$$K_{ijkl} = \frac{1}{2}\mathrm{tr}(\tau^i\sigma^l\tau^j\sigma^k + \tau^j\sigma^l\tau^i\sigma^k). \tag{5.1.9}$$

The action thus takes the form

$$S[A', B'] = \frac{1}{2}K(A', B') + \sum_{k=1}^{l}\frac{\lambda_k}{2kn^k}\mathrm{tr}A'^k + \sum_{k=1}^{l}\frac{n^2\mu_k}{2k}\mathrm{tr}B'^k. \tag{5.1.10}$$

Since A' and B' are real symmetric matrices, they can be diagonalized using orthogonal matrices S and T as

$$A'_{ij} = S_{ik}\,x_k\,S^{-1}_{kj}, \qquad B'_{ij} = T_{ik}\,y_k\,T^{-1}_{kj}. \tag{5.1.11}$$

The partition function Z can be written as an integral over eigenvalues and orthogonal matrices,

$$Z = \int \prod_i dx_i dy_i \, dS dT \, |\Delta(x)\Delta(y)| \, e^{-S[SxS^{-1}, TyT^{-1}]}, \tag{5.1.12}$$

$$S[SxS^{-1}, TyT^{-1}] = \frac{1}{2} K(SxS^{-1}, TyT^{-1}) + \sum_{i=1}^{n^2} \sum_{k=1}^{l} \left(\frac{\lambda_k}{2kn^k} x_i^k + \frac{n^2 \mu_k}{2k} y_i^k \right), \tag{5.1.13}$$

where $\Delta(x)$ is the Vandermonde determinant. We suppose that the upper bound l in the sum of interaction terms is an even integer ($l \geq 4$), and that λ_l and μ_l are positive.[1] Then, the x-integrals and y-integrals converge due to the potential terms x_i^l and y_i^l. After performing the integration over eigenvalues, Z reduces to an integral over orthogonal matrices S and T. Since the space of orthogonal matrices, $\mathrm{SO}(n^2)$, is compact, Z should be finite.

Here, we give a choice of bases τ^i and σ^i in (5.1.4):

$$\tau^i_{ef} = \tau^{ab}_{ef} = \sqrt{n}^{-1} \sqrt{g}_{abef} \,, \tag{5.1.14}$$

$$\sigma^i_{ef} = \sigma^{ab}_{ef} = \sqrt{n} (\sqrt{g})^{-1}_{abef} = \tau^{ab}_{fe} \,, \tag{5.1.15}$$

where index i is represented by double-index (a, b) $(a, b = 1, \ldots, n)$, and the square root of g is given by

$$\sqrt{g}_{abcd} = \sqrt{n} \big(\frac{1+i}{2} \delta_{ac} \delta_{bd} + \frac{1-i}{2} \delta_{ad} \delta_{bc} \big), \tag{5.1.16}$$

$$(\sqrt{g})^{-1}_{abcd} = \frac{1}{\sqrt{n}} \big(\frac{1-i}{2} \delta_{ac} \delta_{bd} + \frac{1+i}{2} \delta_{ad} \delta_{bc} \big). \tag{5.1.17}$$

In the bases, A' and B' are represented as

$$A'_{abcd} = n \, (\sqrt{g})^{-1}_{abef} A_{efgh} (\sqrt{g})^{-1}_{ghcd} \,, \tag{5.1.18}$$

$$B'_{abcd} = n^{-1} \sqrt{g}_{abef} B_{fghe} \sqrt{g}_{ghcd} \,, \tag{5.1.19}$$

and the kinetic term $K(A', B')$ is given by

$$K(A', B') = \frac{1}{2} A'_{abcd} (B'_{dabc} + 2B'_{acbd} - B'_{adcb}). \tag{5.1.20}$$

We use this expression in the following computations.

[1] Although the upper bound is an odd integer, we can again deform the integration contour so that the integration become finite. This is a similar discussion to that for the integration contour of cubic matrix model.

5.2. Notes on K

In this section, we note the properties of K, which is defined in (5.1.9). We regard K given by (5.1.9) as a linear transformation on the space of $n^2 \times n^2$ real symmetric matrices. (We call the space $V_{\rm sm}$, whose dimension is $N(N+1)/2$, $(N = n^2)$.) The components are

$$K_{IJ} = K_{ijkl} = \frac{1}{2}\mathrm{tr}(\tau^i\sigma^l\tau^j\sigma^k + \tau^j\sigma^l\tau^i\sigma^k), \tag{5.2.1}$$

where I denotes a symmetrized pair (ij). Using the identity,

$$\sum_{i=1}^{N}\tau^i_{ab}\tau^i_{cd} = \sum_{i=1}^{N}\sigma^i_{ab}\sigma^i_{cd} = \delta_{ad}\delta_{bc}, \tag{5.2.2}$$

one can find that K satisfies $KK^T = 1_{V_{\rm sm}}$, that is,

$$K_{ijkl}K_{mnkl} = \frac{1}{2}(\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm}). \tag{5.2.3}$$

K is thus an orthogonal transformation on $V_{\rm sm}$, and the eigenvalues should be 1 or -1.

The number of each eigenvalue depends on the choice of bases τ^i and σ^i. If we set $\sigma^i = \tau^{iT}$ as in (5.1.15), we have

$$K_{ijkl} = \frac{1}{2}\mathrm{tr}(\tau^i\tau^{lT}\tau^j\tau^{kT} + \tau^j\tau^{lT}\tau^i\tau^{kT}). \tag{5.2.4}$$

K then satisfies

$$\begin{aligned}\mathrm{tr}_{V_{\rm sm}}K &= \frac{1}{2}(\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm})K_{ijmn}\\ &= \frac{1}{2}\mathrm{tr}(\tau^i\tau^{jT}\tau^j\tau^{iT} + \tau^j\tau^{jT}\tau^i\tau^{iT})\\ &= n\,,\end{aligned} \tag{5.2.5}$$

where we have used the identity (5.2.2). Therefore, the number of eigenvalue ± 1 is

$$\frac{1}{2}\Big[\frac{1}{2}n^2(n^2+1) \pm n\Big]\,. \tag{5.2.6}$$

If we instead set $\sigma^i = \tau^i$, we have

$$\mathrm{tr}_{V_{\rm sm}}K = \frac{1}{2}\mathrm{tr}(\tau^i\tau^j\tau^j\tau^i + \tau^j\tau^j\tau^i\tau^i) = n^3\,. \tag{5.2.7}$$

Thus, the number of eigenvalue ± 1 in this case is

$$\frac{1}{2}\Big[\frac{1}{2}n^2(n^2+1) \pm n^3\Big]\,. \tag{5.2.8}$$

5.3. Simplified triangle–hinge model

As a first step towards understanding critical behaviors in triangle–hinge models, we study a simplified model, which is given by setting all λ_k and μ_k in (5.1.10) to zero except for λ_4 and μ_4. Although we can only search a narrow parameter region, we expect to find a critical point corresponding to (pure) gravity theory, because the critical behavior of matrix models is determined only by the scaling behavior near the critical point and many potentials give the same critical behavior unless we tune the parameter, and because the first-order (i.e. the lowest) critical point of the matrix models corresponds to two-dimensional pure gravity theory. By rescaling variables, the action is written in the form

$$S[A', B'] = \frac{n^2\beta}{2}\big[K(A', B') + \frac{1}{4}\mathrm{tr}A'^4 + \frac{1}{4}\mathrm{tr}B'^4\big], \tag{5.3.1}$$

where the coupling constant β is given by

$$\beta = (n^2\lambda_4\mu_4)^{-\frac{1}{2}}. \tag{5.3.2}$$

We first see the asymptotic behaviors of the free energy of the simplified model,

$$F \equiv -\frac{1}{n^4}\log Z = -\frac{1}{n^4}\log\Big[\int dA'dB'e^{-S[A',B']}\Big], \tag{5.3.3}$$

in the regions $\beta \ll 1$ and $\beta \gg 1$, and in the next section we argue that there is a critical point in the model.

5.3.1. Free energy at $\beta \ll 1$

If we rescale the variables as

$$A' \to \beta^{-\frac{1}{4}}A',\ B' \to \beta^{-\frac{1}{4}}B', \tag{5.3.4}$$

the partition function becomes

$$Z = \text{const.} \times \beta^{-\frac{n^2(n^2+1)}{4}} \int dA'dB' \exp\Big[-\frac{n^2}{2}\big\{\sqrt{\beta}K(A', B') + \frac{1}{4}\mathrm{tr}A'^4 + \frac{1}{4}\mathrm{tr}B'^4\big\}\Big]. \tag{5.3.5}$$

We expand it perturbatively with respect to $\beta \ll 1$, and obtain

$$Z = \text{const.} \times \beta^{-\frac{n^2(n^2+1)}{4}}\big(1 + n^4\sum_k a_k\beta^k\big), \tag{5.3.6}$$

$$a_k = \int dA'dB'e^{-\frac{n^2}{8}\left(\mathrm{tr}A'^4+\mathrm{tr}B'^4\right)}\big[K(A', B')^2\big]^k. \tag{5.3.7}$$

The coefficients a_k are written as integrals over eigenvalues. For example, a_1 is given by

$$\begin{aligned} a_1 =& \int \prod_{i=1}^{N=n^2} (dx_i dy_i) |\Delta(x)\Delta(y)| e^{-\frac{n^2}{8}\sum_i (x_i^4+y_i^4)} x_{p_1} x_{p_2} y_{q_1} y_{q_2} \\ & \times K_{i_1 j_1 k_1 l_1} K_{i_2 j_2 k_2 l_2} I_{i_1 p_1 j_1 p_1, i_2 p_2 j_2 p_2} I_{k_1 q_1 l_1 q_1, k_2 q_2 l_2 q_2}, \end{aligned} \tag{5.3.8}$$

$$\begin{aligned} I_{i_1 j_1 i_2 j_2, i_3 j_3 i_4 j_4} =& \int_{\mathrm{SO}(N)} (dS) S_{i_1 j_1} S_{i_2 j_2} S_{i_3 j_3} S_{i_4 j_4} \\ =& \delta_{j_1 j_2}\delta_{j_3 j_4}(\xi \delta_{i_1 i_2}\delta_{i_3 i_4} - \eta \delta_{i_1 i_3}\delta_{i_2 i_4} - \eta \delta_{i_1 i_4}\delta_{i_2 i_3}) \\ & + \delta_{j_1 j_3}\delta_{j_2 j_4}(-\eta \delta_{i_1 i_2}\delta_{i_3 i_4} + \xi \delta_{i_1 i_3}\delta_{i_2 i_4} - \eta \delta_{i_1 i_4}\delta_{i_2 i_3}) \\ & + \delta_{j_1 j_4}\delta_{j_2 j_3}(-\eta \delta_{i_1 i_2}\delta_{i_3 i_4} - \eta \delta_{i_1 i_3}\delta_{i_2 i_4} - \xi \delta_{i_1 i_4}\delta_{i_2 i_3}), \end{aligned} \tag{5.3.9}$$

where $N = n^2$, $\xi = \frac{N+1}{N(N-1)(N+2)}$ and $\eta = \frac{1}{N(N-1)N+2)}$.

5.3.2. Free energy at $\beta \gg 1$

We show that the free energy of the simplified model behaves as follows in the region $\beta \gg 1$:

$$F \sim -\frac{1}{4}\beta + c_n \log \beta + \text{const.}, \tag{5.3.10}$$

where c_n is a constant depending on n.

In order to derive (5.3.10), we first decompose symmetric matrices into eigenvalues and orthogonal matrices as (5.1.11). The partition function is given by

$$Z = \int \prod_i dx_i dy_i \, dS dT \, |\Delta(x)\Delta(y)| \, e^{-S[SxS^{-1}, TyT^{-1}]}, \tag{5.3.11}$$

$$S[SxS^{-1}, TyT^{-1}] = \frac{n^2\beta}{2}\big[K(SxS^{-1}, TyT^{-1}) + \frac{1}{4}\sum_i (x_i^4 + y_i^4)\big], \tag{5.3.12}$$

Since there is a factor β in the action (5.3.12), it is expected that the configuration which gives the local minimum of the action dominates in the region $\beta \gg 1$. In the bases (5.1.18) and (5.1.19), if we set orthogonal matrices $S = T = 1$, we have conditions for eigenvalues,

$$0 = \frac{\partial S}{\partial x_{ab}} = \begin{cases} \frac{n^2\beta}{2}(y_{aa} + x_{aa}^3) \ (a = b) \\ \frac{n^2\beta}{2}\big(-\frac{1}{2}(y_{ab} - y_{ba}) + x_{ab}^3\big) \ (a \neq b) \end{cases}, \tag{5.3.13}$$

$$0 = \frac{\partial S}{\partial y_{ab}} = \begin{cases} \frac{n^2\beta}{2}(x_{aa} + y_{aa}^3) \ (a = b) \\ \frac{n^2\beta}{2}\big(-\frac{1}{2}(x_{ab} - x_{ba}) + y_{ab}^3\big) \ (a \neq b) \end{cases}. \tag{5.3.14}$$

These are satisfied by the following $\bar{x}$ and $\bar{y}$:

$$\begin{cases} \bar{x}_{aa} = -\bar{y}_{aa} = \pm 1 \\ \bar{x}_{ab} = -\bar{x}_{ba} = \bar{y}_{ab} = -\bar{y}_{ba} = \pm 1 \ (a \neq b) \end{cases}. \tag{5.3.15}$$

The value of the action at this point is $S_0 = -\beta n^4/4$. Expanding x and y around the configurations as,

$$x_{ab} = \bar{x}_{ab} + \frac{1}{\sqrt{\beta}}\hat{x}_{ab}, \tag{5.3.16}$$

$$y_{ab} = \bar{y}_{ab} + \frac{1}{\sqrt{\beta}}\hat{y}_{ab}, \tag{5.3.17}$$

one can show that the quadratic part of the action with respect to $\hat{x}$ and $\hat{y}$ is positive definite. Thus, (5.3.15) gives the local minimum if we fix $S = T = 1$.

We next consider directions of orthogonal matrices. If we represent S and T as

$$S = e^{u/\sqrt{\beta}}, \quad T = e^{v/\sqrt{\beta}}, \tag{5.3.18}$$

where $u^T = -u$ and $v^T = -v$, the expansion of K with respect to u and v at $x = \bar{x}, y = \bar{y}$ is given by

$$\begin{aligned}\beta K(S\bar{x}S^{-1}, T\bar{y}T^{-1}) = \beta K(\bar{x},\bar{y}) &+ \sum_{a,b,c,d,e,f,g,h} K_{abcd,efgh}(\bar{x}_{ab} - \bar{x}_{cd})u_{abcd}(\bar{y}_{ef} - \bar{y}_{gh})v_{efgh} \\ &+ \sum_{a,b,c,d}(\bar{x}_{ab} - \bar{x}_{cd})^2 u_{abcd}^2 + \sum_{a,b,c,d}(\bar{y}_{ab} - \bar{y}_{cd})^2 v_{abcd}^2 + \mathcal{O}\big(\beta^{-1/2}\big). \end{aligned} \tag{5.3.19}$$

We can show that the quadratic term with respect to u and v is non-negative,[2] by using the fact that eigenvalues of K as a linear transformation on the space of symmetric matrices are ± 1 (see Sect. 5.2). Thus, $S = T = 1$ gives a local minimum.

Expanding around the above minimum, we have

$$\begin{aligned} Z =& \text{const. } \beta^{\text{const.}} \int \prod_{ab} d\hat{x}_{ab} d\hat{y}_{ab} \prod_{(ab)<(cd)} du_{abcd} dv_{abcd} \\ &\times |\Delta(x)\Delta(y)| \exp\Big[\frac{n^4}{4}\beta - \delta V(\hat{x},\hat{y},u,v)\Big], \end{aligned} \tag{5.3.20}$$

where $\delta V(\hat{x},\hat{y},\sigma,\tau)$ represents a term depending on $\hat{x},\hat{y},u,v$, The leading term of δV is quadratic with $\mathcal{O}(\beta^0)$ coefficient in the $\beta \gg 1$. Thus, we can find the behavior (5.3.10), although it is hard to determine the coefficient c_n which depends on the nonlinear structure of δV due to zero-eigenvalues of the Hesse matrix.

5.4. Numerical analysis

In this section, we argue that there is a critical point in the model, and a third-order phase transition occurs there. We expect that there is a phase transition in the large N simplified

[2]There are zero-eigenvalues in the Hesse matrix around the point. Some of them correspond to flat directions, namely symmetry of K. However, the other directions are not true symmetry, and lift up at higher order.

triangle–hinge model from the following observation. Due to the expression (5.3.5) for small $\beta \ll 1$, the potential terms $\mathrm{tr}(A'^4 + B'^4)$ dominate compared to $\sqrt{\beta}K(A', B')$. Since the potential has only one minimum at the origin, eigenvalues of A' are distributed continuously in a connected range around the origin, (and those of B' also). On the other hand, for $\beta \gg 1$, eigenvalues should be distributed around the saddle point (5.3.15) of the action (5.3.12). This distribution may have two disconnected supports: one is around 1 and the other is around -1. Therefore, there seems to be a transition between the 1-cut distribution at small β and the 2-cut distribution at large β.

In order to check the above observation, we will compute numerically the following k-point functions I_k corresponding to derivatives of the free energy:

$$I_k(\beta) = \beta^k \frac{d^k F}{d\beta^k} = \big(-\frac{n^4}{2}\big)^{k-1} \beta^k \langle O^k \rangle_c, \tag{5.4.1}$$

$$O = \frac{1}{n^2}\big[K(A', B') + \frac{1}{4}\mathrm{tr}A'^4 + \frac{1}{4}\mathrm{tr}B'^4\big]. \tag{5.4.2}$$

From the behavior (5.3.10) at $\beta \gg 1$, I_k ($k \geq 2$) should be an $\mathcal{O}(1)$ constant at $\beta \gg \beta_c$ for arbitrary n. This may make it clear to see numerical results and to find a critical point β_c.

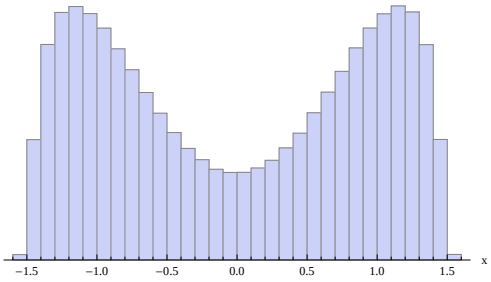
We perform the Monte Carlo simulation to compute k-point functions I_k defined by (5.4.1) in the simplified triangle–hinge model. We used matrices A, B as dynamical variables, and applied Metropolis-Hasting algorithm with the standard deviation $\sigma = 0.05\beta^{-1/4}$ for the transition. We consider three types of $n = 4, 6, 8$ (thus the size of matrix is $16, 36, 64$ respectively), and guess the large-n behavior from them. We calculate β dependence of $\langle O \rangle$, where O is defined in (5.4.2). We start the Monte Carlo calculation from initial values

$$A'_{abcd} = \alpha_{ab}\delta_{ac}\delta_{bd}, B'_{abcd} = \beta_{ab}\delta_{ac}\delta_{bd}, \tag{5.4.3}$$

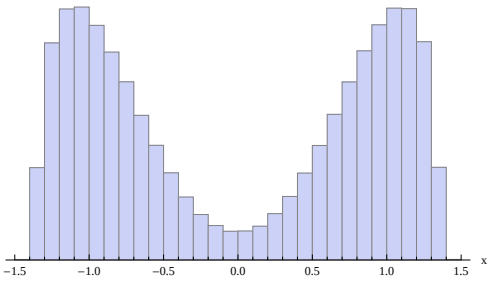
$$\alpha_{aa} = -\beta_{aa} = \begin{cases} +1 \ (a \leq \frac{n}{2}) \\ -1 \ (a > \frac{n}{2}) \end{cases}, \tag{5.4.4}$$

$$\alpha_{ab} = \beta_{ab} = \begin{cases} +1 \ (a > b) \\ -1 \ (a < b) \end{cases}. \tag{5.4.5}$$

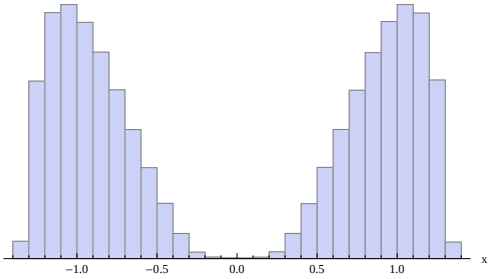
We show the distributions of eigenvalues for various β at $n = 8 (N = 64)$ in Fig. 5.1. For small β, the distribution has a single connected support. For large β, the distribution has two supports as we expected. They indicate that there is a phase transition between the two behaviors. Taking the finite n effect into account, we expect that the transition occurs around $\beta \sim 6$ in the large n limit. We show the 1-point functions in the triangle–hinge model in Fig. 5.2. They seem to be smooth. Thus, the expected transition seems to be more than the second-order. We show the 2-point functions in the triangle–hinge model in Fig. 5.3. It seems that the behavior changes around $\beta \sim 6$. We thus consider that there is a third-order phase transition around $\beta \sim 6$ in the simplified triangle–hinge model.



