ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Finite gauge theories

Thesis presented by

Alberto Zaffaroni

for the degree of Doctor Philosophiae

Supervisor: Pietro Fré and Luciano Girardello

S.I.S.S.A. – I.S.A.S.

Elementary Particle Sector

academic year 1993 - 94

SISSA - SCUOLA
INTERNAZIONALE
SUPERIORE
DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

TRIESTE

Finite gauge theories

Thesis presented by

Alberto Zaffaroni

for the degree of Doctor Philosophiae

Supervisor: Pietro Fré and Luciano Girardello

S.I.S.S.A. - I.S.A.S.

Elementary Particle Sector

academic year 1993 - 94

Contents

Introduction

Two finite limiting case of QCD, two-dimensional pure QCD and four dimensional N=4 supersymmetric Yang-Mills have recently received a lot of attention, althought their physical content is somewhat limited.

Two dimensional QCD is indeed trivial; there are no propagating gluons and therefore the only degrees of freedom are of topological nature. The only way to capture them is to define the theory on non trivial manifold, but also in this case the partition function encodes the triviality of the theory by depending only on the area and the genus of the manifold. Moreover there is an explicit formula [1] for the partition function for every manifold and every gauge group. Before coupling to matter, the dynamics is completely absent.

For what concerns N=4 Yang-Mills, because of the extended supersymmetry one finds that the theory is finite and has zero beta function. This does not mean exact solvability of the theory: the computation of partition functions and of physical quantities is far from trivial. However, the statement, for example, that the coupling constant is small has a definite sense in such a quantum field theory; we can consistently canonically quantize the system in a box (perturbatively) for small coupling and then try to take the thermodynamic limit without fear that the coupling runs with the size of the box. This simplicity has a counterpart in that the theory is not asymptotically free and that probably does not confine. From this point of view, pure Yang-Mills and perhaps also N=1,N=2 Yang-Mills are far more interesting and physical.

However, despite of this lack of physical meaning, both theories have revealed much structure. Two dimensional QCD was shown to have a consistent string interpretation when expanded in 1/N [2] and N=4 Yang-Mills is the right arena in quantum field theory for S-duality.

QCD_2 reveals its string features when expanded in 1/N; just its solvability via Migdal's formula allows to compute exactly the coefficient of the expansion in the large N limit. It happens that these coefficients count the number of maps between a world-sheet and the target (the manifold on which the theory is defined); proposals for the underlying world-sheet lagrangian [3] and links with c=1 string theory were made [4]. The string picture applies to the large area region, the strong coupling regime; it is believed that a phase transition occurs at an intermediate value of the area, leading to a rather different behaviour of the theory in the weak coupling. A crucial quantity to check and interpret this transition is the partition function on the sphere.

For what regards N=4 Yang-Mills, it is an old conjecture that weak and strong coupling are dual, duality being the exchange of elementary particles of the lagrangian with monopoles, i.e. the simultaneous exchange of g with 1/g and of magnetic with electric quantum number. This S-duality has received attention recently [5] in the context of some models of heterotic string theory compactified on a torus; the flat field theory limit is exactly N=4 Yang-Mills. Evidence for S-duality has been given by Sen who found new states with both electric and magnetic numbers and by Witten and Vafa who compute the superpartition function on various manifold finding a modular form [6]. S-duality has been also a crucial ingredient for new results in N=2 Yang-Mills theory [7].

The two theories (at least in the approximation considered in the thesis) show similar features in the canonical quantization approach (on the cylinder for the two-dimensional case and on a box in the four dimensional). It happens that the relevant variables live on the Cartan torus of the gauge group. More precisely, they live on the Cartan torus divided by the Weyl group, a compact manifold with boundary: the manifold of the zero modes of the gauge fields. The Hilbert space of two dimensional QCD, composed by the irreducible representations of the gauge group, can be realized by the free motion on this manifold with null boundary conditions. For what regards N=4 Yang-Mills, the first order perturbative calculation constrains the modes to live on this manifold; this is completely different from non supersymmetric QCD where only a finite numbers of points of the manifold are quantum stable and fluctuations contribute to leading order. The conjugate variables, the momenta, live on the weight lattice. The quantum mechanics on the Cartan torus plays therefore the most consistent role in both theories.

A crucial point not well understood is how conformal field theory enters in these systems. Two dimensional QCD can be equally quantized on the cylinder as a theory of free fermions; it is again a quantum mechanical problem easily related to the previous one of the motion with zero boundary conditions. But in the large N limit, exactly as it happens in c=1 string, the fermions become relativistic and the hamiltonian becomes the perturbation (in 1/N) of a free conformal field theory. Analogously, in N=4 theory a striking resemblance with formulas for affine characters for Kac-Moody algebras appears in the computation of superpartition functions on various manifolds [6]. Why the Cartan torus should be restricted to allowed only integrable representations in some regimes is not clear.

The thesis is divided in two chapters. Both chapters begin with a brief review of the significant literature. In the first chapter we discuss a formulation of QCD_2 on the sphere in terms of variables living on the Cartan torus and solve a slightly modified model (corresponding to a Coulomb gas on the half-line). In the second chapter, we give further evidence for S-duality in N=4 Yang-Mills by proposing an ansatz for the partition function (or for a suitable sector) which is dual and satisfies non trivial physical conditions (like t'Hooft duality).

Chapter 1

Two dimensional QCD

It is an old idea that QCD might be represented as a string theory. The t'Hooft result about 1/N expansion [8] in perturbative QCD shows already string resemblance; the string coupling constant is identified with 1/N, which picks out the topology of the surfaces. The same is true for every matrix-model, whose Feynman graphs capture the triangulation of two-dimensional surfaces. Other evidence comes from the strong coupling lattice formulation of QCD, where the expansion of the free energy can be represented by a sum over surfaces [9]; problems arise in the attempt of finding the correct weights and the right way to take the continuum limit. Moreover, there is a phase transition between weak and strong coupling which is a typical lattice artifact [10].

The existence of a consistent, completely solvable formulation of two dimensional pure QCD, allows to check directly the statement that QCD is equivalent to a string theory, by finding the 1/N expansion of such a solvable theory and comparing it to the string expectation. What type of string theory we would expect? A first check of the QCD partition function shows that it is indeed a double expansion in powers of A and e^{-A}; A is the only quantity, except for the topology of the surface of definition, on which the theory depends. The factor e^{-A} suggests the Boltzman weight associated with an action proportional to the area, the Nambu-Goto action for string. The only known way to quantize Nambu action is to transform it into Polyakov action, which in two dimensions, realized as c=1 string, has a W_∞ symmetry; a very suspect fact if we note the area-preserving diffeomorphisms invariance (i.e. W_∞) of Yang-Mills action in two dimensions. However the argument leading from Nambu action to Polyakov action is suspect just in two-dimensions where not all maps are regular, the generic one being instead singular. Some resemblance with c=1 was however found [4]. The coefficient of the 1/N expansions was identified by Gross [2] as the number of maps with suitable smoothness from the world-sheet to the target space. A proposal for a two-dimensional world-sheet theory was made in [3].

The string representation is valid obviously when e^{-A} is small enough, i.e. for large A, the strong coupling phase of QCD. There is at least one case in which it is known that the string picture breaks down in the weak coupling, which is rather a point-particles theory: on a sphere a phase transition was found [11]. Note that this phenomenon is completely different from the lattice artifact of Gross and Witten in the one plaquette model, and that it is induced by an instanton condensation.

The first half of this chapter contains a review of the known results about the structure of QCD and its string interpretation. We begin with the derivation of Migdal's formula and the pseudo-topological nature of the theory. The 1/N expansion of Migdal formula contains all the informations about the string nature of QCD. In sections 3,4,5 we rewrite the problem in terms of the quantum mechanics on the Cartan torus divided by the Weyl group, revealing that these are indeed the relevant variables. In sections 6,7 we review what is known about the breakdown of string interpretation in the weak coupling regime on the sphere and the related phase transition. Sections 3,4,5,6 contain known results, but several derivation are original.

In the last part of the chapter, we rewrite the problem for the sphere or the disk in terms of an integral over the Cartan torus. In the large N limit the saddle point approximation reproduces the weak coupling partition function and the structure of the disk kernel. The related model of a Coulomb gas on the half-line is explicitly solved.

1.1 Migdal formula

QCD_2 is the next best thing to a topological field theory. In two dimension there are no propagating gluons and the theory becomes not trivial only when formulated on a compact manifold, where winding modes contribute to the physical degrees of freedom. In a topological field theory, the partition function and the observables are independent from the metric of the manifold, on which the theory is defined, and only depend on its topology. In two dimension QCD is not completely independent from the choice of a metric, but it depends only on the area of the manifold.

The partition function

$$Z_M = \int [DA^\mu] e^{-\frac{1}{4g^2}\int_M d^2x\sqrt{\eta}trF^{\mu\nu}F_{\mu\nu}} \tag{1.1}$$

does not really depend on the metric η, but only on the area A and on the topology of the manifold M. In fact, the theory is invariant under area-preserving diffeomorphisms (W_∞ symmetry). In two dimensions, we can always write the field strength as $F_{\mu\nu} = \epsilon_{\mu\nu} f$, where f is a scalar, so that

$$S = \int d^2x\sqrt{\eta}trF^2 = \int d\mu(trf^2) \tag{1.2}$$

makes explicit the fact that the action depends only on a choice of measure and not of metric.

What we have realized is that the partition function Z_M does really depend only on the topology of M (in two dimensions = genus), the combination A/g^2 and the choice of the gauge group.