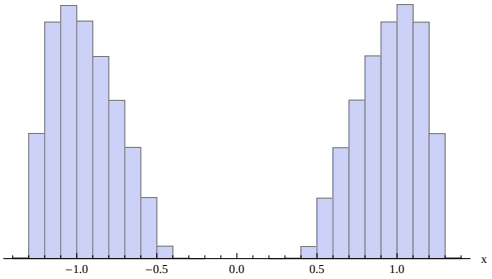
(a) $n = 8$, $\beta = 2.0$



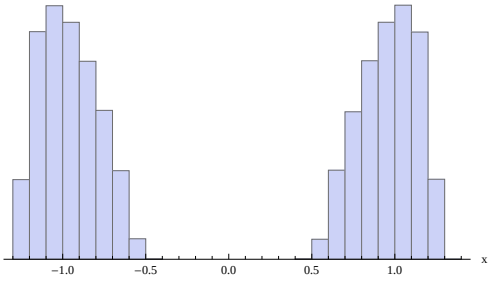
(b) $n = 8$, $\beta = 4.0$



(c) $n = 8$, $\beta = 6.0$



(d) $n = 8$, $\beta = 8.0$



(e) $n = 8$, $\beta = 10.0$

Figure 5.1: Distributions of eigenvalues in the triangle–hinge model (5.3.1).

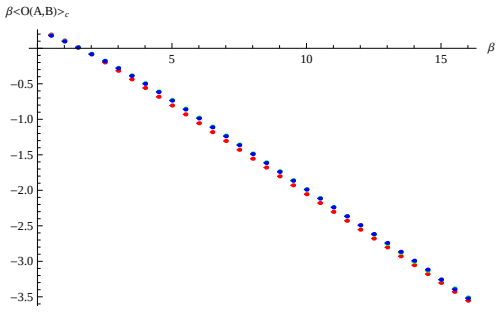


Figure 5.2: 1-point functions for the triangle–hinge model (5.3.1). The red, green and blue dots are the results for $n = 4, 6$ and 8 respectively.

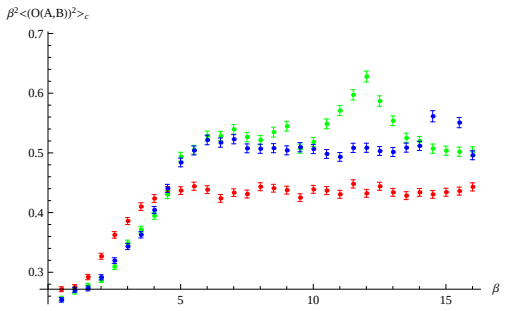


Figure 5.3: 2-point functions in the triangle–hinge model (5.3.1). The red, green and blue dots are the results for $n = 4, 6$ and 8 respectively.

5.5. Interpretation

We saw in the previous section that the triangle–hinge models have a third-order phase transition in the large-n limit. However, this phase transition may not correspond to three-dimensional gravity theory: The phase transition is quite similar to that of a class of one-matrix models such as

$$S(M) = N\beta \mathrm{tr}\left(-\frac{1}{2}M^2 + \frac{1}{4}M^4\right), \tag{5.5.1}$$

where M is $N \times N$ Hermitian matrix. As we have seen in Sect. 2.1.2, the resolvent for the one-cut solution is given as

$$w(z) = \frac{\beta}{2}\left(-z + z^3 - \sqrt{(z^2 + \frac{a^2}{2} - 1)^2(z^2 - a^2)}\right) \tag{5.5.2}$$

$$a^2 \equiv \frac{2}{3}\left(1 - \sqrt{1 + \frac{\beta}{12}}\right). \tag{5.5.3}$$

On the other hands, this model has another class of solution (two-cut solution). Let us consider the case that the eigenvalues distribute in two regions $[-a, b]$ and $[b, a]$ $(0 < b < a)$. The resolvent is then written as

$$w(z) = \frac{\beta}{2}\left(-z + z^3 - |z|\sqrt{(z^2 - c^2)(z^2 - d^2)}\right). \tag{5.5.4}$$

Here, the constant c and d can again be determined by (2.1.31), (2.1.31) and the asymptotics $w(z) \to 1/z (z \to \infty)$, and we obtain

$$w(z) = \frac{\beta}{2}\left(-z + z^3 - \sqrt{z^2(z^2 - 1 + \frac{2}{\sqrt{\beta}})(z^2 - 1 - \frac{2}{\sqrt{\beta}})}\right). \tag{5.5.5}$$

We can easily calculate the second derivative of the free energy from the resolvent as

$$\begin{aligned}\frac{\partial F}{\partial \beta} &= \frac{1}{N^2}\left[\langle\left(\frac{\partial S}{\partial \beta}\right)^2\rangle - \langle\frac{\partial S}{\partial \beta}\rangle^2\right] \\ &= \begin{cases} -\frac{1}{216\beta^2}\left(54+\beta^2+(\beta-6)\sqrt{\beta(\beta+12)}\right) & (\beta \le 4) \\ -\frac{1}{4\beta^2} & (\beta > 4) \end{cases}, \end{aligned} \tag{5.5.6}$$

where the one-cut solution appears for $\beta \le 4$ and the two-cut solution appears for $\beta > 4$. The third derivative is discontinuous at $\beta = 4$, and this model thus has third-order phase transition at $\beta = 4$. Since the two-point function of the triangle–hinge model has a behavior quite similar to that of this matrix model, this critical point may correspond not to three-dimensional gravity but to two-dimensional gravity.

The numerical result shows that the critical point $\beta \simeq 6$ of triangle–hinge models do not depend on n. The critical point thus corresponds to the large-n limit

$$n \to \infty,\ \beta = n^2\lambda_4\mu_4 :\ \text{fixed}. \tag{5.5.7}$$

However, this limit is different from the limit that we have considered in Sect. 3.7, which should be corresponds to three-dimensional gravity. Furthermore, we do notinsert the matrix ω, which we need in order to restrict the diagrams to that represent tetrahedral decompositions as we discussed in Sect. 3.7. Because of that, the diagrams as in Fig.5.4 for example, which represents two-dimensional surface rather than three-dimensional volume, still remains and gives the Boltzmann weight proportional to $n^{52}\lambda_4^{48}\mu_4^{48}$. This is dominant

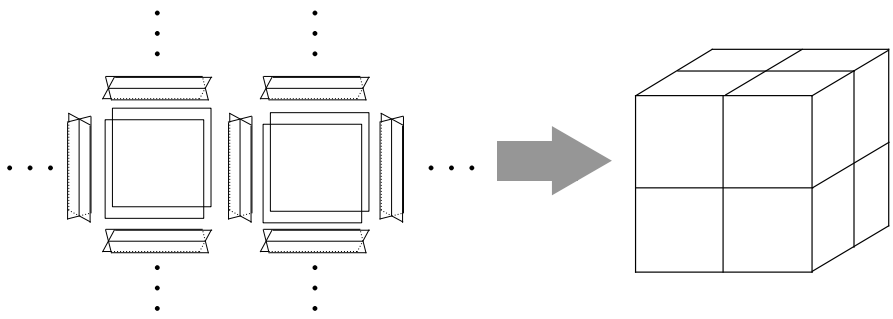


Figure 5.4: An example of singular diagrams that have larger contribution than diagram corressponding to tetrahedral decomposition in the large-n limit. The cube consists of doubloed quadrangles which are glued by four-hinges.

in the limit (5.5.7) compared to the diagrams that represent three-dimensional tetrahedral decompositions. This also implies that the critical point may correspond to two-dimensional gravity.

In order to obtain the phase transition corresponding to three-dimensional gravity, we need to insert ω properly to restrict the diagrams. However, we must do this carefully: The

ω-inserted model is given as

$$\begin{aligned} S(A,B) =& \frac{1}{2}A_{abcd}B^{abcd} \\ & - \frac{\lambda_4}{8n^4}A_{a_1b_1c_1d_1}\cdots A_{a_4b_4c_4d_4}\omega^{d_1a_2}\omega^{b_2c_1}\omega^{d_2a_3}\omega^{b_3c_2}\omega^{d_3a_4}\omega^{b_4c_3}\omega^{d_4a_1}\omega^{b_1c_4} \\ & - \frac{n^2\mu_4}{8}B^{a_1a_2b_2b_1}B^{a_2a_3b_3b_2}B^{a_3a_4b_4b_3}B^{a_4a_1b_1b_4}. \end{aligned} \tag{5.5.8}$$

We can perform similarity transformation

$$A_{abcd} \to A' = A_{a'b'c'd'}(\omega^{-1})^{aa'}(\omega^{-1})^{bb'}(\omega^{-1})^{cc'}(\omega^{-1})^{dd'}, \tag{5.5.9}$$

and obtain

$$\begin{aligned} S(A,B) =& \frac{1}{2}A_{abcd}B^{a'b'c'd'}w^{aa'}w^{bb'}w^{cc'}w^{dd'} \\ & - \frac{\lambda_4}{8n^4}A_{bacd}A_{dcef}A_{fegh}A_{hgab} - \frac{n^2\mu_4}{8}B^{a_1a_2b_2b_1}B^{a_2a_3b_3b_2}B^{a_3a_4b_4b_3}B^{a_4a_1b_1b_4}, \end{aligned} \tag{5.5.10}$$

where ω appears only in the quadratic term (we denote A' simply as A). We can again use the transformations (5.1.18) and (5.1.19), and redefine the models as we discussed in Sect. 3.7, we finally obtain

$$S(A,B) = \frac{1}{2}K_\omega(A,B) - \frac{\lambda_4}{8n^4}\mathrm{tr}A^4 - \frac{n^2\mu_4}{8}\mathrm{tr}B^4 \tag{5.5.11}$$

$$K_\omega(A,B) \equiv A_{abcd}B^{efgh}\begin{bmatrix} \omega^{af}\omega^{gb}\omega^{ch}\omega^{ed} - i\omega^{bf}\omega^{ga}\omega^{ch}\omega^{ed} - i\omega^{af}\omega^{gb}\omega^{dh}\omega^{ec} - \omega^{bf}\omega^{ga}\omega^{dh}\omega^{ec} \\ +i\omega^{ae}\omega^{gb}\omega^{ch}\omega^{fd} + \omega^{be}\omega^{ga}\omega^{ch}\omega^{fd} + \omega^{ae}\omega^{gb}\omega^{dh}\omega^{fc} - i\omega^{be}\omega^{ga}\omega^{dh}\omega^{fc} \\ +i\omega^{af}\omega^{hb}\omega^{cg}\omega^{ed} + \omega^{bf}\omega^{ha}\omega^{cg}\omega^{ed} + \omega^{af}\omega^{hb}\omega^{dg}\omega^{ec} - i\omega^{bf}\omega^{ha}\omega^{dg}\omega^{ec} \\ -\omega^{ae}\omega^{hb}\omega^{cg}\omega^{fd} + i\omega^{be}\omega^{ha}\omega^{cg}\omega^{fd} + i\omega^{ae}\omega^{hb}\omega^{dg}\omega^{fc} + \omega^{be}\omega^{ha}\omega^{dg}\omega^{fc} \end{bmatrix}. \tag{5.5.12}$$

However, since ω is not a Hermitian matrix (Hermiticity conflicts the properties tr ω = tr $\omega^2 = 0$ and tr $\omega^3 \neq 0$), (5.5.11) takes complex value. We then need to face the sign problem, which we will see in the next chapter.

Chapter 6

Numerical tools for complex actions: Lefschetz thimble method

6.1. Sign problem

We first briefly review the problem (so-called "sign problem") that we often encounter when we perform numerical calculation with a complex action.

When we study phase structures of some models, we often perform the Monte Carlo simulation. In Monte Carlo simulation, we calculate the correlation function of operator $O(x)$ as an ensemble average with respect to the configuration generated with probability $P(x) \equiv \frac{1}{Z} e^{-S(x)}$ as

$$\begin{aligned} \langle O \rangle &= \frac{\int dx O(x) e^{-S(x)}}{\int dx e^{-S(x)}} = \frac{1}{Z} \int dx O(x) e^{-S(x)} \\ &\equiv \int dx P(x) O(x) \\ &\simeq \frac{1}{M} \sum_{i=1}^{M} O(x_i). \end{aligned} \tag{6.1.1}$$

In this approach, the statistic error of the correlation function is $\mathcal{O}(M^{-1/2})$, especially independent of the degrees of freedom. This approach is quite useful for real-valued action $S(x)$, however, we are often interested in complex actions, as we already encountered in Chap. 5. Complex actions also appear when we consider the finite density lattice QCD or real time evolution of the system. Since $P(x)$ can no longer be regarded as a probability when the action is complex-valued, we need some prescription to apply Monte Carlo simulation. The naivest prescriptions is so-called reweighting method: we use the real part of the action $S_R(x)$ to generate configurations and incorporate the imaginary part $S_I(x)$ into operators

as

$$\langle O \rangle = \frac{\langle O e^{-iS_I} \rangle_{S_R}}{\langle e^{-iS_I} \rangle_{S_R}} \tag{6.1.2}$$

$$\langle O \rangle_{S_R} \equiv \frac{\int dx O(x) e^{-S_R(x)}}{\int dx e^{-S(x)}} = \frac{1}{Z_{S_R}} \int dx O(x) e^{-S_R(x)}. \tag{6.1.3}$$

Although we can in principle calculate with this prescription, this often fails because the phase $e^{-iS_I(x)}$ oscillates violently when $S_I(x)$ becomes large. For example, let us consider a simple model

$$S(x) = \frac{\beta}{2}(x-i)^2 \tag{6.1.4}$$

$$O(x) = x^2. \tag{6.1.5}$$

We can easily perform the integral and obtain $\langle O \rangle = 1/\beta - 1$. If we apply the reweighting method to this model, $S_R(x) = \frac{\beta}{2}(x^2-1)$, $S_I(x) = -\beta x$ and from (6.1.2) we obtain

$$\langle O e^{-iS_I} \rangle_{S_R} = (\frac{1}{\beta} - 1) e^{-\beta/2} \tag{6.1.6}$$

$$\langle e^{-iS_I} \rangle_{S_R} = e^{-\beta/2}. \tag{6.1.7}$$

Both of them become exponentially small when $\beta \gg 1$. On the other hand, the statistic error can be estimated to be

$$\delta \langle e^{-iS_I} \rangle_{S_R} = \langle \cos(S_I) \rangle_{S_R} - i \langle \sin(S_I) \rangle_{S_R} \tag{6.1.8}$$

$$\delta \langle \cos(S_I) \rangle_{S_R} \sim \frac{1}{\sqrt{M}} \sqrt{\langle \cos(S_I)^2 \rangle_{S_R} - \langle \cos(S_I) \rangle_{S_R}^2}, \tag{6.1.9}$$

where M is the number of data points. In contrast to small $\langle \cos(S_I) \rangle_{S_R}$, $\langle \cos(S_I)^2 \rangle_{S_R}$ remains $\mathcal{O}(1)$ and we thus need exponentially large number of data points enough in order to suppress the error.

To avoid the sign problem, there are two major approaches. One is called complex Langevin method (CLM): In this approach, roughly speaking, we double the degrees of freedom and consider real positive probability $P(x,y)$ that satisfies

$$\int dx P(x) O(x) = \int dx dy P(x,y) O(x+iy) \tag{6.1.10}$$

for any $O(x)$. Since $P(x,y)$ is positive we no longer suffer from the sign problem in this approach. However, CLM suffer from another problem which is called "wrong convergence problem". This means that the Monte Carlo simulations corresponding to $P(x,y)$ may not give the correct answer corresponding to $P(x)$ even when we generate sufficiently large data points. The conditions that judge whether this problem occurs or not is given in [48],

however, the prescription to avoid the wrong convergence problem is still under development. The other approach is called (generalized) Lefschetz thimble method (GLTM): This approach is, roughly speaking, the change of integration contour so that the imaginary part of the action do not oscillates. We will review this approach in the following section.

6.2. Generalized Lefschetz thimble method

In this section, we review the generalized Lefschetz thimble method (GLTM), which is a strong tool to avoid the sign problem.

The idea of GLTM is the change of integration contour from $\mathbb{R}^N$ to N-dimensional submanifold in $\mathbb{C}^N$ where the sign problem becomes milder. In order to do this, we need to define submanifolds in $\mathbb{C}^N$ with good properties. The Lefschetz thimbles J_σ are defined by an antiholomorphic gradient flow with saddle points of the action z_σ as

$$J_\sigma \equiv \left\{ z(t) | \frac{dz_i}{dt} = \frac{\partial \bar{S}(z)}{\partial z_i}, z(-\infty) = z_\sigma \right\}, \tag{6.2.1}$$

where σ labels saddle points. We here put the following assumptions:

- The second derivative of the action (i..e Hesse matrix) do not have zero eigenvalue on the saddle point. ($\det(\frac{\partial S}{\partial z_i \partial z_j}) \neq 0$.)
- The gradient flow do not end at another saddle points. ($\mathrm{Re}(S(z(t)))) \to \infty$ ($t \to \infty$.))

Then the Lefschetz thimble J_σ has the following properties:

- J_σ is N-dimensional submanifold in $\mathbb{C}^N$ including z_σ.
- The imaginary part of the action is constant on J_σ.

The first property can be shown from the fact that J_σ is defined by the gradient flow and the fact that J_σ spans N-dimensional space at least near z_σ since $\det(\frac{\partial S}{\partial z_i \partial z_j}) \neq 0$. The second property can be shown from

$$\begin{aligned} \frac{dS(z(t))}{dt} &= \frac{\partial S(z)}{\partial z_i}\frac{dz(t)}{dt} = \frac{\partial S(z)}{\partial z_i}\frac{\partial \bar{S}(z)}{\partial z_i} \\ &= \left| \frac{\partial S(z)}{\partial z_i} \right|^2 \geq 0, \end{aligned} \tag{6.2.2}$$

by using the flow equation. When we change the dynamical variables to Lefschetz thimbles, we can rewrite the partition function as

$$Z = \sum_\sigma n_\sigma \int_{J_\sigma} dz e^{-S(z)}. \tag{6.2.3}$$

Here, n_σ is integers called "intersection numbers", and can be represented as the number of intersection between another submanifold K_σ defined as

$$K_\sigma \equiv \left\{ z(t) | \frac{dz_i}{dt} = -\frac{\partial \bar{S}(z)}{\partial z_i}, z(-\infty) = z_\sigma \right\} \tag{6.2.4}$$

and real axis $\mathbb{R}^N$. [1] However, since Z is represented as the sum of integrations, it is hard to perform Monte Carlo simulations when more than two thimbles contribute to Z.

In the generalized Lefschetz thimble method, which is proposed in [40], the change of integration contour itself is defined by the gradient flow:

$$\frac{dz_i}{dt} = \frac{\partial \bar{S}(z)}{\partial z_i} \tag{6.2.5}$$

$$z(0)_i = x_i \in \mathbb{R} \tag{6.2.6}$$

When we take the flow time t large, the dynamical variables flow from real axis to neighbor of Lefschetz thimbles, and go to infinity or a singular point of the action. Thus the integration contour sticks to thimbles for large t. The Jacobian can also be calculated by flow equation

$$\begin{aligned} \frac{dJ(t)_{ij}}{dt} &\equiv \frac{d}{dt}\frac{\partial z(t)_i}{\partial x_j} \\ &= \frac{\partial}{\partial x_j}\frac{dz(t)_i}{dt} \\ &= \frac{\partial^2 S(z(\bar{t})}{\partial z_i \partial z_k}\frac{dz(t)_k}{dt} = \frac{\partial^2 S(z(\bar{t})}{\partial z_i \partial z_k} J(t)_{kj} \end{aligned} \tag{6.2.7}$$

$$J(t=0)_{ij} = \delta_{ij}. \tag{6.2.8}$$

We finally rewrite the partition function as

$$Z = \int dx \det J e^{-S(z(t;x))} = \int dx e^{-S_{\text{eff}}(x;t)} e^{-i\text{Im}S(z(t;x)) + i \arg\det J} \tag{6.2.9}$$

$$S_{\text{eff}}(x;t) = \text{Re}S(z(t;x)) - \log|\det J|. \tag{6.2.10}$$

The reweighting method now becomes varid. Although it seems strange that ImS is constant along the flow, ReS increases drastically and $z(t;x)$ is finite only when x is near the point of intersection between the real axis and K_σ (the intersection point flows into the saddle point), and thus ImS is almost constant in this narrow region.