To better deal with the theory and also to make sense of the Feynman path integral, we can start formulating it on a lattice, which triangulates or covers by polygons M. The finite dimensional analogous of the continuous space of connections which defines the path integral is the finite set of group elements U_γ, living on the edges γ of the polygons; if x and y are the endpoints of γ, U_γ can be regarded as the operator of parallel transport from x to y. A lattice gauge transformation is the assignment of a group element g_x to every vertex x, acting on connections as

$$U_\gamma \rightarrow g_y U_\gamma g_x^{-1}. \tag{1.3}$$

The manifold M becomes an union of plaquettes, and for each plaquette we need a lattice form of the action which must reduce to the Yang-Mills action in the limit in which the lattice spacing goes to zero. A famous candidate for this purpose is the Wilson action

$$S = \sum_P \frac{1}{g^2} Tr(U_P + h.c.) \tag{1.4}$$

where the sum is extended to all the plaquettes P and U_P is the ordered product of edges variables around P (the only gauge invariant quantity related to P, the holonomy around the plaquette). It is believed that as the covering becomes finer one can recover the desired continuum Yang-Mills theory, which can be considered as the fixed point of a renormalization group transformation applied to the Wilson action. What is peculiar about two dimensions is that it exists a lattice version of the theory invariant under subdivision of the lattice. Being invariant under renormalization group flow, this theory is already a fixed point = continuum theory and is completely equivalent to the path integral formulation.

The lattice theory can be specified by a choice of a local factor $\Gamma(U_P, A_P)$ for each plaquette P of area A_P,

$$Z_M(A) = \int \Pi_\gamma dU_\gamma \Pi_P \Gamma(U_P, A_P) \tag{1.5}$$

Migdal [1] suggested the factor

$$\Gamma(U, A) = \sum_R (dimR) \chi_R(U) e^{-\frac{Ac_R}{2}} \tag{1.6}$$

where the sum runs over all the representations R of the gauge group G, dimR and c_R are the dimension and the casimir of the representation. We need some explanations.

First of all, we must be sure that the naive continuum limit of this theory is the Yang-Mills lagrangian. It can be checked by a direct computation that it is indeed true and that the Wilson and Migdal actions are completely equivalent from this point of view; since we will provide an alternative hamiltonian derivation of Migdal's formula, we will skip this exercise.

The main reason for Migdal's formula is, as already mentioned, its invariance under subdivision. This means that if we consider a finer covering of the manifold the partition function computed with this factor does not change. We will prove this statement below; let's explain some immediate consequence of this property. We can associate the Migdal's kernel not only with the elementary plaquettes of the lattice, but also with a general closed contour of edges on the manifold: for the invariance under subdivision, it is consistent to consider $\Gamma(U, A)$ as the partition function of QCD restricted to the portion of surface enclosed by the contour of area A, with specified boundary conditions. For gauge invariance, this path integral can depend only on the holonomy of the gauge field around the contour, and more precisely on the coniugacy class of the holonomy U. This means that Γ must be a class function and can be expanded in characters, a complete ortonormal set of class functions, as in Migdal's formula. As for the coefficient, we can note that for $A \to 0$ one has

$$\sum_R dimR \chi_R(U) = \delta(U = 1) \tag{1.7}$$

which is the correct result for a loop of zero area. This explains the factor dimR; the exponential of the casimir is crucial both for obtaining the naive continuum limit and the subdivision invariance property.

To prove invariance under subdivision, consider the case in which a square Q of edges U_1, U_2, U_3, U_4 is divided in two triangles T_1, T_2 of edges U_1, V, U_2 and U_3, V, U_4 by the diagonal V. What we have to prove is that

$$\Gamma(Q) = \int dV \Gamma(T_1)\Gamma(T_2) \tag{1.8}$$

The product of factors for the triangles is

$$\sum_{R,S} dimR dimS \chi_R(U_1 V U_2) \chi_S(U_3 V^{-1} U_4) e^{-\frac{A_1 c_R + A_2 c_S}{2}} \tag{1.9}$$

performing the integration over V with the help of the formula

$$\int dV \chi_R(AV) \chi_S(V^{-1}B) = \delta_{RS} \frac{1}{dimR} \chi_R(AB) \tag{1.10}$$

we obtain

$$\sum_R dimR \chi_R(U_4 U_3 U_2 U_1) e^{-\frac{a c_R}{2}} \tag{1.11}$$

where $A = A_1 + A_2$, which is exactly the factor associated to the square.

What is the main advantage of this lattice formulation is that while the physical meaning is clear when the subdivision becomes finer and finer, the computation can be done and is simpler when the covering is smallest as possible. We can cover a compact manifold of genus G with a single polygon of 4g sides. In the well known case of the torus, we can represent the manifold with a square with identified sides U, V, U^{-1}, V^{-1}. In general, for every non trivial cycle on the surface we have 4 sides and two independent variables U_i, V_i. The partition function on this surface can be computed via Migdal formula,

$$Z_M(A) = \sum_R dimRe^{-\frac{Ac_R}{2}} \int dU_i dV_i \chi_R(U_1 V_1 U_1^{-1} V_1^{-1} ... U_g V_g U_g^{-1} V_g^{-1}) \quad (1.12)$$

The integrals can be computed via the formula (1.10) and the formula

$$\int dU \chi_R(AUBU^{-1}) = \frac{1}{dimR} \chi_R(A) \chi_R(B) \quad (1.13)$$

and ones gets

$$Z_M(A) = \sum_R (dimR)^{2-2g} e^{-\frac{Ac_R}{2}} \quad (1.14)$$

for the Yang-Mills partition function on a surface of genus g.

This is the famous Migdal's formula.

It exists a complementary way to derive this formula based on the hamiltonian quantization of the theory [12], which is strictly related to the almost topological nature of QCD in two dimension and which consents to compute partition functions by gluing manifolds.

Introducing the adjoint scalar field ϕ we can rewrite the lagrangian as

$$-\frac{\epsilon}{8\pi^2} \int d\mu Tr\phi^2 - i\frac{1}{4\pi^2} \int Tr\phi F \quad (1.15)$$

By performing the gaussian integration we recover Yang-Mills action.

Notice that the dependence from the metric (better from the measure) is confined to the quadratic term in ϕ. The term ϕF in fact represents a topological field theory called BF theory; it is equivalent to topological Yang-Mills theory up to the ghost multiplets [12], and it is related to Chern-Simons theory and Ray-Singer torsion [13].

By introducing decoupled fermionic fields in the action [12] we obtain the explicit form of topological Yang-Mills

$$\frac{1}{4\pi^2} \int Tr(-i\phi F - \frac{1}{2}\psi\psi) \quad (1.16)$$

invariant under BRST transformations

$$\delta A_\mu = i\epsilon\psi_\mu \quad (1.17)$$

$$\delta\psi_\mu = -\epsilon D_\mu \phi \quad (1.18)$$

$$\delta\phi = 0 \quad (1.19)$$

Notice that ϕ is an observable of the topological theory and its correlation functions do not depend on the point in which the operator is inserted. By now, we will consider full Yang-Mills theory as the integrated insertion of the BRST closed operator $e^{\int Tr\phi^2}$ in the topological BF theory.

We perform the canonical quantization on the cylinder of the first order BF theory. The hamiltonian is zero and the scalar field is represented by the derivative with respect to the gauge field. In the dual lagrangian formulation we should assign a path integral on the cylinder specifying on the circle at zero time the boundary conditions for the fields (the conjugacy class of the holonomy when gauge invariance is taken into account). So the hilbert space is the space of functions of the element U in the gauge group invariant by coniugation; this space is spanned by the complete set of characters of the group. We have a one-to-one correspondence between state in the hilbert space and representations of the gauge group, and the characters are the corresponding wave functions.

If we want to compute for example the correlation function of ϕ we can move the scalar field to the boundary where it acts as a derivative on the state R represented by the character $\chi_R(U) = Tre^{i\int A^aT^a}$ extracting from the exponential a factor T^a; so $Tr\phi^2$ extracts the casimir. We learn that, when sandwiched between the states corresponding to the representation R the area term $e^{tr\phi F}$ is computed to e^{Ac_R}.

Denoting with f_R the partition function of a disk with the state R inserted on the boundary in BF theory, we have that

$$\sum_R f_R \chi_R(U) = \delta(U = 1) \tag{1.20}$$

which is the partition function of a disk with specified holonomy U at the boundary. We learn that $f_R = dimR$.

The partition function of the cylinder with the states R,S inserted on the two boundaries is simply a delta function $h_{R,S} = \delta_{R,S}$, because the hamiltonian is zero.

As for the sphere with three holes, in which we insert the state R,S,T, we begin observing that if we insert the scalar field ϕ somewhere, moving it to the various boundaries we obtain each time the casimirs c_R, c_S, c_T; it is inconsistent unless $R = S = T$. Denoting the corresponding non null matrix element by g_R, we can compute it by observing that the gluing of a three holes sphere with a disk is a cylinder;so

$$\sum_R f_R g_R \delta_{R,S,T} = \delta_{S,T} \tag{1.21}$$

and we learn $g_R = 1/dimR$.

All two dimensional manifold can be obtained by gluing a certain number of the previous three, two and one-holes spheres; for example, a torus with two holes is the gluing of two three holes spheres and contributes a factor $1/dimR^2$. The generic genus G surface can be obtained by gluing G torus with two holes and closing the remaining holes with two disks, with a total factor $1/dimR^{2(G-1)}$.

Recalling the effect of the quadratic ϕ term, we easily recover Migdal's formula.