Let us see that this method actually works well through the simple model (6.1.4). In this model we can solve the flow equation (6.2.5) and (6.2.7) exactly and obtain

$$z(t;x) = e^{\beta t}x + i(1 - e^{-\beta t}) \tag{6.2.11}$$

$$J(t;x) = e^{\beta t}. \tag{6.2.12}$$

[1] The sign of n_σ can be uniquely defined when we define the direction of integrations. The appearance of these integers can be easily understood when we consider GLTM which we will see below. However, we do not discuss on n_σ anymore because we do not have to calculate n_σ in GLTM.

The partition function is

$$\langle x^2 \rangle = \frac{\langle z(t;x)^2 e^{i\beta t} \rangle}{\langle e^{i\beta t} \rangle} \tag{6.2.13}$$

$$\langle O(x) \rangle_{\text{eff}} \propto \int dx e^{-\frac{\beta}{2}(x^2 e^{2\beta t})} \tag{6.2.14}$$

$$\langle z(t;x)^2 e^{i\beta x} \rangle_{\text{eff}} = e^{-\frac{\beta}{2}e^{-2\beta t}} \left(\frac{1}{\beta} - 1\right) \tag{6.2.15}$$

$$\langle e^{i\beta x} \rangle_{\text{eff}} = e^{-\frac{\beta}{2}e^{-2\beta t}}. \tag{6.2.16}$$

Here, the integrand of (6.2.14) is finite only for an narrow region of x ($z(t;x)$ spread widely, however), and thus $e^{i\beta x}$ do not oscillate and the damping with respect to β is suppressed for large t. Therefore we can obtain the denominator of order $\mathcal{O}(1)$ and avoid the sign problem.

GLTM is quite versatile method in contrast to complex Langevin method, which suffers from wrong convergence problem. However, there are two other problems in GLTM: One is that the configuration would be trapped on a single thimble when we take t large. This is because the effective action (6.2.10) has exponentially or infinitely high potential barrier among thmbles, and this problem (so-called multimodal problem) gives wrong result precisely. Of course this problem can be mild for small t, however, the sign problem comes back to life. We cannot always tune t in general so that both of these problems do not appear. Another problem is the computational cost of this method. The computational cost is $\mathcal{O}(N^3)$ because of the determinant of the Jacobian. Furthermore, we should solve flow equation carefully since $z(t;x)$ moves drastically and flows off to infinity in finite flow time in general. Therefore solving the flow equation takes many computational costs (the flow equation for Jacobian itself need $\mathcal{O}(N^3)$ computational costs for nonlocal action). Of course they are drastically reduced in contrast to the reweighting method, which need exponentially large computational costs, it is still hard to apply GLTM to large-scale numerical calculation such as lattice QCD.

6.3. Parallel tempering algorithm for Lefschetz thimble method

In the previous section we briefly review the generalized Lefschetz thimble method and pointed two problems. To solve the first problem, we proposed to introduce parallel tempering algorithm by regarding the flow time as a tempering parameter. In this section we review our work [4].

6.3.1. Simulated tempering

Before introducing the parallel tempering algorithm, we first review the simulated tempering algorithm [49] which should help the understanding of the idea of tempering algorithm. In simulated tempering, we consider the parameter λ which reduces the multimodality of the action. Usually λ is overall constant of the action, but we here leave λ as a general parameter. In simulated tempering, we expand the configuration space and regard λ as dynamical variables: That is, we consider several $\lambda = \lambda_0 \geq \cdots \geq \lambda_A$ and perform the Monte Carlo simulation corresponding to the equilibrium distribution

$$P_{\rm eq}(x;a) = w_a e^{-S(x;\lambda_a)}. \tag{6.3.1}$$

Here, the original action is corresponding to $\lambda = \lambda_A$, and $\lambda = \lambda_0$ is sufficiently small so that the multimodal problem do not occur for $S(x;\lambda_0)$. The weight w_a is constant which should be tuned by hand or adaptiively so that the appearance ratios of the configuration with a ($= \int dx P_{\rm eq}(x;a) = Z(\lambda_a)w_a$) are finite for all $a = 0, \cdots, A$. We finally obtain the configuration linked to a, and calculate the correlation function by using only configurations corresponding to $a = A$. In this distribution (6.3.1), the configuration can move to other modes through small λ_a, although the original action is highly multimodal.

We consider two types of transition in simulated tempering. One is the transition from (x,a) to (x',a), which is an ordinary transition used in Monte Carlo simulations (such as Metropolis-Hasting algorithm, Langevin algorithm, or hybrid Monte Carlo algorithm) for the action $S(x;\lambda_a)$. The other is the transition from (x,a) to (x,a'), where $a' = a \pm 1$. To keep the detailed balance condition with respect to (6.3.1), we accept the transition with probability

$$\begin{aligned} &\min\{1, \frac{P_{\rm eq}(x;a')}{P_{\rm eq}(x;a)}\} \\ =&\min\{1, \frac{w_{a'}}{w_a} e^{-S(x;\lambda_{a'})+S(x;\lambda_a)}\}. \end{aligned} \tag{6.3.2}$$

The multimodal problem can be solved by using the simulated tempering algorithm, however, we must tune w_a by hands or adaptively, and thus we need to estimate (at least roughly) the partition function $Z(\lambda_a)$. This is in general hard to calculate, and in particular we cannot apply this method to the effective action (6.2.10). The parallel tempering algorithm is the method that overcame this problem, and we need not to calculate w_a (implicitly w_a is tuned to be $1/Z(\lambda_a)$).

6.3.2. Parallel tempering

In the parallel tempering algorithm, unlike in the case of the simulated tempering algorithm, we consider the replica of configuration space as

$$x \in M \to \{x_a\}_{a=0,\cdots,A} \in M^{A+1}. \tag{6.3.3}$$

We perform the Monte Carlo simulation with respect to this extended configuration space with the equilibrium distribution

$$P_{\rm eq}(x_0, \cdots, x_A) = \prod_a \frac{1}{Z(\lambda_a)} e^{-S(x_a;\lambda_a)}. \tag{6.3.4}$$

We again consider two types of transition: One is again the transition along x-direction, such as Metropolis-Hasting algorithm, Langevin algorithm, and so on. Here the configuration x_a evolves with the action $S(x_a;\lambda_a)$ and do not interact other x_a. The other transition is the exchange of configuration among replicas. In parallel tempering algorithm, we exchange x_a and $x_{a'}$ with the probability

$$w_\alpha(x, x') = \min\Big(1, \frac{e^{-S(x';\lambda_\alpha)-S(x;\lambda_{\alpha+1})}}{e^{-S(x;\lambda_\alpha)-S(x',\lambda_{\alpha+1})}}\Big), \tag{6.3.5}$$

due to the detailed balance condition with respect to (6.3.4). We can solve the sign problem as is the case for the simulated tempering algorithm, but (6.3.5) do not require additional parameters to tune such as w_a. All we have to do is to choose λ_a (and A) so that the exchange can be done with finite probability. [2]

6.3.3. Parallel tempering algorithm for GLTM

Our proposal is to take the flow time t as such tempering parameter. The basic algorithm is then as follows.[3]

- Step 0. Fix the maximum flow time T which should be sufficiently large such that the sign problem disappears there, and pick up flow times $\{t_\alpha\}$ from the interval $[0, T]$ with $t_0 = 0 < t_1 < \cdots < t_A = T$. The values of A and t_α are determined manually or adaptively to optimize the acceptance rate in Step 3 below.

[2] In the simulated tempering algorithm, the acceptance rate of the transition along a-direction is

$$\int dx \min\{\frac{1}{Z(\lambda_a)} e^{-S(x;\lambda_a)}, \frac{1}{Z(\lambda_{a'})} e^{-S(x;\lambda_{a'})}\}. \tag{6.3.6}$$

This means that we need much numbers of λ_a if the distribution changes drastically with respect to λ. In the parallel tempering algorithm, we should take λ_a in the same manner for the simulated tempering algorithm.

[3] To make discussions simple, we only take the flow time as a tempering parameter. The algorithm can be readily extended such that other parameters are included as extra tempering parameters.

- Step 1. Choose an initial value $x_\alpha \in \mathbb{R}^N$ for each replica α, and numerically solve the differential equations (6.2.5) and (6.2.7) to obtain the triplet $(x_\alpha, z_{t_\alpha}, J_{t_\alpha})$.

- Step 2. For each α, construct a Metropolis process to update the value of x. Explicitly, we take a value x'_α from x_α using a symmetric proposal distribution, and recalculate the triplet $(x'_\alpha, z'_{t_\alpha}, J'_{t_\alpha})$ using the x'_α as the initial value. We then update x_α to x'_α with the probability $\min(1, e^{-\Delta S_{\text{eff},\alpha}})$, where

$$\begin{aligned}\Delta S_{\text{eff},\alpha} &\equiv S_{\text{eff}}(x'_\alpha, t_\alpha) - S_{\text{eff}}(x_\alpha, t_\alpha) \\ &= S_R(z'_{t_\alpha}) - \ln\left|\det J'_{t_\alpha}\right| - S_R(z_{t_\alpha}) + \ln\left|\det J_{t_\alpha}\right| \end{aligned} \tag{6.3.7}$$

  [recall that $S_{\text{eff}}(x;t) = S_R\big(z_t(x)\big) - \ln\left|\det J_t(x)\right|$, eq. (6.2.10)]. We repeat the process sufficiently many times such that local equilibrium is realized for each α.

- Step 3. Starting from $\alpha = 0$ through $\alpha = A-1$, swap the values of x between two adjacent replicas α and $\alpha+1$ with the probability

$$w_\alpha(x_\alpha, x_{\alpha+1}) \equiv \min\Big(1, \frac{e^{-S_{\text{eff}}(x_{\alpha+1};t_\alpha) - S_{\text{eff}}(x_\alpha;t_{\alpha+1})}}{e^{-S_{\text{eff}}(x_\alpha;t_\alpha) - S_{\text{eff}}(x_{\alpha+1};t_{\alpha+1})}}\Big). \tag{6.3.8}$$

- Step 4. After repeating steps 2 and 3 sufficiently many times, get a triplet $(x_A, z_{t_A=T}, J_{t_A=T})$ from $\alpha = A$ as an element of a sample.

- Step 5. After repeating steps 2 to 4, we obtain a sequence of triplets $\{(x_A^{(a)}, z_{t_A=T}^{(a)}, J_{t_A=T}^{(a)})\}\}$, which we use to estimate the expectation value:

$$\langle \mathcal{O}(x)\rangle \approx \frac{\sum_a e^{i\left[\arg\det J_T^{(a)} - S_I(z_T^{(a)})\right]}\mathcal{O}\big(z_T^{(a)}\big)}{\sum_a e^{i\left[\arg\det J_T^{(a)} - S_I(z_T^{(a)})\right]}}. \tag{6.3.9}$$

Note that the action with $t = t_0 = 0$ is the original action for which the sign problem exists. However, as can be seen from (6.3.9), the sample average is taken only with respect to the action with $t = t_A = T$, so that we need not worry about the original sign problem although we include configurations near $t = 0$. Also, T can be taken to be sufficiently large such that the sign problem disappears if we complement intermediate flow times sufficiently (with larger A). Thus, this simple algorithm solves the two obstacles simultaneously; the original sign problem at $t = 0$ is resolved at $t = T$ and the multimodal problem is resolved even at $t_A = T$ by passing through configurations near $t_0 = 0$.[4] Note that this algorithm is valid for GLTM. The parallel tempering algorithm in general fails when $e^{-S(x;\lambda)}$ is sensitive to λ, because the acceptance rate becomes small. However, if we regard the flow time as a

[4]We have implicitly assumed that the action at $t_0 = 0$ does not cause a multimodality in the configuration space. If this is not the case, we further introduce other parameters (such as the coefficient of the action) as extra tempering parameters or prepare flow times $\{t_\alpha\}$ such that $t_0 < 0$.

tempering parameter, the center of a mode of $e^{-S_{\rm eff}(x;t)}$ is fixed near the intersection point between the real axis and K_σ. Thus we can tune the acceptance rate to be finite even for large T. We also note for the relation between the set of tempering parameter $\{t_a\}$ and the acceptance rate. When we take t large so that $z(t;x)$ is near the thimble, the dispersion of a mode of $\propto e^{-S_{\rm eff}(x;t_a)}$ is controlled by the eigenvalues of Hesse matrix on the saddle point and exponentially small with respect to t. We thus can estimate the acceptance rate r as $\log r \propto N\delta t$, where $\delta t = t_{a+1} - t_a$. This means that we need $A \sim N$ replicas so that we take $\delta t \sim 1/N$ to keep r finite. Therefore the total computational cost of this algorithm is $\mathcal{O}(N^4)$. Although this is still large for practical use, this is at most polynomail of N, in contrast to the reweighting method which need exponentially large computational costs.

6.4. Example: massive Thirring model at finite density

In this section we investigate the (0+1)-dimensional massive Thirring model at finite density [50, 51, 39] to exemplify that the algorithm given in the previous section does work.

The $(0+1)$-dimensional massive Thirring model is defined from the standard $(1+1)$-dimensional massive Thirring model by dimensional reduction. With an auxiliary field $\phi(\tau)$, the continuum representation of the grand partition function $Z = \mathrm{tr}\, e^{-\beta(H-\mu Q)}$ is given by the path integral

$$Z = \int_{\rm PBC} [d\phi(\tau)] \int_{\rm ABC} [d\bar\psi(\tau) d\psi(\tau)]\, e^{-S[\phi,\bar\psi,\psi]}, \tag{6.4.1}$$

where the Euclidean action takes the form

$$S[\phi,\bar\psi,\psi] = \int_0^\beta d\tau \left[-\bar\psi\big(\gamma^0(\partial_0 + i\phi + \mu) + m\big)\psi + \frac{1}{2g^2}\,\phi^2\right], \quad \left[\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right], \tag{6.4.2}$$

and the ϕ integral (the ψ, $\bar\psi$ integral) obeys the periodic (anti-periodic) boundary condition. In realizing the model on the lattice, we discretize the Euclidean time as $\tau = na$ ($n = 1, \ldots, N$) with $\beta = Na$ (N: even), and follow the prescription of [51], where $\phi(\tau) = \phi_n$ is treated as a $U(1)$ gauge potential and is combined with the chemical potential μ to become a link variable of complexified gauge group, $e^{(i\phi(\tau)+\mu)a} = e^{i\phi_n a} e^{\mu a} \equiv U_n\, e^{\mu a}$, as proposed in [52]. Then, by using a staggered fermion formulation, the grand partition function is given by

$$Z = \int (dU)\, e^{-S(U,\bar\chi,\chi)}, \tag{6.4.3}$$

where $(dU) = \prod_n (dU_n) \equiv \prod_n [d(\phi_n a)/2\pi]$ (with $\phi_n a \in (-\pi, \pi]$) and

$$S(U,\bar\chi,\chi) = -\sum_{n,\ell=1}^{N} \bar\chi_n\, D_{n\ell}(U)\, \chi_\ell + \frac{1}{2g^2 a} \sum_{n=1}^{N} \Big[1 - \frac{1}{2}(U_n + U_n^{-1})\Big] \tag{6.4.4}$$

with $D_{n\ell}(U) \equiv (1/2)\big(U_n e^{\mu a}\delta_{n+1,\ell} - U_{n-1}^{-1}e^{-\mu a}\delta_{n-1,\ell} - U_N e^{\mu a}\delta_{n,N}\delta_{\ell,1} + U_N^{-1}e^{-\mu a}\delta_{n,1}\delta_{\ell,N}\big) + ma\,\delta_{n\ell}$. We henceforth set $a = 1$ and treat m, μ and g^2 as dimensionless parameters.

After carrying out the fermion integration, we obtain

$$Z = \int (dU)\, \det D(U)\, e^{-(1/2g^2)\sum_n [1-(1/2)(U_n + U_n^{-1})]}, \tag{6.4.5}$$

which can be calculated analytically [51] as

$$Z = \frac{e^{-N\alpha}}{2^{N-1}}\big[\cosh(N\mu)\, I_1^N(\alpha) + \rho_+\, I_0^N(\alpha)\big], \tag{6.4.6}$$

where $\alpha \equiv 1/(2g^2)$, $2\rho_\pm \equiv (\sqrt{m^2+1}+m)^N \pm (\sqrt{m^2+1}-m)^N$, and $I_n(\alpha)$ is the modified Bessel function of the first kind of order n. Using this expression, we obtain the analytic form of the chiral condensate as

$$\langle\bar\chi\chi\rangle = \frac{1}{N}\frac{\partial \ln Z}{\partial m} = \frac{\rho_-\, I_0^N(\alpha)}{\sqrt{m^2+1}\,\big[\cosh(N\mu)\, I_1^N(\alpha) + \rho_+\, I_0^N(\alpha)\big]}. \tag{6.4.7}$$

We now numerically evaluate the chiral condensate by using the complex action

$$S(U) \equiv \frac{1}{2g^2}\sum_n \Big[1 - \frac{1}{2}\,(U_n + U_n^{-1})\Big] - \ln\det D(U) \tag{6.4.8}$$

as

$$\langle\bar\chi\chi\rangle = \Big\langle \frac{1}{N}\,\mathrm{tr}\,\Big(D^{-1}(U)\frac{\partial D(U)}{\partial m}\Big)\Big\rangle = \Big\langle \frac{1}{N}\,\mathrm{tr}\, D^{-1}(U)\Big\rangle, \tag{6.4.9}$$

where

$$\langle\mathcal{O}(U)\rangle \equiv \frac{\int (dU)\, e^{-S(U)}\,\mathcal{O}(U)}{\int (dU)\, e^{-S(U)}}. \tag{6.4.10}$$

It is easy to check that $[\det D(U;\mu)]^* = \det D(U;-\mu)$, so that the second term of (6.4.8) is complex-valued for real μ, and the sign problem should arise when N is large.

We follow the steps given in Sect. 6.3 with the complexification of the variables $\phi_n \in (-\pi,\pi]$. We first set the largest flow time to $T = 2$, and prepare flow times $\{t_\alpha\}$ ($\alpha = 0,1,2,\ldots,20$) with equal separations as $t_0 = 0$, $t_1 = 0.1$, $t_2 = 0.2$, $\ldots$, $t_{20} = 2.0$.[5] Fig. 6.1 shows $\phi(t) \equiv (1/N)\sum_n \phi_n(t)$ at flow times $t = t_\alpha$ with $\phi_n(0)$'s set to the same value. As Step 1 of our algorithm, we make a cold start ($\phi_n = 0$) for every replica α, and numerically

[5]The fermion determinant at flow time $t = 0$ is given by $\det D = (1/2^N)\,[\zeta + \zeta^{-1} + (\sqrt{2}+1)^N + (\sqrt{2}-1)^N)]$ where $\zeta \equiv e^{i\sum_n \phi_n + N\mu}$. Note that $\det D$ vanishes at $\sum_n \phi_n = \pi \pmod{2\pi}$ when $\mu = \log(\sqrt{2}+1) \sim 0.881$, which gives a multimodal distribution even at $t = 0$. However, this can be handled without introducing another tempering parameter, because each mode has a rather wide distribution (with not-too-small values near the boundary) and thus the whole configuration space can be easily explored by setting the interval of proposal distribution to be a large value as in this paper.