We will now discuss punctured surfaces, i.e. surfaces of genus g with boundary composed by a certain number of circles. In the hamiltonian formulation we impose a state = representation on the given boundaries, as in the previous discussion. On the other hand, in the lagrangian formulation we impose a given coniugacy class as a boundary condition for the path integral. We want to compute the partition function on a surface of genus g with n circle boundaries; the path integral is defined over all the connections with fixed holonomy around the circles.

$$Z(g, A; U_1, ..., U_n) \tag{1.22}$$

The configuration space is dual to the space of representations. We have

$$|U> = \sum_R |R>< R|U> = \sum_R |R> \chi_R(U) \tag{1.23}$$

For example, the kernel for the cylinder is

$$Z(U,V) =< U|V> = \sum_{R,S} < U|R > h_{R,S}(A) < S|V> = \sum_R \chi_R(U)\chi_R(V)e^{-Ac_R} \tag{1.24}$$

and that for the disk

$$Z(U) = \sum_R < U|R > f_R e^{-Ac_R} = \sum_R dimR\chi_R(U)e^{-Ac_R} \tag{1.25}$$

Note that the disk kernel can be obtain from the cylinder kernel by setting V=1; to set the holonomy to 1 is equivalent to the topological operation of closing a boundary. We will use often this procedure in the following. For example, to set U=1 in the disk or U=V=1 in the cylinder kernel gives the partition function on the sphere. On the other hand, for the ortogonality of characters, the torus partition function can be obtained by identifying U and V in Z(U,V) and by integrating over U.

It is easy to compute the partition function in the general case,

$$Z(A, g; U_1, ..., U_n) = \sum_R e^{-Ac_R}(dimR)^{2-2g-n} \prod_{i=1}^{n} \chi_R(U_i) \tag{1.26}$$

This completes our review of Migdal's formulas.

1.2 QCD as a string theory

In this section we briefly review the interpretation of two-dimensional QCD as a string theory [2]. The precise statement is that the free energy of QCD with gauge group $SU(N)$, when expanded for large N, is the partition function of some string theory with string coupling $1/N$ and string tension $\lambda = g^2N$.

A systematic expansion of Migdal's formula can be obtained via the expressions of casimir and dimension of a representation R in terms of the lengths n_i of the rows of the corresponding Young tableaux. In the case of $SU(N)$ the n_i are non negative decreasing integers, $n_1 \geq n_2 \geq ...n_N = 0$. Denoting with n the total number of boxes in the tableaux $((n = \sum n_i)$ and with d_R the dimension of the symmetry group S_n associated to the diagram,

$$c_R = nN + \sum n_i(n_i + 2i - 1) - \frac{n^2}{N}$$
$$dimR = \frac{\Pi_{i<j}(n_i - i - n_j + j)}{\Pi_{i<j}(i-j)} = \frac{d_R N^n}{n!} + O(N^{n-1}) \qquad (1.27)$$

We naively expect from the exponential in Migdal's formula the dominant term $e^{-\frac{\lambda A n}{2}}$ (recall that $\lambda = g^2N$). A string interpretation is easily found if we assume the Nambu-Goto action as the world-sheet theory,

$$Z_M^{Nambu} = \sum_g (g_{st})^{2g-2} \int Dx^\mu(z) e^{\int d^2z\sqrt{g}/2} \qquad (1.28)$$

The term $e^{-\frac{\lambda A n}{2}}$ is the contribution of a map from the world-sheet, covering n times the target space M of area A.

String expectations say that in the expansion

$$lnZ_M = \sum_{g=0}^{\infty} \frac{1}{N^{2g-2}} f_g^G(\lambda A) \qquad (1.29)$$

the coefficient f_g^G must count the number of suitable maps from the world-sheet of genus g to the target space M of genus G.

More precisely, QCD string theory counts the number of coverings of M with a finite number of branchpoints. A non-trivial check comes from Kneser's formula, which for a map from genus g to genus G, with winding number n and i branchpoints, implies

$$2(g-1) = 2n(G-1) + i \qquad (1.30)$$

For example, there are no maps from sphere to torus, or no maps from genus g to genus g which wind around more than once. According to this formula, in the large N expansion of the free energy some terms must be absent, as we will see. Moreover, if we add to a simple smooth covering with winding n (2(g-1)=2n(G-1)) branchpoints or collapsed handles we get also powers of A (corresponding to the arbitrary position of such points on the target); such powers of A may come from the expansion of the exponential of the casimir using formula (1.27).

The general expectation is

$$\sum_{g=0}^{\infty} \left(\frac{1}{N^2}\right)^{g-1} \sum_{n=0}^{\infty} \sum_{i=0}^{g-1-n(G-1)} \omega_{g,G}^{n,i} A^i e^{-nA/2} \qquad (1.31)$$

(for simplicity, by now, we put $\lambda = 1$) For example, a direct computation in the case of a torus shows that:

$$Z_{G=1} = \sum_{n_1 \geq n_2 \geq ... 0} e^{-A\sum n_i/2} = \sum_{k_i \geq 0} e^{-A\sum ik_i/2} = \eta(e^{-\lambda A/2}) + O(\frac{1}{N^2}) \quad (1.32)$$

where η is proportional to the Dedekind function

$$\eta(q) = \Pi_{n=1} \frac{1}{1-q^n} \quad (1.33)$$

To agree with the string interpretation, odd powers of N must be absent; but this is somewhat trivial (for every genus) following from the nice behaviour of casimir and dimensions when $N \to -N$, as one can checks easily from formula (1.27). Less trivial is the absence of term $O(N^2)$ in the torus free energy: it corresponds to the fact that there are no maps from sphere to torus. Moreover it is also true that Dedekind function counts exactly the number of maps from torus to torus; from Kneser's formula it follows that there are no branchpoints and, as a consequence, no powers of A.

Gross [2] gave the general proof that the coefficient $\omega^{n,i}_{g,G}$ is the number of maps described above. For G different from 1 there are some technical complications due to the introduction of so called Ω-points.

We simply note two points.

We have excluded maps with folds: for example, there are no zero winding modes in the QCD partition function ($e^{-0A/2}$). This is not surprising; folds would correspond to the propagation of the string centre of mass, the moving of a string particle, the tachion. But this is consistent with the fact that QCD has not propagating degrees of freedom.

The real structure of the large N expansion is slightly modified. The validity of the N expansion in formula (1.27) relies on the smallness of n_i, n; under this assumption the casimir behaves as nN. Unfortunately, there are also large Young tableaux (for example the dual of the previous ones) which contribute quantities of the order N. Gross proposed to reorganize the sum over representations in order to take in equal account small and dual tableaux.

$$\sum_{R,S} (dim\bar{S}R)^{2-2G} e^{-\frac{A}{2N} c_{\bar{S}R}} \quad (1.34)$$

One finds:

$$c_{\bar{S}R} = c_R + c_S + \frac{2n\tilde{n}}{N} \quad (1.35)$$

$$dim\bar{S}R = dimRdimS(1 + O(1/N)) \quad (1.36)$$

In this double sum we are allowed to sum over small tableaux for which the expansion encoded in formula (1.27) is valid.

Z is the square of the "first version" (chiral) partition function, except from 1) corrections from $dim\bar{S}R$, 2) the coupling term $\frac{2n\tilde{n}}{N}$. In the case of the torus the only corrections come from the small coupling term.

We summary the interpretation of the partition function in the case of the torus (G=1).

For a single chiral sector,

$$\sum_{g=0}^{\infty}\left(\frac{1}{N^2}\right)^{g-1}\sum_{n=0}^{\infty}\sum_{i=0}^{g-1-n(G-1)}\omega_{g,G}^{n,i}A^i e^{-nA/2} \tag{1.37}$$

We already know that the Kneser's bound is satisfied. When $2(g-1) = 2(G-1)+1$, taking into account the $O(1)$ term in the casimir ($\tilde{c}_R = \sum_i n_i(n_i - 2i+1)$), we obtain

$$\omega_{g,G}^{n,i} = \sum_R \left(\frac{n!}{d_R}\right)^{2G-2}\frac{1}{i!}\left(\frac{\tilde{c}_R}{2}\right)^i \tag{1.38}$$

In the case of a torus there is a simple combinatorial proof [2] that shows it is exactly the number of homotopically inequivalent maps from a surface of genus g to the torus.

Let's include the corrections. A first term comes from the factor $\frac{n^2}{N}$ in the casimir; writing

$$\frac{An^2}{2N^2} = \frac{nA}{2N^2} + \frac{n(n-1)A}{2N^2} \tag{1.39}$$

we interpret the first term as the limiting case in which a handle of the world-sheet is mapped to a single point in the target; the factor $1/N^2$ takes into account that the genus g is increased by one unity, A comes from the position of the handle, n from the windings and 1/2 from the indistinguibility of the two ends of the handle. The second term has the interpretation of a small tube from two different sheets at a single point in the world-sheet; the factor n(n-1)/2 takes into account the possible locations of the two ends of the tube. All these contributions clearly exponentiate and give the missing subleading term in the casimir.

Note that all the previous operations preserve the orientation of the surfaces. So we can completely understand the chiral partition function as a sum over orientation preserving maps. The two chiral copies correspond to the two opposite orientations. The small coupling $-An\tilde{n}/N^2$ can be interpreted as an infinitesimal tube from a $\tilde{n}$ sheets covering of given orientation to a n sheets covering of the opposite orientation; for the distinguibility of the two sheets there is no 1/2 symmetry factor. Notice that the theory provides a minus sign for every tube.

Starting from these evidences of the string nature of QCD, one of the main open problem is to find the correct world-sheet lagrangian. It is believed [14] that the correct candidate is some two dimensional topological field theory coupled to topological gravity. There is a proposal [3] to take a sigma model coupled to topological gravity to reproduce all the feature of Yang-Mills string theory in the limit of area zero. The area can be include by a perturbation of this topological theory via the operator corresponding to the kaeler form of the target; the delicate contact term algebra of this operator might reproduce the powers of A in the string expansion.

There is also a correspondence of Yang-Mills theory with c=1 string theory [4], obtained recognizing the fermionic nature of the degrees of freedom of the problem: Yang-Mills canonically quantized on a cylinder is equivalent to a system of free fermions living on a circle (see also below). The difference with c=1 is that in the non-critical string the fermions live on the real line. By bosonization of the degrees of freedom, one can obtain the Das-Jevicki-Sakita hamiltonian.

1.3 Moving on a group manifold

In this section we face with an alternative description of Yang-Mills theory as the quantum mechanics on a group manifold or, equivalently, the motion of free particles in the corresponding Coxeter box with vanishing boundary conditions [15] [16].

Perhaps the simple way to recognize this alternative description is a closer look to Migdal's formula in the case of the cylinder with specified holonomy U and V as boundary conditions,

$$Z(U,V;A) = \sum_R \chi_R(U)\chi_R(V)e^{-\frac{Ac_R}{2}} \tag{1.40}$$

It has the aspect of a kernel in which we can take the area A as the time of the cylinder, c_R as the hamiltonian and the characters χ as the wave functions.

In the case of SU(N), the elements U and V can be diagonalized: $U = (e^{ix_i})$, $V = (e^{iy_i})$, where $\sum x_i = \sum y_i = 0, i = 1, ..., N$. The real configuration space is not the maximal torus T with coordinates x_i, because the hilbert space is not the complete space of function of U, but only of class function; the effect of taking functions invariant under coniugation (a permutation of the x_i) corresponds to take T divided by the Weyl group W as the configuration space. It is, as we will see, a compact manifold with boundary.

A simple hamiltonian defined on T/W with the desired spectrum is the laplacian describing the free motion on this space with vanishing boundary conditions.

For the constrain $\sum x_i = 0$, T/W has dimension N-1 and is spanned by the vectors $f_a = e_{a-1} - e_a, a = 1, ..., N-1$ where $e_i, i = 1, ..., N$ is the canonical basis of R^N. The f_a are the well known simple root and their scalar product is the Cartan metric α_{ab}, the induced metric on T/W. By now, we will always consider both the indices a and i; for example, a vector in the root lattice $r \epsilon \Lambda_R$ has N-1 coordinates with respect to the basis f_a ($r = \sum r^a f_a$), but N component when considered embedded in R^N. Some quantities are simpler when expressed in R^N with redundant variables: for example, the Weyl group becomes simply the permutation group S_N.

In R^N, T/W has the simple aspect of the hyperplane $\sum x_i = 0$ with coordinates defined modulo 2π, moded by arbitrary permutations (the Weyl group $W = S_N$). Let's see how it looks in the independent variables x_a. The hyperplane $\sum x_i = 0$ is the vector space V of dimension N-1 spanned by the vectors f_a; periodicity of the variables is a simple shift $2\pi f_a$, selecting the fundamental cell of the root lattice (modulo a trivial 2π dilatation of unity): this is the maximal torus T. The generic element of the Weyl group is a reflection in the plane orthogonal to a given root

$$x \to x - \frac{2(\alpha x)\alpha}{\alpha^2} \tag{1.41}$$

and is generated by the reflections about simple roots. For SU(N) the set of the roots is composed by the vectors $e_{ij} = e_i - e_j$. It is simple to check that the reflection about e_{ij} corresponds in R^N to the permutation of x_i and x_j. So, as we have already anticipated, the Weyl group is isomorphic to S_N and it acts in R^N exactly as the permutation group of the coordinates x_i while in T is the group of reflections about f_a.

To take into account the action of W, we can label the angles x_i in such a way that moving on the circle counterclockwise starting from 1 we encounter first x_N, then x_{N-1} and so on, until we find x_1 before ending the circle; in this way, we have moded by W and taken into account the periodicity. In terms of the variables $\tilde{x}_a$ (tilded to avoid ambiguity in the next formulas), we have $\tilde{x}_a = x_{a-1} - x_a > 0$ and $\sum \tilde{x}_a = x_1 - x_N < 2\pi$. Denoting $f_0 = -\sum_{a=1}^{N-1} f_a$ T/W is defined by

$$\begin{aligned} x_a = x f_a > 0 \\ -x f_0 < 2\pi \end{aligned} \tag{1.42}$$

So T/W is the simplex defined by the N hyperplanes of the formula (1.42);it is a compact manifold with N boundary. Acting with the Weyl group on this simplex we obtain all the fundamental cell of the root lattice.

The vectors f^a defined by $f^a f_b = \delta_{ab}$ are a basis for the weight lattice Λ_W. The matrix $\alpha_{ab}^{-1} = f^a f^b$ can be computed to

$$\alpha_{ab}^{-1} = inf(a,b) - \frac{ab}{N} \tag{1.43}$$

The natural hamiltonian is the laplacian in the induced metric

$$H = \sum \frac{\partial}{\partial x_a} \alpha_{ab}^{-1} \frac{\partial}{\partial x_b} \tag{1.44}$$

and the natural wavefunction is $e^{ix \times k}$, where k is a generic element of the weight lattice. Since the product of elements of weight and root lattice is an integer, the vawefunction is periodic in x. We have

$$H e^{ix \times k} = k^2 e^{ix \times k} \tag{1.45}$$

Now we recall that an irreducible representation is specified by its highest weight λ, with coefficient $\lambda_a \geq 0$, i.e. a "dominant" weight. There is a simple formula for the casimir of a representation with highest weight λ

$$c_\lambda = \lambda^2 + 2\lambda\rho \tag{1.46}$$

where $\rho = \frac{1}{2}\sum_{\alpha>0}\alpha = \sum_a f^a$ (α is a generic root). So the casimir is, up to a constant, the square of the weight vector $k = \lambda + \rho$; k covers all the set of non-negative coefficient weight. The weights $\lambda = \sum \lambda_a f^a$ are called "strictly dominant" if $\lambda_a > 0$ for every a; they belong to the principal Weyl chamber and are in one-to-one correspondence with the irreducible representations. All other weights belong to the walls of the chambers or can be obtained by the strictly dominant weights acting with the Weyl group.

We have seen that the energy of the plane wave is, up to a constant, the casimir of the corresponding representation; we have again to mode out by the Weyl group. The correct QCD spectrum can be obtained by imposing zero boundary conditions on the walls limiting T/W, i.e. the hyperplanes defined by formula (1.42); this suggests to set

$$\psi_k(x) = \sum_{w \epsilon W} \epsilon(w) e^{ix \times k^w} \tag{1.47}$$

This wavefunction vanishes if k or x are invariant by an odd element of the Weyl group. The first N-1 hyperplanes defined by f_a are trivially invariant under the reflection about simple roots and so $\psi_k(x)$ vanishes on these. For what regards the last hyperplane, it is simple to check that the reflection about it $x \longrightarrow x - (xf_0 + 2\pi)f_0$ changes its sign. So $\psi_k(x)$ vanishes on the boundaries of T/W.

Moreover, it vanishes for k on the walls of Weyl chambers, which are in fact characterized by the fact that some elements of the Weyl group leave them invariant. We obtain a one-to-one correspondence between wavefunctions and strictly dominant weights (irreducible representations) with energy equal, up to a constant, to the casimir. Let's note that the trivial representation $\lambda = 0$ which has energy zero corresponds to $k = \rho$, so we have to subtract the zero point energy ρ^2 from the expression k^2 of the energy.

It is also true that, under the same assumptions, the character of a given representation corresponding to k is

$$\chi_k(x) = \frac{\psi_k(x)}{\psi_\rho(x)} \tag{1.48}$$

(see also next section), so the character is proportional to the wavefunction of our quantum mechanical system.

Under all these assumptions, formula (1.40) can be reinterpreted, up to a proportionality factor, as the kernel

$$Z(U, V; A) =< U|e^{-AH/2N}|V > \tag{1.49}$$

Recalling that with the redundant variables of R^N the Weyl group is simply the permutation group, we can note that the completely antisymmetric wavefunction $\psi_k(x)$ resembles the quantum mechanical wavefunctions of N fermions, one for every x_i; this would be true if all the variables x_i were independent. This is exactly what happens if we enlarge the gauge group to U(N), as we will see in the next section.

Now we will translate some quantities of two-dimensional QCD in the new language.

Denote with Λ_W^+ the subset of the weight lattice corresponding to strictly dominant weights. We have seen that, up to a constant zero point energy, the partition function of QCD on a torus can be written

$$Z_{G=1}(A) = \sum_{k\epsilon\Lambda_W^+} e^{-\frac{A}{2N}k^2} \tag{1.50}$$

Without the restriction to strictly dominant weights, it would be the standard theta function $T(t) = \sum_{\Lambda_W} e^{-\pi t k^2}$ on the weight lattice with nice modular transformation

$$T(t) = \sum_{k\epsilon\Lambda_W} e^{-\pi t k^2} = \frac{\sqrt{N}}{t^{N-1/2}} \sum_{r\epsilon\Lambda_R} e^{-\frac{\pi}{t}r^2} \tag{1.51}$$

Formula (1.51) follows trivially from Poisson's formula

$$\sum_{k\epsilon Z^n} e^{-\pi k A k + 2\pi i k x} = \frac{1}{det A^{n/2}} \sum_{r\epsilon Z^n} e^{-\pi(r-x)A^{-1}(r-x)} \tag{1.52}$$

when one notes that, in coordinates, $k^2 = k_a \alpha_{ab}^{-1} k_b$, if $k = k_a f^a \epsilon \Lambda_W$ and $r^2 = r^a \alpha_{ab} r^b$ if $r = r^a f_a \epsilon \Lambda_R$ (recall $f_a f_b = \alpha_{ab}$) and that $det\alpha = N$.