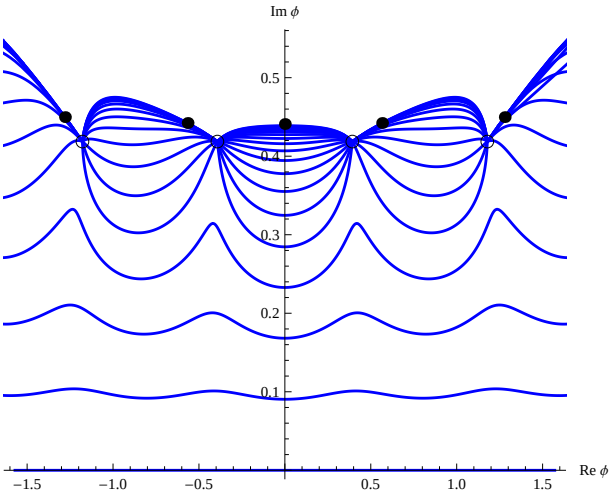


Figure 6.1: $\phi(t) = (1/N)\sum_n \phi_n(t)$ at flow times $t = 0,\ 0.1,\ 0.2, \ldots, 1.9,\ 2.0$ from bottom to top [4]. The parameters are set to $N = 8$, $m = 1$, $\mu = 1.3$, $g^2 = 1/2$. The full circles are the critical points of S, and the empty circles are the log singularities of S.

solve the differential equations (6.2.5) and (6.2.7) with the adaptive 4th-order Runge-Kutta method to obtain the triplet $(x_\alpha, z_{t_\alpha}, J_{t_\alpha})$. We repeat the Metropolis process twenty times in Step 2,[6] which is followed by a single sequence of swapping of Step 3. We then repeat Steps 2 and 3 ten times (as Step 4). With the first 5 data discarded as initial sweeps, we estimate correlation functions with 10^4 data. Fig. 6.2 shows the absolute value of the denominator in (6.3.9) (divided by the sample size) as a function of μ with the other parameters set to $N = 8$, $m = 1$, $g^2 = 1/2$. The blue points are the result for the flow time $T = 0$. They correspond to the usual reweighting calculus, and show that the sign problem actually exists for $\mu \gtrsim 1.0$. The green (red) points are the result for the flow time $T = 2$ without (with) the parallel tempering implemented. The results show that the sign problem disappears at $T = 2$.

Fig. 6.3 shows the chiral condensate $\langle\bar{\chi}\chi\rangle$ as a function of μ. The other parameters are again set to $N = 8$, $m = 1$, $g^2 = 1/2$. The dotted line represents the analytic result (6.4.7). The blue points are the result for the flow time $T = 0$ and exhibit large statistical errors reflecting the sign problem. The green points are the result for the flow time $T = 2$ without the parallel tempering implemented. They have small statistical errors, but exhibit statistically significant discrepancies from the analytic result. This should be attributed, as discussed in detail in [40], to the fact that the dominant contributions come only from a single thimble for such large T. The red points are the result for the flow time $T = 2$ now with

[6] As a proposal distribution we use the uniform distribution within the interval $[-\epsilon, \epsilon]$, where ϵ is chosen randomly from $\{1,\ 10^{-1},\ 10^{-2}, \ldots,\ 10^{-[2t_\alpha+1]}\}$ ($[k]$ is the floor of k).

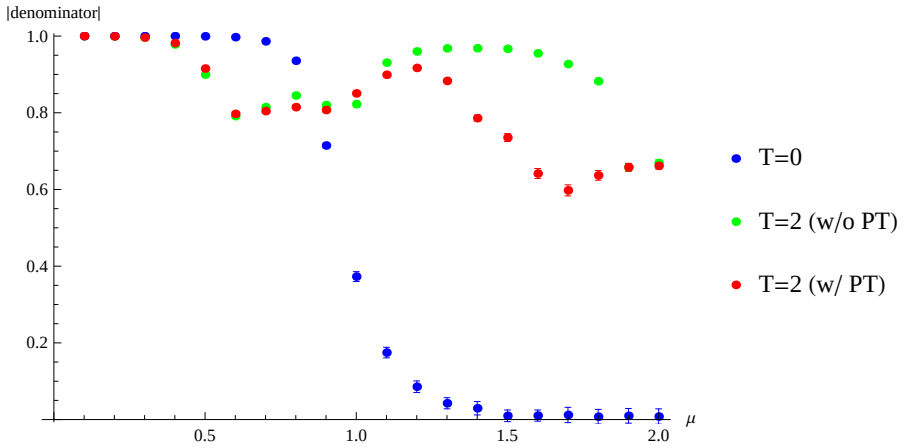


Figure 6.2: The absolute value of the denominator in (6.3.9) (divided by the sample size) as a function of μ with the other parameters set to $N = 8$, $m = 1$, $g^2 = 1/2$ [4]. The blue points are the result for the flow time $T = 0$ and show that the sign problem actually exists for $\mu \gtrsim 1.0$. The green (red) points are the result for the flow time $T = 2$ without (with) the parallel tempering implemented, and show that the sign problem disappears at $T = 2$.

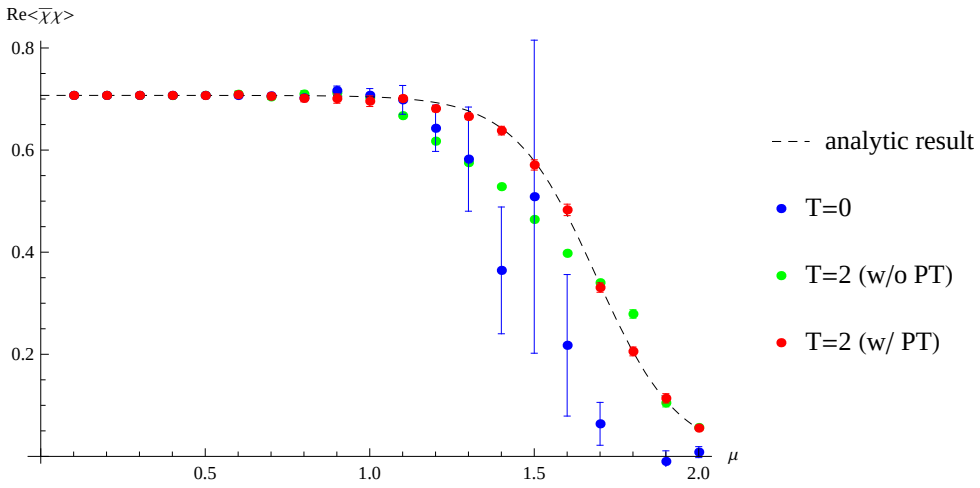


Figure 6.3: Chiral condensate $\langle\bar{\chi}\chi\rangle$ as a function of μ with the other parameters set to $N = 8$, $m = 1$, $g^2 = 1/2$ [4]. The dotted line represents the analytic result (6.4.7). The blue (green) points are the result for the flow time $T = 0$ ($T = 2$) without the parallel tempering implemented. The red points are the result for the flow time $T = 2$ with the parallel tempering implemented.

the parallel tempering implemented. They show a good agreement with the analytic result, which implies that contributions from various thimbles are correctly taken into account through the parallel tempering. This can be confirmed by Fig. 6.4 and Fig. 6.5. Fig. 6.4 exhibits the histogram of $\phi = (1/N)\sum_n \phi_n$ at $T = 2$. We see that the configurations are

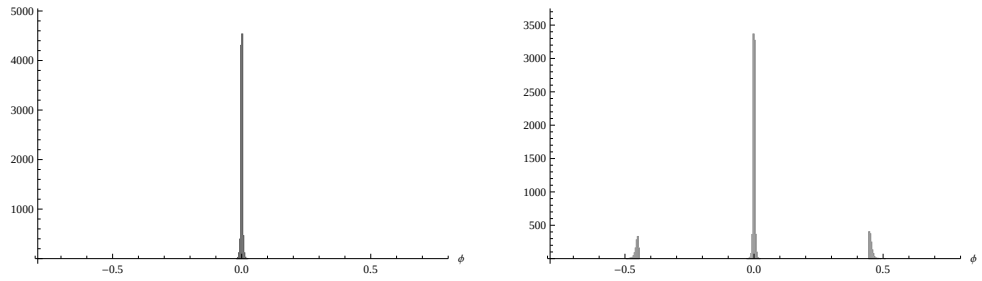


Figure 6.4: Histograms of $\phi = (1/N)\sum_n \phi_n$ at the flow time $T = 2$ without/with the parallel tempering implemented (left/right) [4]. The parameters are set to $N = 8$, $m = 1$, $\mu = 1.3$, $g^2 = 1/2$.

concentrated on a single thimble if the parallel tempering is not implemented (left), while they are spread out on various thimbles if the parallel tempering is implemented (right). Fig. 6.5 shows the average acceptance rates for the swaps between replicas α and $\alpha + 1$ in Step 3. We see that the swapping is carried out very well because the average acceptance

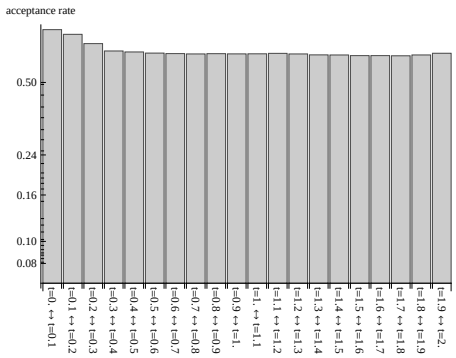


Figure 6.5: Average acceptance rate for the swap between replicas α and $\alpha + 1$ in Step 3 [4]. The parameters are set to $N = 8$, $m = 1$, $\mu = 1.3$, $g^2 = 1/2$.

rate is more than 50% for all pairs $(\alpha, \alpha + 1)$.

6.5. Alternative approach: deep learning method

The generalized Lefschetz thimble method suffer from high computational cost. In this section, we review another promising approach [42] which solves both of the sign problem

and the multimodal problem, while reducing the computational cost. Although this method seems to be quite strong, this is not so strong yet that we can apply to any models because it loses the versatility. Since we apply the above parallel tempering method in the next chapters, the reader can skip this section.

In GLTM we consider the change of variables to near the Lefschetz thimbles to reduce the oscillation of the imaginary part of the action. The effective action becomes highly multimodal because quite narrow regions of x spread to the thimbles by using the gradient flow (6.2.5). To make matters worse, we have to solve the flow equations carefully because $z(t;x)$ moves drastically and flows off to infinity in finite flow time in general. However, we can consider another way of the change of variables such that only the imaginary part of the variables moves (which should make the multimodality of the effective action mild). Although this is hard to represent by flow equations, fortunately we can know the "solutions" of the flow equation (6.2.5). Alexandru et.al. propose to realize the change of variables near the thimbles by using deep learning: We first solve (6.2.5) numerically for a number of $x = x_0$, and estimate the imaginary part y of $z(t;x_0) = x + iy$ from the real part x by deep learning. Although the estimated result contains error from the solution y, we do not need the exact positions of Lefschetz thimbles, and we have only to obtain the change of variables to near the thimbles and the Jacobian of the change. In deep learning method which we will see below, the Jacobian can be calculated by chain rules, and the computational cost is $\mathcal{O}(N^3)$ because of the calculation of the Jacobian and its determinant. Furthermore, we need not to solve the gradient flow after learning, and this also reduces the computational cost.

6.5.1. Deep learning method

In this subsection we briefly review the deep learning and the way to define the change of variables.

In the deep learning method, we consider a function $f(x;\lambda)$ with parameter λ that estimate the solution $t = (t_i(x))$ for all $x \in M$. [7] If we define a distance between $f(x;\lambda)$ and $t(x)$, we can tune parameters λ by using the gradient descent method. In deep learning method, we set $f(x;\lambda)$ as combinations of linear transformations and simple nonlinear transformations, and tune $f(x;\lambda)$ by using the gradient descent method. We call each component as "layer".

For example, $f(x;\lambda)$ can be constructed by combining the following two layers alter-

[7] We only consider a supervised learning. In general, in addition to a supervised learning, we can consider unsupervised learning by using clustering algorithms or anomaly detections. We do not discuss further in this direction.

nately:

$$\mathrm{Affine}(x; w, b)_i = w_{ij} x_j + b_i \tag{6.5.1}$$

$$\mathrm{SoftPlus}(x)_i = \log(1 + e^{x_i}), \tag{6.5.2}$$

where the parameters λ is the set of the weight w and the bias b. The Affine layer $\mathrm{Affine}(x; w, b)$ is a linear transformation, and the Softplus layer is a nonlinear transformation. We can construct $f(x; \lambda) = f(x; w^{(1)}, b^{(1)}, \cdots)$ by combining these layers as in Fig. 6.6. The layers which have index of hidden nodes, which are neither input nodes nor output

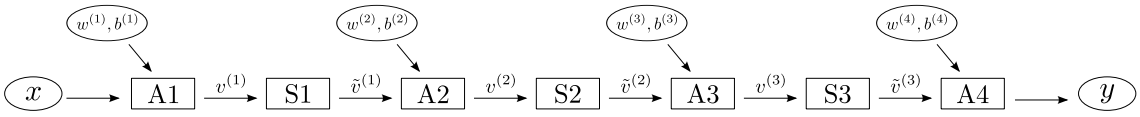


Figure 6.6: A conceptual diagram of the deep learning. The flow represents the order of calculation. The Affine (or Softplus) layers are denoted as A (or S) with boxes. The input (or the output) is denoted as x (or y), and the results of the Affine (or Softplus) layers are denoted as $v^{(k)}$ (or $\tilde{v}^{(k)}$).

nodes, are called "hidden layer" (usually we call a set of linear transformation and nonlinear transformation as a single hidden layer). In the following, we consider as an example the algorithm with three hidden layer as in Fig. 6.6. We then define the cost function which quantifies the distance between $y = f(x; \lambda)$ and $t(x)$ as

$$\mathrm{cost}(\lambda) \equiv \sum_x C(y(x; \lambda), t(x)) \tag{6.5.3}$$

$$C(y(x; \lambda), t(x)) = \sqrt{\sum_i (f(x)_i - t(x)_i)^2}, \tag{6.5.4}$$

where the summation $\sum_x$ is taken over known data point of $t(x)$ (which we call as "teacher data"). The gradient with respect to the parameters $\partial \mathrm{cost}/\partial\lambda = \sum_x \partial C/\partial\lambda$ can be calculated by chain rule. For example,

$$\frac{\partial C(y,t)}{\partial w^{(3)}_{ij}} = \frac{\partial C(y,t)}{\partial y_k} \left.\frac{\partial y(z; w^{(4)}, b^{(4)})_k}{\partial z_l}\right|_{z=\tilde{v}^{(3)}} \left.\frac{\partial \tilde{v}^{(3)}(z)_l}{\partial z_m}\right|_{z=v^{(3)}} \left.\frac{\partial v^{(3)}(z; w^{(2)}, b^{(2)})_m}{\partial w^{(2)}_{ij}}\right|_{z=\tilde{v}^{(2)}}. \tag{6.5.5}$$

This is also called backpropagation method due to the order of the calculation.

By using the gradient descent method with above gradient, we can tune the weight and the bias iteratively and decrease the cost function. However, there is a case for the parameters to reach to a destination with oscillatory behavior, and then this method needs a large number of iteration. This method is also expensive when we consider a large number of

teacher data. To avoid these obstacles, we often choose a subset of teacher data randomly for each iteration (this technique is called stochastic gradient descent), and using an improved method instead of the original gradient descent method. We here review the adaptive momentum estimate method (Adam) [53] as an example of improved methods. In Adam, for s-th update, the parameter λ is updated as

$$\lambda \to \lambda' = \lambda - \frac{\tilde{\eta}_s}{1-\alpha^s}\tilde{\nabla C}_s \tag{6.5.6}$$

$$\tilde{\nabla C}_s = (1-\alpha)\tilde{\nabla C} + \tilde{\nabla C}_{s-1} \tag{6.5.7}$$

$$\tilde{\eta}_s = \frac{\eta}{\sqrt{\frac{\tilde{v}_s}{1-\beta^s}} + \epsilon} \tag{6.5.8}$$

$$\tilde{v}_s = (1-\beta)\left|\tilde{\nabla C}_s\right|^2 \beta \tilde{v}_{s-1} \tag{6.5.9}$$

$$\tilde{\nabla C}_0 = 0 \tag{6.5.10}$$

$$\tilde{v}_0 = 0, \tag{6.5.11}$$

where ∇C is the original gradient and α, β, ϵ and η are constants. α and β play the role of disturbing rapid change of the gradient. ϵ is a small constant which prevents the divergence, and η is the learning rate. These are recommended to be set $\alpha = 0.9, \beta = 0.999, \epsilon = 10^{-8}, \eta = 10^{-3}$ in the original paper [53]. The speed of learning also depends on the initial values of λ. The initial values are often randomly selected from a normal distribution with some standard deviation σ. The standard deviation is taken by referring the number of nodes just before the transformation and the shape of nonlinear function, and it is recommended by He [54] to take $\sigma = \sqrt{2/N}$ for the Softplus layer, where N is the number of nodes. [8]

We can define the change of variables by using above $f(x;\lambda)$ as

$$x \to z = x + if(x;\lambda). \tag{6.5.13}$$

The Jacobian can be calculated by chain rule as

$$J(x;w,b)_{ij} = \frac{\partial h(x;w,b)_i}{x_j} = \frac{\partial y(z;w^{(4)},b^{(4)})_i}{\partial z_{k_1}}\bigg|_{z=\tilde{v}^{(3)}} \cdots \frac{\partial v^{(1)}(z;w^{(1)},b^{(1)})_{k_8}}{\partial z_j}\bigg|_{z=x}. \tag{6.5.14}$$

The total algorithm can be summarized as follows:

- Step 1. We solve the gradient flow (6.2.5) for sufficiently large number of $x = x_0$ to make a training set $\{z(t;x_0^{(k)})\}_k$. Here, the number of x_0 is large so that $\{\mathrm{Re}z(t;x_0^{(k)})\}_k$

[8] Strictly speaking, this is proposed for ReLU layer

$$\mathrm{ReLU}(x) = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}. \tag{6.5.12}$$

We here disregard the difference between ReLU layer and Softplus layer.

covers all region of $x \in M$. If the action has symmetry we should restrict x to representative elements of the symmetry, because $f(x;\lambda)$ cannot reflect the symmetry. x_0 need not to follow some specific distribution such as $e^{-S_{\text{eff}}}$, and we need not to calculate the Jacobian here.

- Step 2. Tuning the weight and the bias by using Adam method and proper initial values.

- Step 3. Define the effective action by using (6.5.13) and (6.5.14) as

$$S_{\text{eff}}(x;w,b) = S(z(x;w,b)) - \log(\det J(x;w,b)) \tag{6.5.15}$$

$$z(x;w,b)_i \equiv x_i + ih(x;w,b)_i, \tag{6.5.16}$$

  and perform Monte Carlo simulation with respect to S_{eff} by using reweighting method.

We can solve the sign problem by using reweighting method since $z(x;w,b)$ should be near Lefschetz thimbles and thus the phase should not oscillate. Furthermore, unlike $z(t;x)$ defined by (6.2.5), wide region fo x represent the shape of thimbles, and thus the multimodal problem should be relatively mild. If we set the number of nodes as $\mathcal{O}(N)$, the learning takes the computational cost of $\mathcal{O}(N^2)$. [9] The most expensive part of the algorithm is the calculation of $\det J$, which takes the computational cost of $\mathcal{O}(N^3)$. However, we need to solve the flow equation (6.2.5) for a small number of teacher data in contrast to the number of data points for Monte Carlo simulation, and this make the total computational cost quite cheap.

6.5.2. Limitations

At first glance the deep learning method which we saw in the previous subsection seems to be of great use. However, this method loses the versatility of the Lefschetz thimble method in three points: First, to define the change of variables (6.5.13), we assume that the imaginary part of $z(t;x_0)$ is a function of the real part. However, it is in general not true and this may give a serious problem that $f(x;\lambda)$ do not reach to $t(x)$ even after learning. Although it can be neglected for small flow time, this may give the upper bound for the flow time. Second, the manifold (which they call "learnifold") where we change the variables from the real axis, contains error from original Lefschetz thimbles. The effective action reflects this error in first order, and this can recover the original sign problem. To solve sign problem, we must suppress the error δz to be $|\delta z| \leq \beta$, where β is the order of the action, but it is hard (at least for now) to suppress the error to such small one in the deep learning method.

[9] The number of layer and the number of nodes should correspond to the nonlinearity of $t(x)$. If the imaginary part of $t(x)$ is sensitive to the real part, we should take the algorithm deeper.

The third is that the deep learning method extrapolate $f(x;\lambda)$ from the region with a large number of teacher data to the region with few teacher data. However, there is no reason that $f(x;\lambda)$ represents $t(x)$ precisely, and to make matters worse, $S(f(x;\lambda))$ can have extra local minimum due to the lack of the precision. They also cause the original sign problem.