The generalization of formula (1.51) is

$$T(x;t) = \sum_{k\epsilon\Lambda_W} e^{-\pi t k^2 + 2\pi i k x} = \frac{\sqrt{N}}{t^{N/2}} \sum_{r\epsilon\Lambda_R} e^{-\frac{\pi}{t}(r-x)^2} \tag{1.53}$$

QCD on a torus is not a theta function since we have excluded in its definition the walls of Weyl chambers in the weight lattice. The interiors of all the other chambers (some k^a negative) can be reached by the action of the Weyl group which is free and transitive; by dividing by N! we can extend the sum over Λ_W^+ to all the weight lattice excluded the walls of the chambers, because the casimir is Weyl invariant,

$$Z_{G=1}(A) = \frac{1}{N!} \sum_{k^a \neq 0} e^{-\frac{A}{2N}k^2} \tag{1.54}$$

We can consistently write this partition function as a sum of theta functions on increasingly smaller lattices composed by the walls, the walls of walls, and so on, with the right combinatorial coefficients to cancel the contribution of the boundaries. We have lost in any case the nice modular properties of weight lattice theta function. Recall that modular properties for QCD on a torus are suggested by the fact that the first term in the large N expansion is a modular form (the Dedekind function); in the next section we will return on this problem.

The case in which we can extend the sum over all the weight lattice is that of the kernel

$$U(x_1, x_2; t) =< x_1|e^{-tH}|x_2 >= \frac{1}{N^{1/2}} \sum_{k\epsilon\Lambda_W^+} \psi_k^*(x_1)\psi_k(x_2)e^{-\pi tk^2} \tag{1.55}$$

As we have already noted, ψ_k vanishes when k belongs to a wall, therefore

$$U(x_1, x_2; t) = \frac{1}{N!N^{1/2}} \sum_{\Lambda_W} e^{-\pi tk^2} \tag{1.56}$$

The factor $N^{1/2}$ comes from the normalization of the wavefunctions,

$$\int_{T/W} \psi_k^*\psi_{\tilde{k}} = \sqrt{N}\delta_{k\tilde{k}} \tag{1.57}$$

following from the fact that the volume of T/W is $\sqrt{N}/N!$ (it is the volume of the elementary cell of root lattice $(det f_a f_b)^{1/2} = (det\alpha)^{1/2} = N^{1/2}$ divided by the order of the Weyl group).

Dividing $U(x_1, x_2; t)$ by ψ_ρ^2 we obtain exactly the kernel for QCD on the cylinder. By comparing the orthogonality conditions for characters we learn that

$$dU = \frac{1}{N^{1/2}}|\psi_\rho(x)|^2 dx \tag{1.58}$$

The cylinder kernel can be used to compute several quantities. If we send $x \to 0$, $\chi_R(U) \to \chi_R(1) = dimR$, and we obtain the partition function on the disk $(V \to 0)$ or on the sphere $(U, V \to 0)$. On the other hand, for the orthogonality conditions of the characters (in this case interpreted as a gluing of manifolds), if we identify the end circles of the cylinder and integrate on U, we obtain the torus partition function.

Let's explore the modular properties (i.e. Poisson formula) from this point of view. We have

$$U(x_1, x_2; t) = \frac{1}{N^{1/2}} \sum_{w\epsilon W} \epsilon(w) T(x_1 - x_2^w; t) \tag{1.59}$$

Integrating on $x_1 = x_2$ we obtain (see also [17])

$$\sum_{\Lambda_W^+} e^{-\pi tk^2} = t^{-(N-1)/2} \sum_{W,\Lambda_R} \epsilon(w) \int_{T/W} dx e^{-\frac{\pi}{t}(x-x^w-r)^2}. \tag{1.60}$$

This is the aspect of Poisson's formula on the torus. It does not seem in the more convenient form to perform the large N expansion, and does not say a lot about modular properties of the various terms.

1.4 More Lie algebras

In this section we want to clarify some technical points about the relation between Young tableaux, highest weights and Dinkin labels, the modular properties of SU(N) theta functions and the relation between SU(N) and U(N) gauge theory.

Just the U(N) gauge theory will have a main role in the following.

As we have already said, an irreducible representation of SU(N) is determined by its highest weight $\lambda = \sum_{a=1}^{N-1} m_a f^a$, where the coefficients m_a are non-negative integers. The relation with the previous section is that $k = \lambda + \rho, k_a = m_a + 1$.

In terms of Young tableaux, m_a is the number of columns of lenght a. The dual description in terms of the lengths of the rows, $n_1 \geq n_2 \geq ... \geq n_N = 0$ is obtained via

$$\begin{aligned} m_a &= n_a - n_{a-1} \\ n_a &= \sum_{b=a}^{N-1} m_b \end{aligned} \tag{1.61}$$

Notice that $n = \sum am_a$.

The second casimir is $c_R = \lambda^2 + 2\rho\lambda = k^2 - \rho^2$. Let's compute the zero point energy. Recall that

$$\begin{aligned} \rho &= \sum f^a \\ f^a f^b &= \alpha_{ab}^{-1} = inf(a,b) - \frac{ab}{N} \end{aligned} \tag{1.62}$$

We have $\rho = \sum_{ab} \alpha_{ab}^{-1}$. Now,

$$\begin{aligned} &\sum_a \alpha_{ab}^{-1} = \sum_{a=1}^{b} a + \sum_{a=b+1}^{N-1} b - b \sum_{a=1}^{N-1} a/N = \\ &b(b+1)/2 + (N-b-1)b - b(N-1)/2 = b(-b+N)/2 \end{aligned} \tag{1.63}$$

and for the zero point energy we find

$$\rho^2 = N(N-1)(N+1)/12 \tag{1.64}$$

It is a very boring but simple exercise to compute the casimir in terms of the lengths of the rows,

$$c_R = \lambda^2 + 2\rho\lambda = nN + \sum_{a=1}^{N-1} n_i(n_i - 2i + 1) - \frac{n^2}{N} \tag{1.65}$$

our old friend.

The characters of a given representation of SU(N) are given by the Weyl formula as a ratio of sum over the Weyl group; the numerator and the denominator are exactly the wavefunctions met in the previous section. If we introduce the decreasing sequence of positive integers $h_1 > h_2 > ... > h_N = 0$, defined through $h_i = \sum_{b=i}^{N-1} k_b, k_b = h_b - h_{b-1}$ we find

$$k = \sum_{i=1}^{N} \left(h_i - \frac{1}{N} \sum_{j=1}^{N} h_j \right) e_i$$

$$E_k = k^2 = \sum_{i=1}^{N} h_i^2 - \frac{1}{N} \left(\sum_{j=1}^{N} h_j \right)^2 = \sum_{i=1}^{N} \left(h_i - \frac{1}{N} \sum_{j=1}^{N} h_j \right)^2 \quad (1.66)$$

From the previous formulas, taking into account that $\sum_{i=1}^{N} x_i = 0$, we find that $kx = \sum_{i=1}^{N} h_i x_i$. So, in R^N the wavefunction ψ_k looks like a $N \times N$ determinant of one coordinate plane waves with momentum h_i, $e^{ih_i x_j}$,

$$\psi_k(x) = det(e^{ih_i x_j})_{ij} \quad (1.67)$$

looking like the wavefunction of N free fermions.

Recalling that $k_a = h_a - h_{a+1}, m_a = n_a - n_{a+1}$ we find that the correct identification in terms of row lengths is

$$h_a = n_a - a + N \quad (1.68)$$

and that for the identity representation we have $h_a^0 = N - a$.

The set of arbitrary positive decreasing integers h_a can be used to label the irreducible representations; the character of the representation h is given by,

$$\chi_{h_i}(x_i) = \frac{\psi_{h_i}(x_i)}{\psi_{h_i^0}(x_i)} \quad (1.69)$$

Weyl's trick of formally posing $x_i = \epsilon i$ can be used to compute the dimension $dim(h_i) = \chi_{h_i}(0)$ when $\epsilon \to 0$ from the leading term of this expression,

$$\chi_{h_i} =_{\epsilon \to 0} (i\epsilon)^{N(N-1)/2} \Pi_{i<j}(h_i = h_j) \quad (1.70)$$

giving the old formula

$$dim(h_i) = \frac{\Pi_{i<j}(h_i - h_j)}{\Pi_{i<j}(i - j)} \quad (1.71)$$

which can be interpreted also as a ratio of sums over positive roots

$$dim(k) = \Pi_{\alpha>0} \frac{k\alpha}{\rho\alpha} \quad (1.72)$$

Now we come to the main point of this section, the right understanding of the U(N) theory. The U(N) representations [18] are labelled by N independent non decreasing integers (not necessarily positive) $n_1 \geq n_2 \geq ... n_N$. The casimir is

$$c_R(U(N)) = nN + \sum_{i=1}^{N} n_i(n_i - 2i + 1) \quad (1.73)$$

Note that it is identical to SU(N) casimir without the $O(1/N)$ term; the similarity is misleading because now we have N integers m_i against the N-1 integers of the SU(N) case. The further degree of freedom is obviously an U(1) charge. Let's see how it works.

The equivalence classes $\{n_i - n_i' = s\}$, with s fixed integer, realize the reduction to SU(N). In each of these classes there is one element with $n_N = 0$; it corresponds to the associated SU(N) representation. In general, $\tilde{n}_i = n_i - n_N$ are the row lengths of the Young tableaux of the associated SU(N) representation.

In terms of the basis matrices $(E^{kl})_{ij} = \delta_{ki}\delta_{lj}$ the generator for the fundamental representation of U(N) are

$$H^i = E^{ii} - E^{i+1,i+1},\ E^{ij} + E^{ji},\ i(E^{ij} - E^{ji}),\ Q = \sum_i E^{ii} \tag{1.74}$$

The Killing metric $g^{ij} = TrT^iT^j$ is block-diagonal for $SU(N) \times U(1)$, it is the well known Cartan metric α_{ab} for SU(N) and $g^{QQ} = N$.