In this subsection, we see two examples. The deep learning method is applicable to one example, and inapplicable to the other example. The first example which we can apply the deep learning method, is the (0+1)-dimensional massive Thirring model at finite density, which we discussed in Sect. 6.4. The dynamical variable is periodic (especially compact) in this model, and we can apply the deep learning method while keeping the $U(1)^N$ symmetry explicitly by treating $(sin(x), cos(x))$ as an input. If we take $N = 8, \mu = 1.3, m = 1, g = 1/\sqrt{2}$ and apply the above deep learning method, the learnifold can represent Lefschetz thimbles well, and thus the numerical result precisely agree with the analytic solution.

The other example which we cannot apply the deep learning method, is the effective action of one-matrix model with complex coefficient, which we will discuss in the next chapter. If we consider $N = 2$ case for simplicity, the effective action can be given as

$$S(x_1, x_2; \beta, c) = \beta \sum_{i=1,2} (-\frac{c}{2} x_i^2 + \frac{1}{4} x_i^4) - \log((x_1 - x_2)^2). \tag{6.5.17}$$

where β is positive and $c = e^{i\theta}$ is complex. This model suffers from the sign problem when $\beta \gg 1$ and $c \neq \pm 1$. If we take $\beta = 20, c = e^{0.3\pi i}$ and apply the deep learning method, the learnifold can be obtained as in Fig. 6.7. When we generate the teacher data naively, there

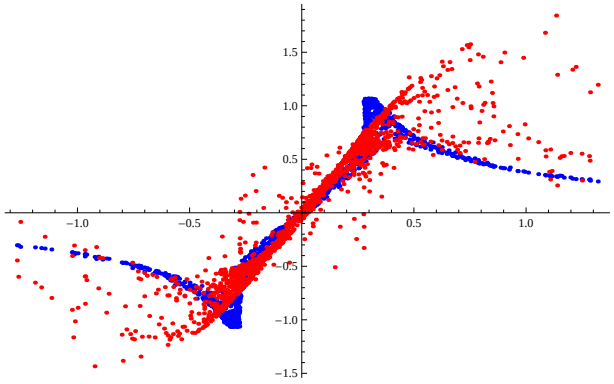


Figure 6.7: The learnifold which we obtain by the above deep learning method. We consider three hidden layers with ten nodes for each layer, and we prepare 10^4 teacher data points. We repeat the learning process 10^5 times while 10^2 teacher data points are chosen randomly as a mini batch. Unfortunately the learnifold fails to reconstruct the Lefschetz thimbles with enough precision.

are few data points in $|x_i| \geq 1$ and this causes the gap between $f(x;\lambda)$ and $t(x)$. Therefore the numerical result suffers from the sign problem.

Chapter 7

Complex one-matrix models

We are now ready to perform the Monte Carlo simulation for complex action. Although we are interested in triangle–hinge models in the case that ω is inserted in triangles, it should be helpful to see the case of matrix models with complex coefficients. In this chapter, we study the critical behavior of a simple one-matrix models

$$S(M) = N\beta \mathrm{tr} \left(-\frac{c}{2} M^2 + \frac{1}{4} M^4 \right), \tag{7.0.1}$$

where β is real positive and $c = e^{i\theta}$ is complex. This model returns to the ordinary one-matrix model with double-well potential (5.5.1) for $c = 1$, and with single-well potential for $c = -1$. This model suffers from the sign problem when $\beta \gg 1$ and $c \neq \pm 1$.

7.1. Three-cut solution

In the case $c = \pm 1$, the action (7.0.1) is real-valued and thus the classical solution is also real. The eigenvalues of the classical solution distribute around the minimum of the potential $V(x) = -c/2x^2 + 1/4x^4$ while feeling the Vandermonde repulsion. If $c = -1$ the potential is unimodal (i.e. $V(x)$ has a single minimum) and thus only the one-cut solution can be arrowed as the classical solution. If $c = +1$ the potential has two local minima, and the classical solution can be either one-cut solution or two-cut solution. As we discussed in Sect. 5.5, these solutions appear for $\beta < 4$ and $\beta > 4$ respectively, and the third-order phase transition occurs at $\beta = 4$.

The one-cut solution can be given for arbitrary c as

$$w(z) = \frac{\beta}{2}\left(-cz + z^3 - \sqrt{(z^2 + \frac{a^2}{2} - c)^2(z^2 - a^2)}\right) \tag{7.1.1}$$

$$a^2 \equiv \frac{2c}{3}\left(1 - \sqrt{1 + \frac{12}{\beta c^2}}\right), \tag{7.1.2}$$

and the two-cut solution as

$$w(z) = \frac{\beta}{2}\left(-cz + z^3 - \sqrt{z^2(z^2 - c + \frac{2}{\sqrt{\beta}})(z^2 - c - \frac{2}{\sqrt{\beta}})}\right). \tag{7.1.3}$$

However, in the case $c \neq \pm 1$, there can be another solution: The potential $V(x)$ has three saddle point in general, and thus the resolvent $w(z)$ can have cuts up to three. The three-cut solution can be written as

$$w(z) = \frac{\beta}{2}\left(-cz + z^3 - \sqrt{(z^2 - a_1^2)(z^2 - z_2^2)(z^2 - z_3^2)}\right), \tag{7.1.4}$$

where $a_1, a_2, a_3 \in \mathbb{C}$ are the endpoints of the cuts, and satisfy

$$a_1^2 + a_2^2 + a_3^2 = 2c \tag{7.1.5}$$

$$a_1^2 a_2^2 + a_2^2 a_3^2 + a_3^2 a_1^2 = c^2 - \frac{\beta}{4}. \tag{7.1.6}$$

Unlike the one-cut solution and the two-cut solution, we require additional property to determine the parameter $u \equiv a_1^2 a_2^2 a_3^2 \in \mathbb{C}$. According to the discussion given by David [55], the solution which actually appears is the one that minimizes (the real part of) the free energy while keeping the filling fraction n_α

$$n_\alpha \equiv \int_{A_\alpha} \frac{dz}{2\pi i} w(z) \tag{7.1.7}$$

to satisfy $0 \leq n_\alpha \leq 1$ (in particular, $n_\alpha \in \mathbb{R}$). Here α labels the cuts of the resolvent, and A_α is the integration contour around the cut I_α as in Fig. 7.1. The condition that minimize the real part of the free energy can be translated to the condition that the corresponding chemical potentials of cut I_α have the same real part. The difference of the chemical potential for the cut I_α and $I_{\alpha+1}$ can be given as the integral of the resolvent such as

$$\Delta\mu_\alpha \equiv \int_{B_\alpha} \frac{dz}{2\pi i} w(z), \tag{7.1.8}$$

where B_α is another integration contour across the cut as in Fig. 7.1. In summary, we can determine $u = u(\beta, c)$ by

$$\mathrm{Im}\left[\int_{A_\alpha} \frac{dz}{2\pi i} w(z)\right] = 0 \tag{7.1.9}$$

$$\mathrm{Re}\left[\int_{B_\alpha} \frac{dz}{2\pi i} w(z)\right] = 0. \tag{7.1.10}$$

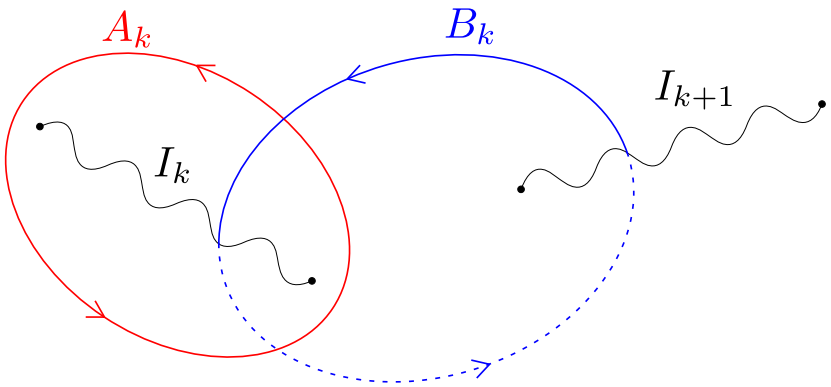


Figure 7.1: The integration contours A_α and B_α. A_α is around the cut I_α, and B_α pass through the cuts I_α and $I_{\alpha+1}$ onto another branch.

Since $V(x)$ is even, we take $0 \leq \mathrm{Re}\, a_1 \leq \mathrm{Re}\, a_2 \leq \mathrm{Re}\, a_3$ and the cut A_1, A_2 and A_3 as the line between $(-a_1, a_1), (a_2, a_3)$ and $(-a_3, -a_2)$ respectively, and then we obtain two independent condition

$$\mathrm{Im}\left[\int_{A_1} \frac{dz}{2\pi i} w(z)\right] = 0 \tag{7.1.11}$$

$$\mathrm{Re}\left[\int_{B_1} \frac{dz}{2\pi i} w(z)\right] = 0. \tag{7.1.12}$$

7.2. Numerical results

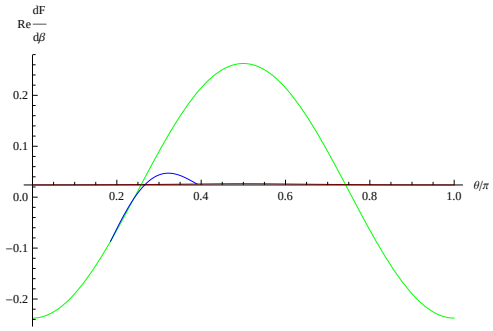
As we saw in the previous section, either one-cut or two-cut or three-cut solutions appears for given β and c. Although we can further know which of them appears by studying the stability of the solutions, we do not do this and only check numerically that the three-cut solution can actually appear.

Let us consider the case $\beta = 20$ for example. If $\theta = 0$ ($c = 1$), the two-cut solution appears. On the other hand, if $\theta = \pi$ ($c = -1$) the one-cut solution appears. There should be phase transition(s) in $0 < \theta < \pi$. However, $\frac{\partial F}{\partial \beta} = \langle V(M) \rangle$ is given as

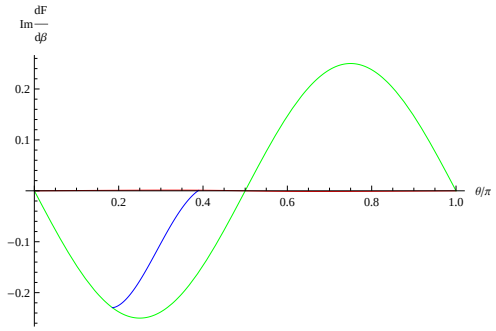
$$\langle V(M) \rangle = \begin{cases} \frac{\beta a^4}{1024}(9a^4 - 40ca^2 + 32c^2) & (\mathrm{one-cutsolution}) \\ -\frac{c^2}{4} + \frac{1}{4\beta} & (\mathrm{two-cutsolution}) \end{cases}, \tag{7.2.1}$$

respectively, and these real parts and imaginary parts cannot be continuous at the same time. This suggests that the three-cut solution should appear in some region.

We solve (7.1.11) and (7.1.12) for $u \in \mathbb{C}$, and obtain one-point function $\langle O \rangle = \langle V(M) \rangle$ in Fig. 7.2. Here, $n_1 = 1$ at $\theta \simeq 0.2\pi$ and $n_1 = 0$ at $\theta \simeq 0.4\pi$. This suggests that the three-cut solution appears in $0.2\pi < \theta < 0.4\pi$, and interpolates $\langle V(M) \rangle$ between that of the one-cut solution and that of the two-cut solution. The two-point function can be given in



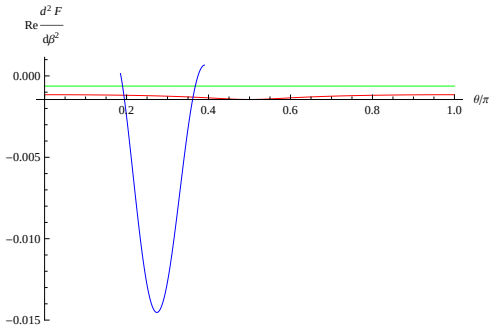
(a) The real part.



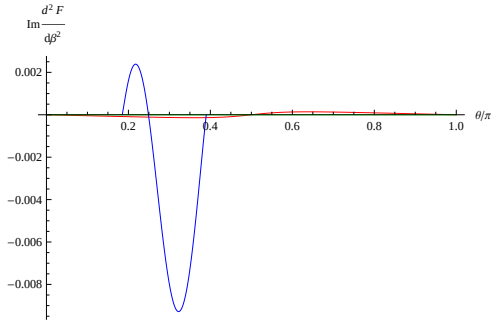
(b) The imaginary part.

Figure 7.2: The one-point functions for $\beta = 20$ in the large-N limit. The red (green, blue) line represents the result for one-cut (two-cut, three-cut) solution. The three-cut solution is given by solving (7.1.11) and (7.1.12) numerically.

Fig. 7.3. Since this is discontinuous at $\theta \simeq 0.2\pi$ and $\theta \simeq 0.4\pi$, the phase transitions should



(a) The real part.



(b) The imaginary part.

Figure 7.3: The two-point functions for $\beta = 20$ in the large-N limit. The red (green, blue) line represents the result for one-cut (two-cut, three-cut) solution.

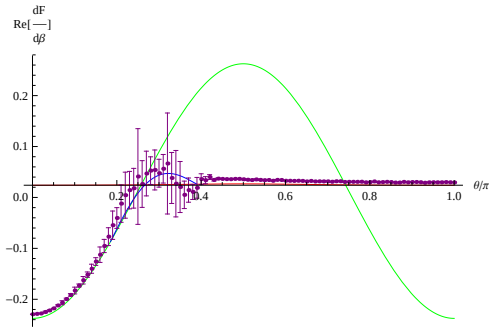
be second-order.

We perform Monte Carlo simulation for the action (7.0.1), by using the generalized Lefschetz thimble method.[1] We introduce the parallel tempering method which we discussed in Sect. 6.3. Although this model can be rewritten to the effective action of the eigenvalues, which has lower degrees of freedom and cheap to calculate, we purposely use the matrices as dynamical variables, looking toward the application to the triangle–hinge models, where we do not know such an effective action. [2] We set $N = 4$, $\beta = 20$ and the largest

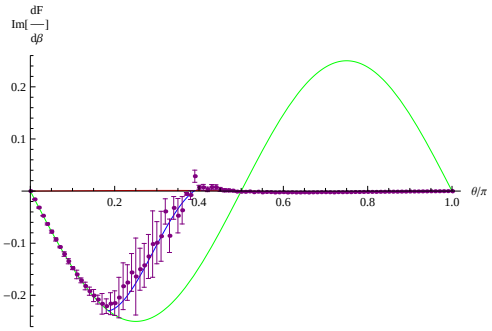
[1]Exactly speaking, we study the action (7.0.1) where M is $N \times N$ real symmetric matrix. Since the result is simply a half of that for Hermitian matrix in the large-N limit and we are only interested in the large-N limit, we doubled the obtained result.

[2]This action has $SU(N)$ continuous symmetry. Although the Lefschetz thimbles themselves are ill-defined (not manifestly N^2-dimensional submanifold), we should be able to apply the generalized Lefschetz thimble method to such a model because this method do not require the explicit properties of the thimbles.

flow time T =, and prepare flow times $\{t_\alpha\}$ ($\alpha = 0, 1, \cdots,$) with equal separations as $T_0 = 0, t_1 =, \cdots, t = T$. We calculate $\langle V(M) \rangle$ for $\theta = 0, 0.05\pi, \cdots \pi$. We take 10000 data points for each calculation, and discard the first 1000 data points for thermalization. We repeat Metropolis algorithm 20 times during the interval for the parallel tempering, and take a data point just before the tempering. We show the result in Fig. 7.4. This agree



(a) The real part.



(b) The imaginary part.

Figure 7.4: The numerical results of the one-point function for $N = 4$, $\beta = 20$. The red (green, blue) line represents the result for one-cut (two-cut, three-cut) solution in the large-N limit, and the purple dots are the numerical results.

with Fig. 7.2 and the three-cut solution appears in $0.2\pi < \theta < 0.4\pi$.

Chapter 8

Numerical analysis of triangle–hinge models with restriction to tetrahedral decompositions

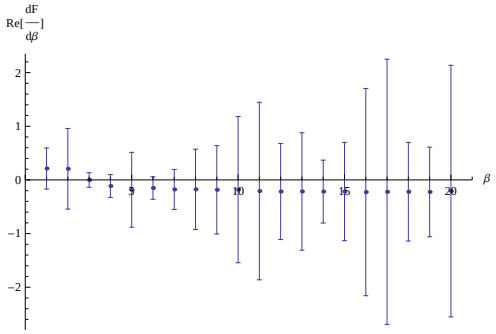
We are now ready to study the phase structure of the triangle–hinge model with the restriction to tetrahedral decompositions (5.5.11). In this chapter, we make a numerical study of the model (5.5.11).

We take $n = 3$ (that is, $m = 1$ in Sect. 3.7.1) and calculate $\frac{dF}{d\beta} = \langle \frac{1}{n^4\beta} S(A, B; \beta) \rangle$ for various $\beta = 1, 2, \cdots, 20$. We use the generalized Lefshcetz thimble algorithm, and take the flow time as $T = \frac{1}{n^2\beta}$.[1] We take $A = n^2 = 9$ and introduce the parallel tempering algorithm which we saw in Sect. 6.3 where we take the set of flow time as $\{t_\alpha\} = \{t_0 = 0, t_1 = T/A, \cdots, t_A = T\}$. We also introduce the parallel tempering algorithm for β-direction (i.e. we consider 200 replicas) and calculate various β simultaneously. We solve the flow equations (6.2.5) and (6.2.5) by using adaptive Runge-Kutta method, where we tune the stepsize of the time evolution so that the discretization errors of $z(t; x)_i$ and $J(t; x)_{ij}$ are less than 10^{-3} and $10^{-}1$ respectively. Here, we regard that the configuration reach to infinity when the stepsize becomes less than $10^{-3} \times t_\alpha$ in order to reduce the computational cost. We repeat the Metropolis process 20 times for each replica before we apply the tempering process for t-direction, and apply the tempering processes for t-direction and β-direction, and take a data point just after performing the tempering process for β-direction. Although we made a cold start in Chap. 5, we now make a hot start because the action (5.5.11) do not have a good classical solution on the real axes. We take the initial value for each replica by repeating the Metropolis process for small flow time and enlarge the flow time slowly. For the proposal distribution of the Metropolis process we use the uniform distribution, where

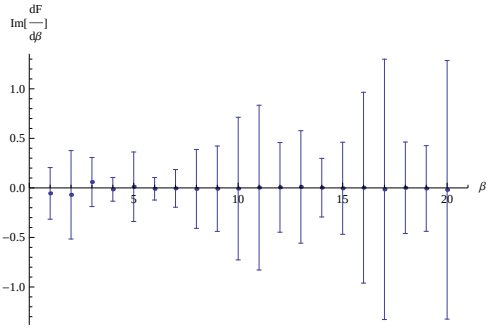
[1] In general, $\frac{dS}{dz}$ determines the typical scale of the flow time. In this case, $\frac{dS}{dz}$ is of order $\mathcal{O}(n^2\beta)$.

the width is chosen adaptively so that the acceptance rate is around 0.8.

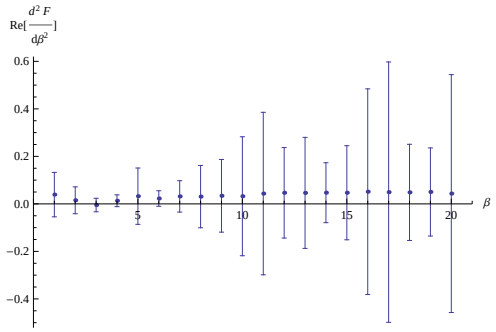
We first repeat the above sequence 100 times for some ϵs, and tune ϵs adaptively. We then take 10^4 data points by repeating the sequence, and discard the first 2×10^3 data points as initial sweeps. We show the result in Fig. 8.1. The one-point function and the two-point



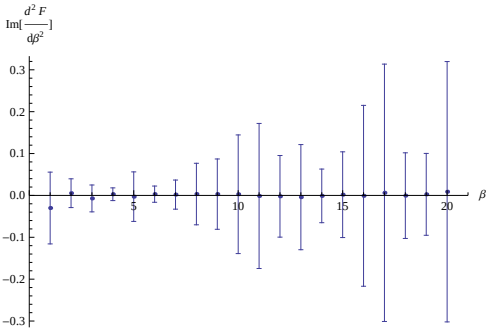
(a) The real part of one-point function.



(b) The imaginary part of one-point function.



(c) The real part of two-point function.



(d) The imaginary part of two-point function.

Figure 8.1: The numerical results for the triangle–hinge model (5.5.11).

function are real error ranges, and the one-point function is continuous. Unfortunately, the results have quite large statistical error and thus we cannot discuss the phase structure of the model. The average of the phase $\langle e^{-i\mathrm{Im}S+i\arg\det J}\rangle$ is shown in Fig. 8.2. This means that

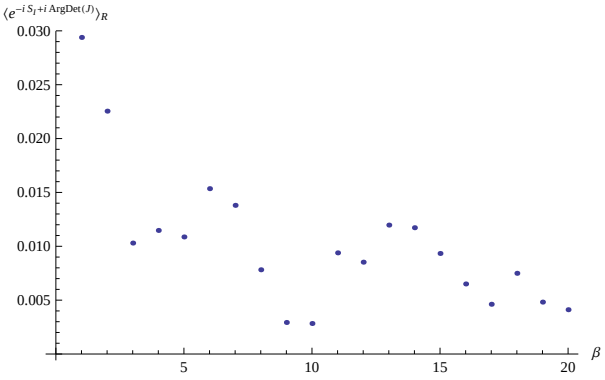


Figure 8.2: The numerical results of $|\langle e^{-i\mathrm{Im}S+i\arg\det J}\rangle|$. These take small values of $\mathcal{O}(10^{-2})$.

the large error is simply because of the sign problem, and that the flow time $T = \frac{1}{n^2\beta}$ is not sufficient to avoid the sign problem. Furthermore, of course, the result is only for small $n = 3$, and we need larger n results in order to discuss the phase transitions.

Although the original sign problem can be solved by taking large flow time, There are some obstacles: One is that another sign problem can then occur because the Jacobian takes quite large value in this case as we show in Fig. 8.3. Since the Jacobian term dominates

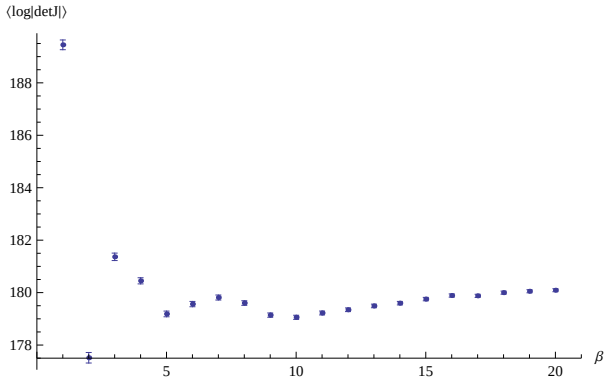


Figure 8.3: The numerical result of $\langle|\log\det J|\rangle$. These take quite large values of $\mathcal{O}(10^2)$, while the action S_R takes values of $\mathcal{O}(10^1)$.

in the effective action, the phase is no longer controllable even on the thimbles. Another obstacle is the computational cost. Our algorithm takes the computational cost of $\mathcal{O}(n^8)$, which is too expensive for larger n. To make the matter worse, We must solve the flow equations carefully due to reduce the discretization error of the Jacobian, and it makes the computational cost expensive. The next smallest n is $n = 6$, however, it spends a large amount of time unless we improve the algorithm.

Part III

Distance between configurations in MCMC simulations

Chapter 9

Distance between configurations in MCMC simulations

As we saw in Part II, Monte Carlo simulation is powerful tools. We usually use Markov chain Monte Carlo (MCMC) simulation, where we update configurations irrespective of their histories. However, in MCMC simulations we often encounter a situation where the equilibrium distribution is multimodal and the computation requires an extraordinarily long computer time to make the system reach global equilibrium. In order to speed up the relaxation to global equilibrium, one usually implements an additional method such as the overrelaxation [56] or the simulated/parallel tempering algorithm. However, it is not always possible to find a nice overrelaxation method. Also, the tempering method requires an adjustment of tempering parameter such that each local move has a finite amount of acceptance rate, which is usually done in a manual or adaptive way (and sometimes in an empirical way). Thus, it should be useful if there is a numerical measure that quantifies how much two configurations are separate for a given MCMC algorithm and how fast the system approaches global equilibrium. Such measure then can be used to numerically evaluate how the situation gets improved by the introduction of additional methods, which will make the above adjustment easier.

So far there have been proposed various methods for the above purpose, but they only quantify the separation between two different probability distributions (such as the L_1 distance) or measure the similarity between two data points for a given observable (such as the autocorrelation), and there has still not been a useful apparatus that directly measures an effective distance between two different configurations. For a given MCMC algorithm, we introduce a distance between two configurations that quantifies the difficulty of transition from one configuration to the other configuration in [5].

The introduction of such distance opens a way to investigate relaxation processes in an MCMC algorithm in terms of the geometry of the configuration space itself (which should

not be confused with the geometry of the set of probability distributions). As an example, we argue that, when the original distribution is highly multimodal with large number of degenerate vacua, our distance can be regarded as a geodesic distance with respect to an anti-de Sitter-like (AdS-like) metric. Such a geometrical viewpoint enables us to adjust the tempering parameter in a purely geometrical way; the adjustment can be carried out by requiring that the resulting geodesic distance be minimized. Thus this distance would be helpful for the reduction of computational cost for the parallel tempering algorithm which we discussed in Sect. 6.3.

9.1. Definition of distance

In this section, we define the distance between two configurations for a given MCMC algorithm. The distance will be written only with the kernel of a positive, symmetric matrix. We argue that the distance should take a universal form for the class of MCMC algorithms that generate local moves in the configuration space.

9.1.1. Preparation

Let $\mathcal{M} = \{x\}$ be a configuration space, and suppose that we are given an MCMC algorithm which generates a new configuration x from a configuration y with the conditional probability $P(x|y)$ at each step. We assume that this yields a stochastic process which has suitable ergodic properties such that $P_n(x|x_0) \equiv (P^n)(x|x_0)$ converges to a unique equilibrium distribution $p_{\rm eq}(x)$ in the limit $n \to \infty$, irrespectively of the initial value x_0. We further make two assumptions:

1. The algorithm satisfies the detailed balance condition for a given real-valued action $S(x)$,

$$P(x|y)\, e^{-S(y)} = P(y|x)\, e^{-S(x)}, \tag{9.1.1}$$

   which ensures that the equilibrium distribution is given by $p_{\rm eq}(x) = e^{-S(x)}/Z$ ($Z = \int dx\, e^{-S(x)}$).

2. The eigenvalues of P are all positive.[1]

[1]Note that the second condition is not too restrictive. In fact, when a transition matrix does not satisfy this condition (i.e., when some of the eigenvalues are negative), one can consider a new transition matrix $P_{\rm new} \equiv (P_{\rm old})^2$, for which the eigenvalues are all positive and the same equilibrium distribution is reached. Note also that this condition is always satisfied for the Langevin algorithm [see, e.g., (9.2.2)–(9.2.6) below].

By using the bra-ket notation $P(x|y) = \langle x|\hat{P}|y\rangle$ and the configuration operator $\hat{x} \equiv \int dx\, x\, |x\rangle\langle x|$, the above two conditions can be rephrased as a single statement that the operator $\hat{P}\, e^{-S(\hat{x})}$ is positive and symmetric. This also means that the "transfer matrix"

$$\hat{T} \equiv e^{S(\hat{x})/2}\, \hat{P}\, e^{-S(\hat{x})/2} = e^{S(\hat{x})/2} \left(\hat{P}\, e^{-S(\hat{x})}\right) e^{S(\hat{x})/2} \tag{9.1.2}$$

is positive and symmetric. $\hat{T}$ shares the same set of eigenvalues as $\hat{P}$, and according to our assumptions, all the eigenvalues are positive and the largest eigenvalue is unity with no degeneracy. Note that $P_n(x_1|x_2) = \langle x_1|\hat{P}^n|x_2\rangle$ can be expressed in the form

$$P_n(x_1|x_2) = K_n(x_1, x_2)\, e^{-(1/2)S(x_1)+(1/2)S(x_2)} \tag{9.1.3}$$

with the kernel of $\hat{T}$,

$$K_n(x_1, x_2) \equiv \langle x_1|\, \hat{T}^n|x_2\rangle. \tag{9.1.4}$$

If we introduce the spectral decomposition of $\hat{T}$ as[2]

$$\hat{T} = \sum_{k\geq 0} \lambda_k\, |k\rangle\langle k| = |0\rangle\langle 0| + \sum_{k\geq 1} \lambda_k\, |k\rangle\langle k| \tag{9.1.5}$$

with $\lambda_0 = 1 > \lambda_1 \geq \lambda_2 \geq \cdots > 0$, then the kernel is expressed as

$$K_n(x_1, x_2) = \sum_{k\geq 0} c_k(x_1, x_2)\, \lambda_k^n \tag{9.1.6}$$

with

$$c_k(x_1, x_2) \equiv \langle x_1|k\rangle\, \langle k|x_2\rangle. \tag{9.1.7}$$

The relaxation of the system to global equilibrium in the stochastic process corresponds to the decoupling of higher modes from $\hat{T}^n = |0\rangle\langle 0| + \sum_{k\geq 1} \lambda_k^n\, |k\rangle\langle k|$ as n increases. Thus, the large scale behavior of relaxation[3] is determined by the wave functions $\langle x|k\rangle$ with small eigenvalues λ_k. Note that the decoupling occurs earlier for modes with larger k.

If we further introduce the Hamiltonian $\hat{H}$ by using a small time increment ϵ as $\hat{T} = e^{-\epsilon\hat{H}}$, then $\hat{T}^n$ is expressed as $e^{-t\hat{H}}$ with $t \equiv n\epsilon$. For generic MCMC algorithms, $\hat{H}$ is a nonlocal operator acting on functions over $\mathcal{M}$. However, for the class of MCMC algorithms that generate only local moves of configuration, $\hat{H}$ should become local in the limit $\epsilon \to 0$ (see Sect. 9.1.4 for further arguments on locality).

We would like to introduce a distance $d_n(x_1|x_2)$ for a pair of configurations, x_1, $x_2 \in \mathcal{M}$, such that it enjoys the following properties in addition to the usual axioms of distance:

[2]The ground state wave function is given by $\langle x|0\rangle = e^{-(1/2)S(x)}/\sqrt{Z}$.

[3]By "large scale behavior (or structure)" we mean the behavior (or structure) both at large step numbers and at scales larger than the increment of configuration at each Monte Carlo step.

- **(P1)** The distance $d_n(x_1|x_2)$ vanishes in the limit $n \to \infty$ for any pairs x_1 and x_2, in order to reflect the fact that any configuration can be reached from every configuration in finite steps.
- **(P2)** If x_1 can be easily reached from x_2, then the distance is small even for finite n.
- **(P3)** If the distribution $e^{-S(x)}/Z$ is multimodal, and if x_1 and x_2 belong to different modes, then the distance is very large for finite n.

As we see below, such a distance can be naturally introduced if we look at transitions in $\mathcal{M}$ that is already in global equilibrium.

9.1.2. Half-time overlap

Let the system be in global equilibrium with probability distribution $p_{\rm eq}(x) = e^{-S(x)}/Z$. We denote by $\mathcal{X}_n$ the set of sequences of n processes in $\mathcal{M}$ and by $\mathcal{X}_n(x_1, x_2)$ the subset of $\mathcal{X}_n$ that consists of sequences which start from x_2 and end at x_1. We then define $f_n(x_1, x_2)$ to be the ratio of the sizes of two sets:

$$f_n(x_1, x_2) \equiv \frac{|\mathcal{X}_n(x_1, x_2)|}{|\mathcal{X}_n|}, \tag{9.1.8}$$

which can be expressed as

$$f_n(x_1, x_2) = P_n(x_1|x_2)\,\frac{1}{Z}\,e^{-S(x_2)} = K_n(x_1, x_2)\,\frac{1}{Z}\,e^{-(1/2)S(x_1)-(1/2)S(x_2)}. \tag{9.1.9}$$

The latter expression shows that $f_n(x_1, x_2)$ is a symmetric function of x_1 and x_2. $f_n(x_1, x_2)$ gives the probability to find a sequence of n processes from x_2 to x_1 (or from x_1 to x_2) out of all the possible n processes, and thus expresses "mobility" between two configurations x_1 and x_2. Note that $f_n(x_1, x_2)$ should take a very small value if the equilibrium distribution is multimodal and x_1 and x_2 belong to different modes.

Since our interest is in the mobility between two different configurations, we normalize the mobility such that it takes the maximal value (= 1) for closed loops which start from and end at the same configuration:

$$F_n(x_1, x_2) \equiv \frac{f_n(x_1, x_2)}{\sqrt{f_n(x_1, x_1) f_n(x_2, x_2)}} = \sqrt{\frac{P_n(x_1|x_2)\,P_n(x_2|x_1)}{P_n(x_1|x_1)\,P_n(x_2|x_2)}}. \tag{9.1.10}$$

We will call the normalized mobility $F_n(x_1, x_2)$ the *half-time overlap* of configurations x_1 and x_2 for n steps. In fact, by using (9.1.9) and introducing the "half-time elapsed state" $|x, n/2\rangle \equiv \hat{T}^{n/2}\,|x\rangle$, the function $F_n(x_1, x_2)$ can actually be expressed as the overlap of two normalized half-time elapsed states:

$$F_n(x_1, x_2) = \frac{K_n(x_1, x_2)}{\sqrt{K_n(x_1, x_1)\,K_n(x_2, x_2)}} = \frac{\langle x_1, n/2\,|\,x_2, n/2\rangle}{||\,|x_1, n/2\rangle\,||\,||\,|x_2, n/2\rangle\,||}. \tag{9.1.11}$$

One can easily prove that $F_n(x_1, x_2)$ enjoys the following properties:

- $F_n(x_1, x_2) = F_n(x_2, x_1)$, (9.1.12)
- $0 \leq F_n(x_1, x_2) \leq 1$, (9.1.13)
- $F_n(x, x) = 1$, (9.1.14)
- $\lim_{n\to\infty} F_n(x_1, x_2) = 1$. (9.1.15)

Furthermore, from the properties of $f_n(x_1, x_2)$, we expect that

- $F_n(x_1, x_2) \simeq 1$ when x_1 can be easily reached from x_2 in n steps. (9.1.16)
- $F_n(x_1, x_2) \ll 1$ when x_1 and x_2 are separated by high potential barriers. (9.1.17)

9.1.3. Definition of distance

Based on the half-time overlap given above, we define a distance between x_1, $x_2 \in \mathcal{M}$ as follows:

$$\theta_n(x_1, x_2) \equiv \arccos(F_n(x_1, x_2)). \tag{9.1.18}$$

This should satisfy the properties (P1)–(P3) due to (9.1.15)–(9.1.17), as well as the following axioms of distance:[4]

- $\theta_n(x_1, x_2) \geq 0$, (9.1.19)
- $x_1 = x_2 \;\Leftrightarrow\; \theta_n(x_1, x_2) = 0$, (9.1.20)
- $\theta_n(x_1, x_2) = \theta_n(x_2, x_1)$, (9.1.21)
- $\theta_n(x_1, x_2) + \theta_n(x_2, x_3) \geq \theta_n(x_1, x_3)$. (9.1.22)

It is often convenient to introduce other distances from the half-time overlap:[5]

$$d_n^2(x_1, x_2) \equiv -2\log(F_n(x_1, x_2)), \tag{9.1.23}$$
$$D_n^2(x_1, x_2) \equiv 2(1 - F_n(x_1, x_2)). \tag{9.1.24}$$

Although these distances do not satisfy the triangle inequality, they agree with $\theta_n(x_1, x_2)$ up to higher order corrections when $\theta_n(x_1, x_2)$ is small enough. For the rest of this paper we will mainly use $d_n(x_1, x_2)$ as the definition of distance, because it gives the simplest expressions for the following examples.

[4]The axiom of nondegeneracy (i.e., $\theta_n(x_1, x_2) = 0 \;\Rightarrow\; x_1 = x_2$) holds for all the algorithms that involve relaxations. Note that the algorithm consisting of the global updates based on the global heat bath is not in this class, because the system transits to the equilibrium state at the first Monte Carlo step (the transfer matrix is given by the projection, $T = |0\rangle\langle 0|$). In fact, for this case, the distance $\theta_n(x_1, x_2)$ vanishes for any pairs x_1 and x_2, which is also consistent with our definition of distance.

[5]Note the similarity of the definition of distance between configurations to that between states in quantum information. There, $F_n(x_1, x_2)$ corresponds to the fidelity of two pure states $\rho_{1,2} = |x_{1,2}\rangle\langle x_{1,2}|/||\,|x_{1,2}\rangle\,||^2$, and the distances $\theta_n(x_1, x_2)$ and $D_n(x_1, x_2)$ to Bures length and Bures distance, respectively.

9.1.4. Universality of distance

We close this section with a comment on the universality of our distance. As long as the chosen algorithm generates only local moves of configuration, we expect that the large scale structure of distance $d_n(x_1, x_2)$ takes a universal form, in the sense that differences of distance between two such algorithms can always be absorbed into a rescaling of n. In fact, our distance is totally expressed with the kernel of the transfer matrix $\hat{T}$, and, when the transition is sufficiently local, the corresponding Hamiltonian $\hat{H}$ is a local operator acting on functions over $\mathcal{M}$ in almost the same way. We thus expect that the eigenvalues λ_k and the wave functions $\langle x|k\rangle$ are almost the same for small k's, which ensures the same large scale structure of the kernel $K_n(x_1, x_2)$ and thus that of the distance $d_n(x_1, x_2)$. This statement is explicitly checked in Appendix for the Langevin and Metropolis algorithms using a simple one-dimensional model. Note that the argument for universality are more trustworthy when the degrees of freedom of system become larger.

This universality will not hold for algorithms that generate nonlocal moves of configuration (such as the overrelaxation algorithm).[6] The Hybrid Monte Carlo (HMC) algorithm [57] is marginal in this sense, and we leave it for future study to investigate whether the distance for the HMC algorithm exhibits the same behavior as local algorithms.

9.2. Examples

Expecting the universality of distance commented in the previous section, we consider the Langevin method as an MCMC algorithm, and write down the explicit form of distance for a configuration space $\mathcal{M} = \mathbb{R}$. We show that the resulting distance actually satisfies the properties (P1)–(P3).

Let ν_t be the Gaussian white noise with diffusion coefficient D, $\langle \nu_t \, \nu_{t'} \rangle_\nu = 2D \, \delta(t - t')$, and $x_t = x_t(x_0, [\nu])$ be the solution to the Langevin equation

$$\dot{x}_t = \nu_t - D \, S'(x_t), \qquad x_t|_{t=0} = x_0. \tag{9.2.1}$$

Then, the probability distribution[7] $P_t(x|x_0) \equiv \langle \delta(x - x_t(x_0, [\nu])) \rangle_\nu$ is expressed as

$$P_t(x|x_0) = \langle x|e^{-t\hat{H}_{\rm FP}}|x_0\rangle \tag{9.2.2}$$

[6] The tempering algorithms are nonlocal when viewed from the original configuration space $\mathcal{M} = \{x\}$, but they actually generate local moves in the extended configuration space, so that the universality of the distance is expected to hold also for these algorithms. See arguments below (9.3.3) in Sect. 9.3.2 for details.