As a consequence, the casimir $c = g_{ij}T^iT^j$ splits into

$$c = c(SU(N)) + g_{QQ}T^QT^Q = c(SU(N)) + \frac{1}{N}Q^2 \tag{1.75}$$

Let's compute the charge of the representation $\{n_i\}$. Setting $n_i = \tilde{n}_i + \lambda, n_N = \lambda$, we obtain

$$c = \tilde{n}N + \sum \tilde{n}_i(\tilde{n}_i - 2i + 1) - \frac{\tilde{n}^2}{N} + \frac{1}{N}(\tilde{n} + \lambda N)^2 = c(SU(N)) + \frac{1}{N}Q^2 \tag{1.76}$$

so we can identify the U(1) charge of the representation

$$Q = \tilde{n} + \lambda N \tag{1.77}$$

We have learned that the charge of the U(1) factor has been chosen not in an independent way, but is congruent modulo N to the total number of boxes of the associated SU(N) diagram. As we will see in a moment, this means the the U(1) factor is linked to the center of SU(N).

The center of SU(N) is Z_N and is isomorphic to the ratio Λ_W/Λ_R; this means that the weight lattice contains exactly N copies of the root lattice, which we can obtain via the shift vector f^1.

Since α_{ab} are integers, Λ_R is a sublattice of Λ_W; moreover

$$\begin{aligned} &f^i - if^1 \epsilon \Lambda_R \\ &Nf^1 \epsilon \lambda_R \end{aligned} \tag{1.78}$$

We conclude that Λ_W consists of the N copies of the root lattice $\mu f^1 + \Lambda_R, \mu = 0, ..., N-1$, with a structure isomorphic to the discrete group Z_N.

We want to show that a given equivalence class corresponds to the representations with total number of boxes congruent to μ modulo N.

In fact, if $\lambda = \mu f^1 + r, \lambda \epsilon \Lambda_W, r \epsilon \Lambda_R$ we have

$$\begin{aligned} \lambda &= \mu f^1 + r^a \alpha_{ab} f^b \\ m_a &= \mu \delta_{a1} + r^b \alpha_{ba} \end{aligned} \tag{1.79}$$

Using $\sum a \alpha_{ba} = N \delta_{b,N-1}$ we find

$$n = \sum a m_a = \mu + N r^{N-1} = \mu (mod N) \tag{1.80}$$

In correspondence with the center Z^N, the SU(N) theta function T(t) splits into N sublattice sums,

$$T_\mu(t) = \sum_{r \epsilon \Lambda_R} e^{-\pi t (r + \mu f^1)^2} \tag{1.81}$$

Using

$$\begin{aligned} k &= r + \mu f^1 = r^a f_a + \mu \alpha^{1a} f_a \\ k_a &= \mu \alpha^{1a} + r^a \\ k_a &= \mu - \frac{a\mu}{N} + r^a \end{aligned} \tag{1.82}$$

and defining new integers $s^a = \mu + r^a$, we find the explicit expression

$$T_\mu(t) = \sum_{s^a \epsilon Z} e^{-\pi t \left(s^a - \frac{\mu a}{N}\right) \alpha_{ab} \left(s^b - \frac{\mu b}{N}\right)} \tag{1.83}$$

It is now an exercise in Poisson resummation to prove the following nice modular covariance property,

$$T_\mu(t) = \frac{1}{N^{1/2} t^{N-1/2}} \sum_{\nu=1}^{N} e^{2\pi i \frac{\mu\nu}{N}} T_\nu(1/t) \tag{1.84}$$

In fact,

$$\begin{aligned} &\sum_{\nu,s} e^{2\pi i \frac{\mu\nu}{N} - \pi t^{-1} \left(s^a - \frac{\mu a}{N}\right) \alpha_{ab} \left(s^b - \frac{\mu b}{N}\right)} =_{Poisson} \\ &t^{N-1/2} det\alpha^{-1/2} \sum_{\nu,k} e^{-\pi t k_a \alpha^{-1}_{ab} k_b + 2\pi i \frac{\nu}{N} (\mu - \sum a k_a)} = \\ &t^{N-1/2} N^{1/2} \sum_k e^{-\pi t k^2} \delta_{\sum a k_a = \mu (mod N)} = t^{N-1/2} N^{1/2} T_\mu(t) \end{aligned} \tag{1.85}$$

The modular property (1.84) will be very useful in the understanding of the electric flux sectors in N=4 Yang-Mills theory. For what regards QCD_2 we will simply note some points.

First of all, the condition $\sum a k_a = \mu (mod N)$ is equivalent to $n = \mu - N(N-1)/2 (mod N)$; this means that the total number of boxes is congruent to the label of the central sector μ only if N is odd. For simplicity, by now, we will assume that this is indeed the case; for N even there are some 1/2 factors of difference.

Defining the N integers

$$\begin{aligned}
&\phi_1 = s^1 \\
&\phi_2 = s^2 - s^1, \ldots \\
&\phi_{N-1} = s^{N-1} - s^{N-2} \\
&\phi_N = \mu - s^{N-1} \\
&\sum_{i=1}^{N} \phi_i = \mu
\end{aligned} \tag{1.86}$$

and using the explicit form of α_{ab}, we find the R^N expression of the T_μ,

$$T_\mu(t) = \sum_{\sum_{i=1}^{N} \phi_i = \mu} e^{-\pi t \sum_{i=1}^{N} (\phi_i - \frac{\mu}{N})^2} \tag{1.87}$$

We have seen that the U(1) charge is coniugated to the central sectors of SU(N); this implies that in every sector the contribution of the U(1) factor to the partition function factorizes. The U(1) contribution to the casimir $A(\lambda + n/N)^2$ if $n = \mu(modN)$ clearly factorizes in a theta function of the form

$$\theta^k(q) = \sum_{n \epsilon Z} q^{\frac{1}{2}(n-k)^2} \tag{1.88}$$

giving

$$T^{U(N)}\left(\frac{A}{N}\right) = \sum_{\mu=1}^{N} \theta^{\mu/N}(e^{-A}) T_\mu\left(\frac{A}{N}\right) \tag{1.89}$$

1.5 QCD on a torus, free fermions and conformal field theory

We have already compute the leading term of QCD on a torus. At genus one, it is simple to find the representations which contribute leading terms to the casimir

$$k_a \alpha_{ab}^{-1} k_b - N(N-1)(N+1)/12 \tag{1.90}$$

It is a positive quadratic form with k_a greater or equal to 1. The smallest weight $\rho, k_a = 1$ exactly cancels the zero point energy. Representations in which all $k_a > 1$ give a contribution of order $O(N^3)$, suppressing the exponential. The leading contribution comes from representations with only a finite number of $k_a > 1$ with $a \ll N$ and their dual. In terms of $k_a = m_a + 1$, the condition is $n = \sum a m_a \ll N$

$$\begin{aligned}
c = O(1) + \frac{2}{N}\Big(\sum_{i \geq j} j(N-i+1) m_i + \sum_{i<j} i m_i (N-j+1)\Big) = \\
O(1) + \frac{2}{N}\Big(\sum i m_i\Big)\frac{N(N+1)}{2} = nN + O(1)
\end{aligned} \tag{1.91}$$

Summing only on representations with $n \ll N$, it trivially follows that

$$Z_{G=1} = \sum_{n_1 \geq n_2 \geq \ldots 0} e^{-A \sum n_i/2} = \sum_{k_i \geq 0} e^{-A \sum i k_i/2} = \eta(e^{-A/2}) + O(\frac{1}{N^2}) \quad (1.92)$$

In the case of U(N), under the same conditions, the U(1) charge factorizes, giving a theta function

$$Z_{G=1}^{U(N)} = \theta_3(e^{-A})\eta(e^{-A/2}) \quad (1.93)$$

The U(N) theory has a natural interpretation in terms of free fermions [16]; as we have already seen, the simplest way to motivate this relation is a close look to the Weyl character formula. For U(N), we have

$$\chi_h(x_i) = \frac{det(e^{ix_i h_j})}{det(e^{ix_i h_j^0})} \quad (1.94)$$

where now x_i are independent angles. We start with the hamiltonian on the group U(N)

$$H = \sum_a E^a E^a \quad (1.95)$$

where $E^a = TrT^a U \partial/\partial U$. The invariant measure is again related to the denominator of Weyl formula, the vacuum wavefunction, which has a simple expression in terms of the variables $z_i = e^{ix_i}$,

$$\psi_\rho(z_i) = \Pi_{i<j}(z_i - z_j) \quad (1.96)$$

(the Vandermonde determinant). If R is the matrix needed to diagonalize U, we have the jacobian factor

$$dU = dRdx_i|\Delta(z)|^2 = dRdx_i\tilde{\Delta}(z)^2 \quad (1.97)$$

where $\tilde{\Delta}(z) = \Pi_{i<j}\sin(x_i - x_j)/2$. It is standard [19] to compute the hamiltonian in the sector independent of R,

$$H = -\sum \frac{1}{\tilde{\Delta}} \frac{d^2}{dx_i^2} \tilde{\Delta} - \frac{N(N-1)(N+1)}{12} \quad (1.98)$$

We are interested in the singlet sector corresponding to T/W; the space of functions of the eigenvalues x_i invariant under permutations; the wavefunctions of H in this sector are the characters $\chi(U)$; after redefining $\chi \to \tilde{\Delta}\psi$ we end with a free hamiltonian with totally antisymmetric wavefunctions. A system of N free fermions, defined on a circle.