[7] In this section we exclusively use a continuous notation; namely, the distribution $P_n(x|x_0)$ will be written as $P_t(x|x_0)$ with $t = n\epsilon$, where ϵ is the time increment.

with the Fokker-Planck Hamiltonian (we will set $D = 1$ below):

$$\hat{H}_{\rm FP} = -D\,\frac{\partial}{\partial x}\Big[\frac{\partial}{\partial x} + S'(\hat{x})\Big]. \tag{9.2.3}$$

The transfer matrix is expressed as

$$\hat{T} = e^{-\epsilon \hat{H}}, \tag{9.2.4}$$

where the positive, symmetric Hamiltonian $\hat{H}$ is given by

$$\begin{aligned}\hat{H} &= e^{S(\hat{x})/2}\,\hat{H}_{\rm FP}\,e^{-S(\hat{x})/2} \\ &= \Big[-\frac{\partial}{\partial x} + \frac{1}{2}\,S'(\hat{x})\Big]\Big[\frac{\partial}{\partial x} + \frac{1}{2}\,S'(\hat{x})\Big] \\ &= -\frac{\partial^2}{\partial x^2} + V(\hat{x})\end{aligned} \tag{9.2.5}$$

with

$$V(x) \equiv \frac{1}{4}\,[S'(x)]^2 - \frac{1}{2}\,S''(x). \tag{9.2.6}$$

The corresponding kernel is then given by

$$K_t(x, x_0) = \langle x|e^{-t\hat{H}}|x_0\rangle \tag{9.2.7}$$

with which the half-time overlap is expressed as

$$F_t(x_1, x_2) = \frac{K_t(x_1, x_2)}{\sqrt{K_t(x_1, x_1)\,K_t(x_2, x_2)}}. \tag{9.2.8}$$

9.2.1. Example 1: Gaussian

We first consider the action

$$S(x) = \frac{\omega}{2}\,x^2, \tag{9.2.9}$$

for which the Hamiltonian $\hat{H}$ takes the form

$$\hat{H} = -\frac{\partial^2}{\partial x^2} + \frac{\omega^2}{4}\,\hat{x}^2 - \frac{\omega}{2}. \tag{9.2.10}$$

Note that the last term removes the zero-point energy of this harmonic oscillator. By using the kernel

$$K_t(x_1, x_2) = \sqrt{\frac{\omega}{2\pi\,(1 - e^{-2\omega t})}}\,\exp\Big[-\frac{\omega}{4\sinh\omega t}\big[(x_1^2 + x_2^2)\,\cosh\omega t - 2x_1x_2\big]\Big], \tag{9.2.11}$$

the distance $d_t(x_1, x_2)$ is easily obtained to be

$$d_t^2(x_1, x_2) = \frac{\omega}{2\sinh(\omega t)} |x_1 - x_2|^2. \tag{9.2.12}$$

This gives a flat and translation invariant metric in the entire configuration space $\mathcal{M} = \mathbb{R}$.[8] Note that the distance decreases exponentially as $d_t^2 \sim e^{-\omega t}$, from which we find that the relaxation time of the system is given by $\sim 1/\omega$.[9] Since the manner of relaxation is almost the same for unimodal distributions, we see that the distance rapidly decreases to zero when the action gives a unimodal distribution. We thus confirm that the properties (P1) and (P2) hold in this example.

9.2.2. Example 2: perturbation around the Gaussian

We now consider the case where the action (9.2.9) is perturbed with a quartic term:

$$S(x) = \frac{\omega}{2} x^2 + \frac{\lambda}{4} x^4. \tag{9.2.13}$$

The perturbative calculation of distance can be done easily, and we find to the first order perturbation:

$$\begin{aligned} d_t^2(x_1, x_2) = |x_1 - x_2|^2 \Big\{ \frac{\omega}{2s} - \frac{\lambda}{8\omega s^4} \Big[& 12(s^3 - 3s^2 c + 3\omega t s + 2\omega t s^3 - \omega t s^2 c) \\ & + \omega\,(s^3 + 3s - 3\omega t c)(x_1 - x_2)^2 \\ & + 3\omega\,(s^3 + 3s - 3\omega t c + 3\omega t - 3sc + 2\omega t s^2)(x_1 + x_2)^2 \Big] + O(\lambda^2) \Big\}, \end{aligned} \tag{9.2.14}$$

where $s \equiv \sinh(\omega t)$ and $c \equiv \cosh(\omega t)$. This shows that the distance generically does not take a flat or translation invariant form when the unimodal distribution is not Gaussian.

9.2.3. Example 3: double-well potential

Finally, in order to confirm the property (P3), we consider the case where a high potential barrier exists between local minima:

$$S(x) = \frac{\beta}{2} (x^2 - 1)^2. \tag{9.2.15}$$

[8] Our discussion can be easily generalized to the case $\mathcal{M} = \mathbb{R}^N = \{\mathbf{x} = (x^i)\}$ and $S(\mathbf{x}) = \sum_i \omega_i (x^i)^2$. The distance is then given by $d_t^2(\mathbf{x}_1, \mathbf{x}_2) = \sum_i (\omega_i / 2\sinh(\omega_i t))\, |x_1^i - x_2^i|^2$.

[9] If we take the limit $\omega \to 0$ (corresponding to the pure Brownian motion), the distance is given by $d_t^2(x_1, x_2) = (1/2t)\, |x_1 - x_2|^2$, and thus the exponential damping disappears. This reflects the fact that the relaxation time becomes infinite in the limit $\omega \to 0$ in the sense that there is no normalizable equilibrium distribution for the pure Brownian motion.

We assume that β takes a large value so that the equilibrium distribution $e^{-S(x)}/Z$ is multimodal. The potential $V(x) = (1/4)[S'(x)]^2 - (1/2)S''(x)$ becomes sextic (see Fig. 9.1),

$$V(x) = \beta^2 x^6 - 2\beta^2 x^4 + (\beta^2 - 3\beta)\, x^2 + \beta, \tag{9.2.16}$$

and has local minima at $x = 0$ and $x = \pm\, x_+$ with $x_+ \equiv \left[\left(2 + \sqrt{1 + 9/\beta}\right)/3\right]^{1/2} \simeq 1$. One

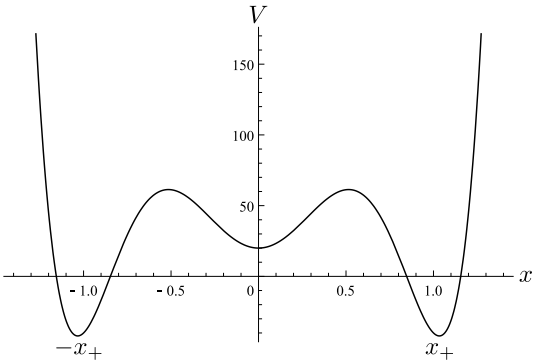


Figure 9.1: The sextic potential $V(x)$ [eq. (9.2.16)] with $\beta = 20$, which has two global minimum at $x = \pm\, x_+$ and a local minimum at $x = 0$ [5].

can easily find that $x = \pm\, x_+$ are global minima by looking at the values of $S''(x)$ (note that $S'(x)$ vanishes there).

The eigenvalues E_k ($E_0 = 0 < E_1 < E_2 \cdots$) of $\hat{H}$ can be roughly estimated as follows. We first note that the Gaussian approximation around the global minima should be effective when $\beta \gg 1$, and thus the first two eigenstates of $\hat{H}$ can be approximated as superpositions of the ground states $|0_+\rangle$ and $|0_-\rangle$ of the approximated Gaussian potential $V(x) \simeq (1/2)\, V''(\pm x_+)\, (x \mp x_+)^2 \simeq 4\beta^2\, (x \mp 1)^2$ around the right ($x = x_+ \simeq +1$) and left ($x = -x_+ \simeq -1$) minima:

$$|0\rangle \simeq \frac{1}{\sqrt{2}}\big(|0_+\rangle + |0_-\rangle\big), \tag{9.2.17}$$

$$|1\rangle \simeq \frac{1}{\sqrt{2}}\big(|0_+\rangle - |0_-\rangle\big). \tag{9.2.18}$$

The energy difference between two states is exponentially small, $E_1 = O(e^{-\beta/2})$, as can be estimated by an instanton calculation.

As for the second excitation of $\hat{H}$, we expect the relations $E_0 = 0 \lesssim E_1 = O(e^{-\beta/2}) \ll E_2 = O(\beta)$. In fact, there are two possible approximations for the second excitation; one is to represent the second excitation as a superposition of the first excited states of the approximated Gaussian potentials around the global minima $x = \pm\, x_+$, and the other is to represent it as the ground state of the approximated Gaussian potential around the local minimum $x = 0$. In our case, the latter approximation is applicable, because the former gives $E_2 \simeq 4\beta$ and the latter gives $E_2 \simeq 2\beta$.

By using the expansion of the kernel $K_t(x_1, x_2) = \langle x_1|e^{-tH}|x_2\rangle$ [see (9.1.6)],

$$K_t(x_1, x_2) = \sum_{k\geq 0} c_k(x_1, x_2)\, e^{-E_k t} \quad \big(c_k(x_1, x_2) = \langle x_1|k\rangle\, \langle k|x_2\rangle\big), \tag{9.2.19}$$

the distance $d_t(x_1, x_2)$ is expressed as

$$\begin{aligned} d_t^2(x_1, x_2) = \sum_{k\geq 1} \frac{(-1)^{k-1}}{k} \Big[\Big(\frac{c_1(x_1, x_1)}{c_0(x_1, x_1)}\Big)^k + \Big(\frac{c_1(x_2, x_2)}{c_0(x_2, x_2)}\Big)^k - 2\Big(\frac{c_1(x_1, x_2)}{c_0(x_1, x_2)}\Big)^k \Big] e^{-kE_1 t} \\ + \Big[\frac{c_2(x_1, x_1)}{c_0(x_1, x_1)} + \frac{c_2(x_2, x_2)}{c_0(x_2, x_2)} - 2\frac{c_2(x_1, x_2)}{c_0(x_1, x_2)} \Big] e^{-E_2 t} + \cdots . \end{aligned} \tag{9.2.20}$$

We now consider the following two cases:

$$\begin{aligned} &\text{Case 1} : x_1 \simeq x_2 \simeq 1, \\ &\text{Case 2} : x_1 \simeq -x_2 \simeq 1. \end{aligned}$$

We first discuss Case 1 where x_1 and x_2 belong to the same mode. If we expand $d_t^2(x_1, x_2 = x_1 + \delta x)$ to the second order of δx, the coefficient of $e^{-kE_1 t}$ is given by

$$\begin{aligned} &\Big(\frac{c_1(x_1, x_1)}{c_0(x_1, x_1)}\Big)^k + \Big(\frac{c_1(x_2, x_2)}{c_0(x_2, x_2)}\Big)^k - 2\Big(\frac{c_1(x_1, x_2)}{c_0(x_1, x_2)}\Big)^k \\ &\simeq 4k^2\delta x^2 \left(\frac{\langle x_1|0_+\rangle - \langle x_1|0_-\rangle}{\langle x_1|0_+\rangle + \langle x_1|0_-\rangle}\right)^{2k} \left(\frac{\langle x_1|0_+\rangle\, \langle x_1|0_-\rangle' - \langle x_1|0_+\rangle'\, \langle x_1|0_-\rangle}{\langle x_1|0_+\rangle^2 - \langle x_1|0_-\rangle^2}\right)^2, \end{aligned} \tag{9.2.21}$$

which is vanishingly small as can be understood by considering the supports of the wave functions $\langle x|0_\pm\rangle$. Thus, the dominant contribution to the distance comes from the second term in (9.2.20), and the two configurations x_1 and x_2 can be reached from each other with a rather short time $\sim 1/E_2 = O(1/\beta)$. This confirms that two configurations are close to each other even for a multimodal distribution *if* they belong to the same mode. In contrast, as for Case 2 where x_1 and x_2 belong to different modes, the coefficient of $e^{-E_1 t}$ is not small, so that the dominant contribution to the distance comes from the first excited state, and the transition between x_1 and x_2 requires an exponentially long time $\sim 1/E_1 = O(e^{\beta/2})$. Accordingly, the distance $d_t(x_1, x_2)$ between $x_1 \simeq +1$ and $x_2 \simeq -1$ has a large value $d_t^2(x_1, x_2) \propto \beta$,[10] which decreases only very slowly as t elapses. We thus see that the property (P3) certainly holds in this example.

9.3. Distance for the simulated tempering and the emergence of AdS-like geometry

In this section, we show that the value of distance $d_n(x_1, x_2)$ for a multimodal probability distribution gets dramatically reduced when one introduces a tempering algorithm. We

[10] $F_t(x_1, x_2) = e^{-(1/2)\, d_t^2(x_1, x_2)}$ for large β can be estimated to be $e^{-\text{const.}\beta}$ by using an instanton analysis.

further argue that the effective distance defined for the extended configuration space gives an AdS-like geometry when the original distribution is highly multimodal with large number of degenerate vacua.

9.3.1. Distance for the simulated tempering

Suppose that we are considering a multimodal distribution, and that we implement a simulated tempering method [49]. We take as the tempering parameter the overall coefficient β of the action and introduce the parameter set $\mathcal{A} \equiv \{\beta_a\}_{a=0,1,\ldots,A}$ such that $\beta_0 > \beta_1 > \cdots > \beta_A$, where β_0 is the overall coefficient of the original action. We denote by $S(x;\beta_a)$ the action with β_0 replaced by β_a.

Since the distribution with smaller β_a is less multimodal, two configurations belonging to different modes at β_0 can now be easily reached from each other by moving around in the extended configuration space, as long as moves in the β direction occur frequently (as we assume here). This improves the convergence to global equilibrium, and accordingly, the distance between two configurations $X_1 = (x_1, \beta_0)$ and $X_2 = (x_2, \beta_0)$ will be reduced. To demonstrate that this actually happens, we calculated the distance between $X_1 = (x_1 = +1, \beta_0)$ and $X_2 = (x_2 = -1, \beta_0)$ for the action $S(x;\beta_0) = (\beta_0/2)\,(x^2-1)^2$ with $\beta_0 = 20$, by using the simplest setting $A = 1$ and $\beta_1 = 1$. As for Step 1 above, we adopted the Metropolis algorithm using Gaussian proposal distribution with variance $\sigma^2 = 0.01$, where the configuration space is restricted to the interval $[-3, 3]$ and is latticized with cutoff $a = 0.001$. Furthermore, by denoting the transition matrix for Step s by $\hat{P}_{(s)}$ $(s = 1, 2)$, we set the transition matrix $\hat{P}$ in the simulated tempering algorithm to $\hat{P} = \hat{P}_{(1)}\hat{P}_{(2)}\hat{P}_{(1)}$ so that the combined transition matrix satisfies the detailed balance condition. Below is the result we obtained:[11]

n	$d_n^2(X_1, X_2)$ (without tempering)	$d_n^2(X_1, X_2)$ (with tempering)
10	39.1	26.5
50	19.2	7.16
100	16.9	4.35
500	13.2	0.708
1,000	11.7	0.106
5,000	8.46	2.78×10^{-8}

We clearly see that the introduction of the simulated tempering method dramatically reduces the distance for such a multimodal distribution.

[11] In order to align the scale of n, the transition matrix $\hat{P} = \hat{P}_{(1)}^2$ is used for the original local algorithm without tempering (central column in the table).

9.3.2. AdS-like geometry of the extended configuration space

When a system has very large degrees of freedom (as in the case of the large-volume or low-temperature limit), the issue related to the multimodality of probability distribution becomes serious. However, it will then become a good approximation to coarse-grain the elements of configuration space $\mathcal{M}$ by regarding configurations in the same mode as a single configuration, and to introduce a new configuration space $\bar{\mathcal{M}}$ that consists of coarse-grained configurations which are separate from each other. We assume that the way of separation between two neighboring configurations is uniform over $\bar{\mathcal{M}}$, keeping in mind a situation where the original distribution has highly degenerate vacua.

Suppose that we introduce the simulated tempering algorithm to such system. The original configuration space $\mathcal{M}$ is then extended to $\mathcal{M} \times \mathcal{A}$, as discussed in the previous subsection. When the degrees of freedom are very large, the spacing in the parameter set $\mathcal{A} = \{\beta_a\}$ must be sufficiently small such that two adjacent configurations (x, β_a) and (x, β_{a+1}) has an acceptance rate of $O(1)$. Then, we may (and we will) regard $\mathcal{A}$ as a continuous set, $\mathcal{A} = \{\beta \,|\, \beta_A \le \beta \le \beta_0\}$.

Now let us consider the distance in the extended, coarse-grained configuration space: $\bar{\mathcal{M}} \times \mathcal{A} = \{X = (x, \beta)\}$. We define the metric $ds^2 = g^{(n)}_{MN}(X)\, dX^M dX^N$ $[(X^M) = (x, \beta)]$ such that its geodesic distance between two arbitrarily chosen points $X_1 = (x_1, \beta_1)$ and $X_2 = (x_2, \beta_2)$ gives our distance $d_n(X_1, X_2)$. Such metric will take the following form due

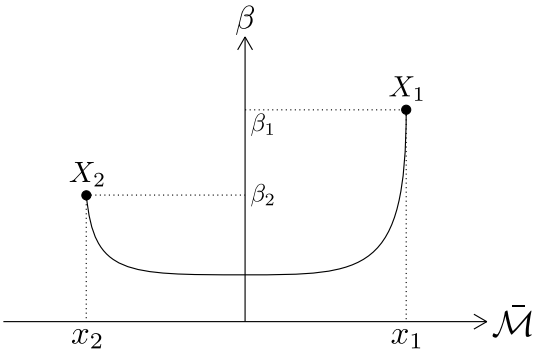


Figure 9.2: The geodesic line for the geometry (9.3.1) with two ends at $X_1 = (x_1, \beta_1)$ and $X_2 = (x_2, \beta_2)$ [5].

to the uniformity over $\bar{\mathcal{M}}$:[12]

$$ds^2 \simeq f(\beta)\, d\beta^2 + g(\beta)\, dx^2. \tag{9.3.1}$$

The functions $f(\beta)$ and $g(\beta)$ will be studied in detail in the subsequent paper [58], and we here make a simple analysis of their properties. Since two different configurations (x_1, β)

[12]We have assumed that the coefficient of $d\beta\, dx$ is relatively small and can be neglected.

and (x_2, β) in $\bar{\mathcal{M}} \times \mathcal{A}$ belong to two different modes of the distribution $e^{-S(x;\beta)}/Z(\beta)$, the transition between them becomes more difficult as β increases. Thus, $g(\beta)$ should be an increasing function at least when β is large, and we assume that it takes an asymptotic form $g(\beta) \propto \beta^q$ for large β with some positive number q.[13] On the other hand, since the transition in the β direction has no obstacle, we expect that $f(\beta)$ has a simple behavior. Note that, when the measure in the β direction is scale invariant [i.e., when $f(\beta) \propto 1/\beta^2$] for large β, the above metric has the following asymptotic form:

$$ds^2 \simeq \text{const.} \frac{\mathrm{d}\beta^2}{\beta^2} + \text{const.}\, \beta^q\, dx^2 \quad (\beta\text{: large}), \tag{9.3.2}$$

which is nothing but the Euclidean AdS metric, as can be easily seen by the redefinition of variable, $z \equiv \text{const.}\, \beta^{-q/2}$:

$$ds^2 \propto \frac{1}{z^2}\,(dz^2 + dx^2) \quad (z\text{: small}). \tag{9.3.3}$$

We here argue that the appearance of AdS-like geometry [eq. (9.3.1) with $g(\beta) \sim \beta^q$] should be universal. In fact, as discussed above, when implementing the tempering method to an MCMC simulation for a large system, the spacing in the parameter set $\mathcal{A} = \{\beta_a\}$ must be sufficiently small, which allows us to regard $\mathcal{A}$ as a continuous set. Then, the stochastic processes in the x and β directions given in Sect. 9.3.1 can be regarded as local moves in the extended configuration space $\mathcal{M} \times \mathcal{A}$, which in turn define local moves in the extended, coarse-grained configuration space $\bar{\mathcal{M}} \times \mathcal{A}$. Thus, combining with arguments in Sect. 9.1.4, we expect that the distance takes a universal form for $\bar{\mathcal{M}} \times \mathcal{A}$ in the sense that it does not depend on the particular MCMC algorithms that generate local moves along the x and β directions.[14] This is why we expect that the emergence of AdS-like geometry is universal.

[13] Without a tempering method, the distance between two configurations belonging to different modes for fixed large β can be roughly estimated by an instanton analysis to be proportional to β [see discussions below (9.2.21)]. This implies that the function $g(\beta)$ increases more slowly than β when the simulated tempering is implemented, because configurations at larger β will get more benefit from the tempering method. We thus expect that q satisfies the inequality $0 < q \leq 1$. Our on-going work with numerical calculation [58] shows that q is actually in this range, excluding a logarithmic growth of $g(\beta)$.

[14] However, the distance should depend on the way of preparing the parameter set $\{\beta_a\}$ even though the set $\mathcal{A}$ can be regarded as almost continuous.