The energy is, up to a constant,

$$\sum_i h_i^2 \quad (1.99)$$

This is indeed true (as one can checks from the expression of the U(N) casimir given in the previous section) up to a term linear in $\sum h_i$ which has a trivial effect, as we will see in the next sections. In a second quantized formalism [16] we introduce the creator and destroying operators B_n, with $\psi(x) = \sum e^{inx} B_n = \sum z^n B_n$. Then $H = \int dx \partial\psi^* \partial\psi$.

The h_i are strictly decreasing integers, thus enforcing Pauli principle. We can expand oscillators near the positive and negative Fermi surface (see [16]), obtaining (in correspondence with Gross sector) left and right-moving modes,

$$\psi(z) = \sum z^{-n} b_n, \bar{\psi}(\bar{z}) = \sum \bar{z}^{-n} \bar{b}_n \tag{1.100}$$

with the standard commutation relation of a two-dimensional field theory.

An alternative basis in the space of the class functions is given by the Schur polynomials,

$$W_n = TrU^n = \sum_n z^n \tag{1.101}$$

As position operators they are expressed as

$$W_n = \int dz z^{-1-n} \psi^*(z)\psi(z) + \int d\bar{z} \bar{z}^{-1-n} \bar{\psi}^*(\bar{z})\bar{\psi}(\bar{z}) \tag{1.102}$$

and we recognize the standard bosonization procedure

$$W_n = \alpha_{-n} + \bar{\alpha}_n = \int d\theta e^{in\theta} \partial_\tau \phi(z, \bar{z}) \tag{1.103}$$

The hamiltonian in these variables reads [16]

$$H^{U(N)} = N(L_0 + \bar{L}_0) + \int dz z^2 : \partial\psi^* \partial\psi : + h.c. \tag{1.104}$$

while the U(1) charge is $Q = L_0 - \bar{L}_0$.

We see that the leading term in the large N expansion is the hamiltonian of a conformal free fermion; this may explain the appareance of modular function in the torus expansion.

The system of free fermions (on a line instead of a circle) is well known in c=1 string theory [20]; the previous construction is similar to that which conduces to a free relativistic fermion in that problem. Also the bosonization is a well known procedure in c=1, leading to the Das-Jevicki-Sakita hamiltonian (see also [4]).

The same step in our problem gives the bosonized hamiltonian

$$H = N(L_0 + \bar{L}_0) + \frac{1}{3} \int dz z^2 : (\partial\phi)^3 : + h.c. \tag{1.105}$$

The large N expansion can be obtained considering the cubic term as an interaction in a free conformal field theory. In this way, Douglas has computed the torus subleading term, finding a quasi modular function. He argued that some simple modification of QCD action [21] would lead to modular results on a torus at any order (see also [22]). An invariance under $A \to 1/A$ in QCD on a torus would be the analogous of T-duality $R \to 1/R$ in string theory.

1.6 QCD on a sphere

The main difference in extracting the leader order contribution to the sphere partition function is the appareance of the dimR in the Migdal's formula. Representations with casimir/N of order $O(N^2)$ are no longer suppressed by exponentially small terms with respect to the O(1) contribution of small tableaux, since the $\sum_R dimR^2$ often contributes $O(e^{N^2})$. In fact, the leading term of the expansion is of order $O(N^2)$, corresponding to the number of maps from the sphere to the sphere.

As an example, we can compute the partition function when we consider the casimir linearized to nN, like on the torus. A first derivation was given by Gross [2], using orthogonal polinomials. A quick derivation can be obtained using Schur polinomials, a method which we will employ often in the following.

Denoting by α, β the eigenvalues of the unitary matrices, we have that

$$S_k(\alpha) = TrU^k = \sum_{i=1}^{N} \alpha_i^k \tag{1.106}$$

(called Schur polynomials) are an alternative basis to the characters for the space of class functions. We have,

$$\begin{aligned}
&\sum_R \chi_R(\alpha)\chi_R(\beta) = \sum_{h_1>h_2>\ldots} \chi_h\chi_h = \frac{1}{N!}\sum_{h_i} \chi_h(\alpha)\chi_h(\beta) = \\
&\frac{1}{N!}\frac{1}{\Delta(\alpha)\Delta(\beta)}\sum_{h_i,P,P'}(-1)^P(-1)^{P'}\alpha_{P(1)}^{h_1}\ldots\alpha_{P(N)}^{h_N}\beta_{P'(1)}^{h_1}\ldots\beta_{P'(N)}^{h_N} = \\
&\frac{1}{N!}\sum(-1)^{P+P'}\Pi_i\frac{1}{1-\alpha_{P(i)}\beta_{P'(i)}}\frac{1}{\Delta(\alpha)\Delta(\beta)} = \\
&\frac{1}{\Delta(\alpha)\Delta(\beta)}det_{ij}\frac{1}{1-\alpha_i\beta_j}
\end{aligned} \tag{1.107}$$

Now, using Chauchy identity [18]

$$\frac{1}{\Delta(\alpha)\Delta(\beta)}det_{ij}\frac{1}{1-\alpha_i\beta_j} = \frac{1}{\Pi_{ij}(1-\alpha_i\beta_j)} \tag{1.108}$$

we find

$$\frac{1}{\Pi_{ij}(1-\alpha_i\beta_j)} = e^{-\sum ln(1-\alpha_i\beta_j)} = e^{\sum_k \frac{1}{k}\sum_{ij}(\alpha_i\beta_j)^k} = e^{\sum_k \frac{1}{k}S_k(\alpha)S_k(\beta)} \tag{1.109}$$

We conclude that the sum over SU(N) representations can be rewritten in terms of Schur polynomials as

$$\sum_R \chi_R(\alpha)\chi_R(\beta) = e^{\sum_k \frac{1}{k}S_k(\alpha)S_k(\beta)} \tag{1.110}$$

The sum with linearized casimir which we want to compute is

$$\sum_R \chi_R(\alpha)\chi_R(\beta)e^{-An/2} \tag{1.111}$$

It can be obtained by dilatating $\alpha \to e^{-A/2}\alpha$, since $\sum h_i = n + N(N-1)/2$,

$$\sum_R \chi_R(\alpha)\chi_R(\bar{\beta})e^{-An/2} = e^{N(N-1)/2}e^{\sum_k \frac{e^{kA/2}}{k} S_k(\alpha)S_k(\beta)} \tag{1.112}$$

The partition function on the sphere is obtained when $\alpha, \beta \to 0$, and since $S_k \to 0$, reads (up to a constant)

$$Z_{G=0} = e^{\sum_{k=0}^{\infty} \frac{e^{-Ak/2}}{k} N^2} = e^{-N^2 ln(1-e^{-A/2})} \tag{1.113}$$

Expanding in powers of $e^{-A/2}$,

$$Z_{G=0} = \frac{1}{(1-e^{-A/2})^{N^2}} = \sum_{n=0}^{\infty} e^{-An/2} C_n^{N^2+n-1}$$

$$\sum_R dim R^2 = \sum_{n=0}^{\infty} C_n^{N^2+n-1} \tag{1.114}$$

where C_n^k are the bynomial coefficients. From this formula we see that the sum over representations of the dimensions is dominated by terms with n of the order N^2, i.e. representations with n_i of order N. For this tableaux it is no longer right to suppress the formally subleading terms in the casimir with respect to the linear one.

The computation of the leading term of the 1/N expansion can be performed by various methods; the most surprising result is that there is a phase transition of third order at a certain value of the area [11] [2]. Despite the strong similarity, this phase transition is of completely different nature from the old third order phase transition of the lattice model for QCD [10]. In this section we will review this lattice transition, and what was believed true for the continuum theory, developing at the same time some techniques we will use often in the following.

We briefly review the formulation for Wilson QCD action.

$$S(U) = \sum_P \frac{1}{g^2} Tr(U_P + h.c.) \tag{1.115}$$

As discussed in section (1.1), this action has the right continuum limit when the lattice spacing goes to zero, but is not equivalent to QCD but only to an QCD discretization.

A convenient gauge is $A_0 = 0$, which means that $U_\gamma = 1$ for every temporal link γ.

Z can be easily evaluated by the change of variables

$$W_\gamma = U_{\gamma+e_0}^{-1} U_\gamma \tag{1.116}$$

where e_0 is the lattice versor in the temporal direction.

$$Z = \int \Pi_\gamma dW_\gamma e^{\sum_\gamma \frac{1}{g^2} Tr(W_\gamma + h.c.)} \tag{1.117}$$

factorizes in L^2 copies of the unitary matrix model,

$$Z = \int [dU] e^{\frac{1}{g^2} Tr(W + W^{-1})} \tag{1.118}$$

The solution of this matrix model follows the standard steps [10]; diagonalizing the unitary matrix U, we find

$$Z = \int_0^{2\pi} \Pi_i d\alpha_i \Delta(\alpha)^2 e^{\frac{2}{g^2} \sum_i^N \cos\alpha_i} \tag{1.119}$$

We are interested in the large N limit for $\lambda = g^2 N$ fixed.