Chapter 10

Conclusion

We proposed the triangle–hinge models, which is a class of models that generate random diagrams consisting of triangles and multiple hinges. The models are completely characterized by semisimple associative algebras $\mathcal{A}$ and tensors C^{ijklmn}. When C^{ijklmn} are chosen as in (3.3.1) or (3.7.1), each Feynman diagram can be expressed as a collection of index networks around vertices. The contribution $\mathcal{F}(\gamma)$ from each diagram γ to the free energy is expressed as the product of the index functions $\zeta(v)$ of vertices v of γ, and $\zeta(v)$ depends only on the topology of the index network around v besides the structure of the defining associative algebra. Although most of the Feynman diagrams do not represent three-dimensional manifolds, we give a general prescription to automatically reduce the set of possible diagrams such that only (and all of the) manifolds are generated. We implement the strategy for the models with $\mathcal{A}$ set to matrix rings, by introducing a color structure and taking the direct sum of K copies of matrix ring. We show that every diagram actually gives a tetrahedral decomposition where each vertex has a neighborhood homeomorphic to B^3 (ensured by the statement that the index network around each vertex has the topology of S^2). We further demonstrate that there is a novel strong-weak duality in the models which interchanges the roles of triangles and hinges. We also investigate the models where the defining associative algebras are group rings, and show that most of their analytic properties can be understood as a straightforward generalization of those for matrix rings.

We also gave a general prescription to introduce matter degrees of freedom to triangle–hinge models. This is achieved by setting the defining associative algebra $\mathcal{A}$ to be a tensor product of the form $\mathcal{A}_{\mathrm{grav}} \otimes \mathcal{A}_{\mathrm{mat}}$ and by modifying the interaction terms appropriately. The matter fields thus obtained have local interactions, since a colored tetrahedron can only interact with the neighbors. We can assign colors not only to tetrahedra but also to simplices of arbitrary dimensions. We further show that there exists a duality between matter fields on a tetrahedral lattice and those on its dual lattice, which can be realized by applying the duality transformation which interchanges the roles of triangles and hinges. When we take

the set of colors to be $\mathbb{R}^D$ and assign colors to tetrahedra as in (4.1.2), the matter fields represent the target space coordinates of membranes in D dimensions. By taking different sets of colors, we can also construct various spin systems on random volumes, including the Ising model, the q-state Potts models and the RSOS models. A wider class of models can be further obtained as triangle–hinge models by coloring lower dimensional simplices as well as tetrahedra. For example, three-dimensional colored tensor models can be obtained by coloring tetrahedra, triangles and edges at a time. This is shown in Sect. 4.1.7 by explicitly demonstrating that the same Feynman rules are obtained.

We also realized unoriented membrane theories, which we defined in terms of tetrahedral decompositions, as triangle–hinge models. Unoriented membrane theories are obtained from oriented open membrane theories of disk topology by gauging the worldvolume parity transformation Ω. For each triangle Δ in a tetrahedral decomposition, we have introduced two types of triplets, $(\Gamma_1, \Gamma_2, \Gamma_3)$ and $(\tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$, which respectively correspond to two ways of identification at Δ, (4.2.8) and (4.2.9). The transformation Ω is then defined as the interchange between $(\Gamma_1, \Gamma_2, \Gamma_3)$ and $(\tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$. After gauging Ω, an unoriented membrane theory treats all the tetrahedral decompositions in the sextet $(\Gamma_1, \Gamma_2, \Gamma_3, \tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3)$ equally. An unoriented membrane theory is realized as a triangle–hinge model with the action (4.3.1). It generates Feynman diagrams representing unoriented tetrahedral decompositions. We gave explicitly in (4.3.18)–(4.3.23) the sextet of Feynman diagrams $(\gamma_1, \ldots, \tilde{\gamma}_3)$ corresponding to $(\Gamma_1, \ldots, \tilde{\Gamma}_3)$, and showed that these six diagrams appear with unit coefficient up to factors of coupling constants if we treat the common part X as a set of distinguished external vertices and sum over all Wick contractions giving the same diagram. We further showed that matter degrees of freedom can be introduced to unoriented triangle–hinge models by coloring tetrahedra as carried out in [2]. Although we only discussed the coloring of tetrahedra in this paper, we can set matter degrees of freedom to simplices of any dimensions (i.e. tetrahedra, triangles, edges and/or vertices) as in [2].

Although the free energies of original triangle–hinge models diverge, we can redefine the models by using (5.1.18) and (5.1.19) and regard that A' and B' run over real symmetric matrices, without changing the original perturbation series. Since the higher-order terms are the trace of the matrices in the redefined models, they give finite free energies when the coefficients of the highest-order term are positive. Although the bilinear term $K(A, B)$ becomes complicated, the models may be solved even analytically because $K(A, B)$ have many good properties as we discussed in Sect. 5.2. Furthermore, the redefinition enables us study the models numerically. We performed the Monte Carlo simulation to the simplified triangle–hinge models (5.3.1), and found that a third-order phase transition occurs at $\beta \simeq 6$. But unfortunately this critical point may not correspond to three-dimensional gravity theory. This is because of the lack of the restriction to tetrahedral decompositions as we discussed in Sect. 3.7, and this suggests that we should study the triangle–hinge models with the

restriction (5.5.11). However, we then need to face the sign problem in the large-n limit because the action is complex-valued.

We used the generalized Lefschetz thimble method [40] as a prescription to solve the sign problem. Although this method is versatile, there is a serious multimodal problem. To improve this method, we proposed a Lefschetz thimble algorithm where the flow time is used as an auxiliary variable for the highly multimodal distribution. We in particular implemented the parallel tempering method by taking the flow time t as a tempering parameter. There, we prepare flow times $\{t_\alpha\}$ such that $t_0 = 0$ and $t_A = T$. The largest flow time T can be taken to be sufficiently large such that the sign problem disappears there. Although the algorithm includes configurations at $t = 0$, the original sign problem at $t = 0$ does not enter the calculation because the sample average is taken only with respect to $t = T$, where the sign problem disappears. We investigated the $(0 + 1)$-dimensional massive Thirring model at finite density to exemplify that the algorithm does work, and showed that contributions from multi thimbles are correctly taken into account even for such a large flow time as $T = 2$.

We should investigate to what extent this algorithm is actually versatile. We then study, as a check of versatility and as a warming up for the application to triangle–hinge models, the phase structure of the one-matrix model with complex coefficients (7.0.1). The action takes complex values when $c \neq \pm 1$, and three-cut solutions may appear, while only one-cut solution or two-cut solution may appear when $c = \pm 1$. We performed the generalized Lefschetz thimble method with the parallel tempering algorithm to this matrix model, and checked that the three-cut solution actually appear for a certain parameter region. It is our future work to determine the phase diagram of this model. In particular, it should be interesting to study the critical behavior around $\beta = 12, c = i$, because this point should be a critical point corresponding to two-dimensional gravity.

We then performed the Monte Carlo simulation of the triangle–hinge models with the restriction to tetrahedral decomposition (5.5.11). However, unfortunately the results still suffered from the sign problem, since the flow time we have set was not sufficiently large. Our analysis was made only for the $n = 3$ case, and of course it is insufficient for discussing the large-n limit. In order to take the flow time large and to study larger n, we must improve the generalized Lefschetz thimble method to reduce the computational cost.

In order to reduce the computational cost, we need to tune the tempering parameters so that the system thermalizes rapidly. As a tool to determine the optimal choice of the set of tempering parameters, we defined the distance $d_n(x_1, x_2)$ between two configurations x_1 and x_2 for a given MCMC algorithm in general. The distance is written only with the kernel of the transfer matrix $\hat{T}$, and various properties of the distance (including the universality for local MCMC algorithms) are easily understood as the reflection of properties of the transfer matrix. We made a detailed study of the distance for multimodal distributions,

and showed that distances between two different modes get dramatically reduced by the introduction of the simulated tempering method. The introduction of distance between configurations opens a way to investigate relaxation processes in an MCMC algorithm in terms of the geometry of the configuration space itself. We also considered the effective distance in the extended configuration space by regarding configurations in the same mode as a single configuration, and argued that the distance between two configurations may then be regarded as the geodesic distance in the extended configuration space with respect to an AdS-like metric. This geometrical interpretation should give us more insight into the optimization of the parameters.

We now list the future studies of these works. The first is to determine the phase structure of the triangle–hinge model with the restriction to tetrahedral decompositions (5.5.11). The model should have a critical point which corresponds to three-dimensional gravity theory. Since $K(A, B)$ have good properties as we discussed in Sect. 5.2, we may be able to solve the model even analytically. Of course we can study the phase structure numerically, but we then improve the generalized Lefschetz thimble method to reduce the computational cost. The second is the improvement of the generalized Lefschetz thimble method. The computational cost is expensive due to the flow equations (6.2.5) and (6.2.7). In particular, (6.2.7) have the cost of $\mathcal{O}(N^3)$. Although it is expensive to obtain J_{ij}, we need only the determinant $\det J$. We should be able to reduce the cost when we obtain the flow equation for $\det J$ itself. Furthermore, since we do not need to solve the flow equations exactly, we can consider another approach to reach near the thimbles. The deep learning method [42] would be suggestive, since this obtains the change of variables with small Jacobian. Unfortunately this may fail due to the lack of precision to reach the thimbles, but similar approaches can make a success if they have enough precision. The third is on the distance that we discussed in Chap. 9. We have seen that the distance actually quantifies the difficulty of transition from one configuration to the other configuration, and that AdS-like geometry emerges when we introduce the simulated tempering method. It should be interesting to study the geometry induced by the distance in detail, and such a geometrical viewpoint should enable us to adjust tempering parameters in a purely geometrical way.

Acknowledgments

I first thank my supervisor Masafumi Fukuma very much. He always discussed with me, and gave many lecture inflating my ideas. I am also grateful to my collaborators Sotaro Sugishita and Nobuyuki Matsumoto for very worthwhile discussion.

This work was supported by JSPS KAKENHI Grant Numbers JP17J08709.

Bibliography

[1] M. Fukuma, S. Sugishita and N. Umeda, "Random volumes from matrices," JHEP **1507** (2015) 088 [arXiv:1503.08812 [hep-th]].

[2] M. Fukuma, S. Sugishita and N. Umeda, "Matter fields in triangle–hinge models," PTEP **2016** (2016) no.5, 053B04 [arXiv:1504.03532 [hep-th]].

[3] M. Fukuma, S. Sugishita and N. Umeda, "Triangle–hinge models for unoriented membranes," PTEP **2016** (2016) no.7, 073B01 [arXiv:1603.05199 [hep-th]].

[4] M. Fukuma and N. Umeda, "Parallel tempering algorithm for integration over Lefschetz thimbles," PTEP **2017**, no. 7, 073B01 (2017) [arXiv:1703.00861 [hep-lat]].

[5] M. Fukuma, N. Matsumoto and N. Umeda, "Distance between configurations in Markov chain Monte Carlo simulations," JHEP **1712** (2017) 001 [arXiv:1705.06097 [hep-lat]].

[6] P. H. Ginsparg and G. W. Moore, "Lectures on 2-D gravity and 2-D string theory," In *Boulder 1992, Proceedings, Recent directions in particle theory*, 277-469 [hep-th/9304011].

[7] P. Di Francesco, P. H. Ginsparg and J. Zinn-Justin, "2-D Gravity and random matrices," Phys. Rept. **254** (1995) 1 [hep-th/9306153].

[8] V. A. Kazakov and A. A. Migdal, "Recent progress in the theory of noncritical strings," Nucl. Phys. B **311** (1988) 171.

[9] G. 't Hooft, "A planar diagram theory for strong interactions," Nucl. Phys. B **72** (1974) 461.

[10] E. Brézin and V. A. Kazakov, "Exactly solvable field theories of closed strings," Phys. Lett. B **236** (1990) 144.

[11] M. R. Douglas and S. H. Shenker, "Strings in less than one dimension," Nucl. Phys. B **335** (1990) 635.

[12] D. J. Gross and A. A. Migdal, "Nonperturbative two-dimensional quantum gravity," Phys. Rev. Lett. **64** (1990) 127.

[13] M. Fukuma, H. Kawai and R. Nakayama, "Continuum Schwinger-Dyson equations and universal structures in two-dimensional quantum gravity," Int. J. Mod. Phys. A **6** (1991) 1385.

[14] R. Dijkgraaf, H. L. Verlinde and E. P. Verlinde, "Loop equations and Virasoro constraints in nonperturbative 2-D quantum gravity," Nucl. Phys. B **348** (1991) 435.

[15] S. H. Shenker, "The Strength of nonperturbative effects in string theory," In "Cargèse 1990, Proceedings, Random surfaces and quantum gravity" 191–200.

[16] M. R. Douglas, "Strings in less than one dimension and the generalized KdV hierarchies," Phys. Lett. B **238** (1990) 176.

[17] M. Fukuma, H. Kawai and R. Nakayama, "Infinite dimensional Grassmannian structure of two-dimensional quantum gravity," Commun. Math. Phys. **143** (1992) 371.

[18] M. Fukuma, H. Kawai and R. Nakayama, "Explicit solution for p–q duality in two-dimensional quantum gravity," Commun. Math. Phys. **148** (1992) 101.

[19] J. Ambjorn, B. Durhuus and T. Jonsson, "Three-dimensional simplicial quantum gravity and generalized matrix models," Mod. Phys. Lett. A **6** (1991) 1133.

[20] N. Sasakura, "Tensor model for gravity and orientability of manifold," Mod. Phys. Lett. A **6** (1991) 2613.

[21] M. Gross, "Tensor models and simplicial quantum gravity in $>$ 2-D," Nucl. Phys. Proc. Suppl. **25A** (1992) 144.

[22] D. V. Boulatov, "A model of three-dimensional lattice gravity," Mod. Phys. Lett. A **7** (1992) 1629 [hep-th/9202074].

[23] L. Freidel, "Group field theory: An overview," Int. J. Theor. Phys. **44** (2005) 1769 [hep-th/0505016].

[24] R. Gurau and J. P. Ryan, "Colored tensor models - a review," SIGMA **8** (2012) 020 [arXiv:1109.4812 [hep-th]].

[25] R. Gurau, "The complete $1/N$ expansion of colored tensor models in arbitrary dimension," Annales Henri Poincare **13** (2012) 399 [arXiv:1102.5759 [gr-qc]].

[26] V. Bonzom, R. Gurau and V. Rivasseau, "Random tensor models in the large N limit: Uncoloring the colored tensor models," Phys. Rev. D **85** (2012) 084037 [arXiv:1202.3637 [hep-th]].

[27] S. Dartois, R. Gurau and V. Rivasseau, "Double scaling in tensor models with a quartic interaction," JHEP **1309** (2013) 088 [arXiv:1307.5281 [hep-th]].

[28] V. Bonzom, R. Gurau, J. P. Ryan and A. Tanasa, "The double scaling limit of random tensor models," JHEP **1409** (2014) 051 [arXiv:1404.7517 [hep-th]].

[29] S. w. Chung, M. Fukuma and A. D. Shapere, "Structure of topological lattice field theories in three dimensions," Int. J. Mod. Phys. A **9** (1994) 1305 [hep-th/9305080].

[30] J. Ambjorn, B. Durhuus and T. Jonsson, "Quantum geometry. A statistical field theory approach," Cambridge University Press, 1997.

[31] M. Cristoforetti *et al.* [AuroraScience Collaboration], "New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble," Phys. Rev. D **86**, 074506 (2012) [arXiv:1205.3996 [hep-lat]].

[32] M. Cristoforetti, F. Di Renzo, A. Mukherjee and L. Scorzato, "Monte Carlo simulations on the Lefschetz thimble: Taming the sign problem," Phys. Rev. D **88**, no. 5, 051501 (2013) [arXiv:1303.7204 [hep-lat]].

[33] A. Mukherjee, M. Cristoforetti and L. Scorzato, "Metropolis Monte Carlo integration on the Lefschetz thimble: Application to a one-plaquette model," Phys. Rev. D **88**, no. 5, 051502 (2013) [arXiv:1308.0233 [physics.comp-ph]].

[34] H. Fujii, D. Honda, M. Kato, Y. Kikukawa, S. Komatsu and T. Sano, "Hybrid Monte Carlo on Lefschetz thimbles - A study of the residual sign problem," JHEP **1310**, 147 (2013) [arXiv:1309.4371 [hep-lat]].

[35] M. Cristoforetti, F. Di Renzo, G. Eruzzi, A. Mukherjee, C. Schmidt, L. Scorzato and C. Torrero, "An efficient method to compute the residual phase on a Lefschetz thimble," Phys. Rev. D **89**, no. 11, 114505 (2014) [arXiv:1403.5637 [hep-lat]].

[36] A. Alexandru, G. Başar and P. Bedaque, "Monte Carlo algorithm for simulating fermions on Lefschetz thimbles," Phys. Rev. D **93**, no. 1, 014504 (2016) [arXiv:1510.03258 [hep-lat]].

[37] T. Kanazawa and Y. Tanizaki, "Structure of Lefschetz thimbles in simple fermionic systems," JHEP **1503**, 044 (2015) [arXiv:1412.2802 [hep-th]].

[38] Y. Tanizaki, Y. Hidaka and T. Hayata, "Lefschetz-thimble analysis of the sign problem in one-site fermion model," New J. Phys. **18**, no. 3, 033002 (2016) [arXiv:1509.07146 [hep-th]].

[39] H. Fujii, S. Kamata and Y. Kikukawa, "Lefschetz thimble structure in one-dimensional lattice Thirring model at finite density," JHEP **1511**, 078 (2015) Erratum: [JHEP **1602**, 036 (2016)] [arXiv:1509.08176 [hep-lat]].

[40] A. Alexandru, G. Başar, P. F. Bedaque, G. W. Ridgway and N. C. Warrington, "Sign problem and Monte Carlo calculations beyond Lefschetz thimbles," JHEP **1605**, 053 (2016) [arXiv:1512.08764 [hep-lat]].

[41] A. Alexandru, G. Başar, P. F. Bedaque, G. W. Ridgway and N. C. Warrington, "Monte Carlo calculations of the finite density Thirring model," Phys. Rev. D **95**, no. 1, 014502 (2017) [arXiv:1609.01730 [hep-lat]].

[42] A. Alexandru, P. F. Bedaque, H. Lamm and S. Lawrence, "Deep Learning Beyond Lefschetz Thimbles," Phys. Rev. D **96** (2017) no.9, 094505 doi:10.1103/PhysRevD.96.094505 [arXiv:1709.01971 [hep-lat]].

[43] M. Fukuma, S. Hosono and H. Kawai, "Lattice topological field theory in two dimensions," Commun. Math. Phys. **161** (1994) 157 [hep-th/9212154].

[44] C. Bachas and P. M. S. Petropoulos, "Topological models on the lattice and a remark on string theory cloning," Commun. Math. Phys. **152** (1993) 191 [hep-th/9205031].

[45] D. Weingarten, "Euclidean quantum gravity on a lattice," Nucl. Phys. B **210** (1982) 229.

[46] C. Itzykson and J. M. Drouffe, "Statistical Field Theory. Vol. 1: From Brownian Motion To Renormalization And Lattice Gauge Theory," Cambridge, UK: Univ. Pr. (1989) 1-403

[47] J. Polchinski, "String theory. Vol. 1: An introduction to the bosonic string," Cambridge University Press (1998).

[48] K. Nagata, J. Nishimura and S. Shimasaki, "Argument for justification of the complex Langevin method and the condition for correct convergence," Phys. Rev. D **94**, no. 11, 114515 (2016) [arXiv:1606.07627 [hep-lat]].

[49] E. Marinari and G. Parisi, "Simulated tempering: A New Monte Carlo scheme," Europhys. Lett. **19**, 451 (1992) [hep-lat/9205018].

[50] J. M. Pawlowski and C. Zielinski, "Thirring model at finite density in 0+1 dimensions with stochastic quantization: Crosscheck with an exact solution," Phys. Rev. D **87**, no. 9, 094503 (2013) [arXiv:1302.1622 [hep-lat]].

[51] J. M. Pawlowski, I. O. Stamatescu and C. Zielinski, "Simple QED- and QCD-like Models at Finite Density," Phys. Rev. D **92**, no. 1, 014508 (2015) [arXiv:1402.6042 [hep-lat]].

[52] P. Hasenfratz and F. Karsch, "Chemical Potential on the Lattice," Phys. Lett. **125B**, 308 (1983).

[53] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," [arXiv:1412.6980 [cs.LG]].

[54] K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," [arXiv:1502.01852 [cs.CV]]

[55] F. David, "Nonperturbative effects in matrix models and vacua of two-dimensional gravity," Phys. Lett. B **302** (1993) 403 doi:10.1016/0370-2693(93)90417-G [hep-th/9212106].

[56] M. Creutz, "Overrelaxation and Monte Carlo Simulation," Phys. Rev. D **36**, 515 (1987).

[57] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, "Hybrid Monte Carlo," Phys. Lett. B **195**, 216 (1987).n

[58] M. Fukuma, N. Matsumoto and N. Umeda, work in progress.