We can employed the sleepest-descendent method to the action

$$\frac{2}{g^2} \sum_{i=0}^{N} \cos\alpha_i + \sum_{i \neq j} log \left| \sin \frac{\alpha_i - \alpha_j}{2} \right| \tag{1.120}$$

which has the stationarity condition

$$\frac{2}{g^2} \sin\alpha_i = \sum_{j \neq i} cot \left| \frac{\alpha_i - \alpha_j}{2} \right| \tag{1.121}$$

In the large N limit we use continuous variables $x = i/N, \alpha(x)$ for which

$$F(\lambda) = \frac{2}{\lambda} \int_0^1 dx \cos\alpha(x) + P \int_0^1 dx \int_0^1 dylog \left| \sin \frac{\alpha(x) - \alpha(y)}{2} \right| \tag{1.122}$$

where P is the principal part of the integral, and

$$\frac{2}{\lambda} \sin\alpha(x) = P \int_0^1 dycot \frac{\alpha(x) - \alpha(y)}{2} \tag{1.123}$$

Introducing the usual density of eigenvalues,

$$\rho(\alpha) = \frac{dx}{d\alpha} \geq 0$$
$$\int_{-\alpha_c}^{\alpha_c} \rho(\alpha) = \int_0^1 dx = 1 \tag{1.124}$$

we obtain the integral equation

$$\frac{2}{\alpha} \sin\alpha = P \int_{-\alpha_c}^{\alpha_c} d\beta \rho(\beta) cot \left(\frac{\alpha - \beta}{2} \right) \tag{1.125}$$

We allow for a symmetric range in which the angular eigenvalues are spread, $(-\alpha_c, \alpha_c)$, $\alpha_c \leq \pi$.

The model presents a phase transition when $\alpha_c = \pi$. A solution for $\alpha_c = \pi$ is easily found to be

$$\rho(\alpha) = \frac{1}{2\pi}[1 + (2/\lambda)\cos\alpha] \tag{1.126}$$

with $F(\lambda) = 1/\lambda^2$. Such solution is positive and physical only for $\lambda \geq 2$. The solution for $\alpha_c < \pi$ is obtained with the standard method of solution of such integral equations [23] [19], searching for a periodic analytic function in the cut complex plane which has ρ as jump singularity. One finds

$$\rho(\alpha) = \frac{2}{\pi\lambda}\cos\frac{\alpha}{2}\left(\frac{\lambda}{2} - \sin^2\frac{\alpha}{2}\right)^{1/2} \tag{1.127}$$

Note that this solution exists only when $\lambda < 2$, and for $\lambda = 2$ it equals the previous solution for the strong phase.

The Wilson loop of this model are clearly related to the moments of the measure

$$w_n = \frac{1}{N} < TrW^n >= \frac{1}{NZ}\int dW TrW^n e^{\frac{N}{\lambda^2}Tr(W+h.c.)} \tag{1.128}$$

which in the large N limit become

$$w_n = \int_0^{2\pi} \rho(\alpha)e^{in\alpha} \tag{1.129}$$

the Fourier components of the density ρ.

If we define the generating functional for the moments [24]

$$\Phi(t) = \frac{1}{N} < Tr\frac{1}{1-tW} >= \sum_{n=0}^{\infty} w_n t^n \tag{1.130}$$

which is an holomorfic function when $|t| \leq 1$ since $w_n \leq 1$, we find that

$$\rho(\alpha) = \frac{1}{\pi}Re\Phi(e^{i\alpha}) \tag{1.131}$$

We are far from the real QCD. This theory is not invariant under subdivision and so has a non trivial continuum limit, obtained by reiterated application of the renormalization group. In this way, new couplings appear in the action

$$S_{eff} = N\sum_{n=1}^{\infty}\frac{1}{\lambda_n}Tr(W^n + h.c.) \tag{1.132}$$

and to obtain the correct continuum limit we have to fit λ_n in the correct way when the lattice spacing goes to zero.

We are not able to find the explicit form of the action at the fixed point, but we know everything of the continuous theory through Migdal's formula and the loop equations. To know the Wilson loop, we would need the saddle point solution for the continuum action, of which we do not know even the form. However, by solving the large N loop equation one is able [25] to find the Wilson loop for a simple region of the plane of area A,

$$w_n = e^{-nA/2}\frac{1}{n}L^1_{n-1}(nA) \tag{1.133}$$

with generating functional defined implicitly by

$$A\Phi(t) - 2coth^{-1}2\Phi(t) = logt \tag{1.134}$$

It is now an interesting question to ask if this transition survives the continuum limit or it is only a lattice artifact. It happens that this transition is washed out by the continuum limit, but a new, different transition appears.

We are interesting in computing quantities in the continuum limit, to search for the existence of this transition. The simpler object to this purpose is the kernel on a disk [26],

$$\zeta(U,t) =< U|e^{-tH}|1 >= \sum_R dimR\chi_R(U)e^{-tc_R} \tag{1.135}$$

Note that the Wilson action was actually defined on the plane; the disk kernel is not so far, since for example we can compare Wilson loop for the two theories [25] for a disk of area $A = 1/\lambda$. Now, if we pose [27]

$$F(x_i,t) = \Delta(e^{ix_i})\zeta(x_i, t = A/2N) = \sum_R dimR\psi_R(e^{ix_i})e^{-tc_R} \tag{1.136}$$

We know that ψ_R is the wavefunction corresponding to the laplacian on the unitary group, with eigenvalue the casimir. By comparing the numerator of formula (1.94) (which is ψ_R) and formula (1.73) one can easily identify

$$H = -\sum \partial_i^2 + i(N-1)\partial_i - E_0 \tag{1.137}$$

And we obtain that F satisfies the slightly modified heat equation

$$\frac{\partial F}{\partial t} = \sum_i(\partial_i^2 F + i(N-1)\partial_i F) \tag{1.138}$$

The linear term can be reabsorbed by the shift $ix_i \to ix_i + A/2(N - 1/N)$ (for large N, $ix_i \to ix_i + A/2$) in the initial condition,

$$F(x_i, 0) = \sum_R dimR\psi_R(e^{ix_i+A/2(N-1/N)}) \tag{1.139}$$

The equation has the standard kernel

$$\frac{1}{\sqrt{t}}e^{-\frac{1}{4t}x^2} \tag{1.140}$$

and, with the given initial condition, we find the well-known result

$$F(x_i,t) = \frac{1}{t^{N/2}} \int dy e^{-\frac{1}{4t}(y_i-x_i)^2} F(y_i,0) \tag{1.141}$$

We see that most of the informations are encoded in the initial condition $F(x_i,0)$.

On the other hand, ζ satisfied

$$\begin{aligned}
&\partial_t \Delta(e^{ix_i})\zeta = -\sum \partial_i^2(\Delta\zeta) + i(N-1)\sum \partial_i(\Delta\zeta) - E_0\Delta\zeta = \\
&-\sum \partial_i(\partial_i\Delta\zeta + \Delta\partial_i\zeta) + i(N-1)\sum \partial_i\Delta\zeta + i(N-1)\sum \Delta\partial_i\zeta - E_0\Delta\zeta \\
&-2\sum \partial_i\Delta\partial_i\zeta - \Delta\sum \partial_i^2\zeta + i(N-1)\sum \Delta\partial_i\zeta
\end{aligned} \tag{1.142}$$

since $H\Delta = 0$ by definition,

$$\frac{\partial}{\partial t}\zeta = -\frac{1}{\Delta^2}\sum_i \partial_i(\Delta^2\partial_i\zeta) + i(N-1)\sum \partial_i\zeta \tag{1.143}$$

Introducing,

$$\begin{aligned}
&\Delta(e^{ix_i}) = (2i)^{N(N-1)/2} e^{(N-1)\sum ix_i/2} \Pi_{i<j} \sin\frac{(x_i - x_j)}{2} \\
&= (2i)^{N(N-1)/2} e^{(N-1)\sum ix_i/2} \tilde{\Delta}(e^{ix_i})
\end{aligned} \tag{1.144}$$

we easily verify the alternative formulas

$$H = -\frac{1}{\tilde{\Delta}^2}\sum_i \partial_i \tilde{\Delta}^2 \partial_i = \frac{1}{\tilde{\Delta}}\sum \partial_i^2 \tilde{\Delta} + R_o \tag{1.145}$$

where $R_0 = N(N^2-1)/12$. This is the standard laplacian over unitary matrices when restricted to the eigenvalues.

A crucial observation is that the laplacian over hermitian matrices, which has a simpler well-known kernel, is very similar,

$$\frac{1}{\Delta(x_i)}\sum \partial_i^2 \Delta(x_i) \tag{1.146}$$

Note the difference between $\Delta(x_i)$ for hermitian matrices and $\Delta(e^{ix_i})$ for unitary matrices.

Now, the evolution problem for hermitian matrices is the simple heat problem. The laplacian reads

$$D = \sum_i \frac{\partial^2}{\partial M_{ii}^2} + \frac{1}{2}\sum_{i<j}\left(\frac{\partial^2}{\partial ReM_{ij}^2} + \frac{\partial^2}{\partial ImM_{ij}^2}\right) \tag{1.147}$$

as a sum over quadratic derivatives. The fact that the metric on the matrix entries is trivial is the main difference with the unitary case. It can be considered the heat equation in R^{N^2} with variables $M_{ii}, ReM_{ij}, ImM_{ij}, i < j$; in this way we find the trivial kernel

$$Z(M_1, M_2) = < M_1|e^{-tD}|M_2 > = \frac{1}{(4\pi t)^{N^2/2}} e^{-\frac{1}{4t}tr(M_1-M_2)^2} \tag{1.148}$$

solution of the heat equation

$$(\partial_t - D)Z(M_1, M_2) = 0 \tag{1.149}$$

which reduces to a delta function for hermitian matrices when $t \to 0$. It is exactly the diffusion problem in an euclidean space, since the metric on the lie algebra of hermitian matrices is flat.

Now the kernel for unitary matrices can be obtained by comparing the two laplacian

$$H = -\frac{\Delta(x_i)}{\tilde{\Delta}(e^{ix_i})} D \frac{\tilde{\Delta}(e^{ix_i})}{\Delta(x_i)} + R_o \tag{1.150}$$

So we find the solution

$$\zeta(x_i, t) = \frac{1}{(4\pi t)^{N^2/2}} \frac{\Delta(x_i)}{\tilde{\Delta}(e^{ix_i})} e^{-\frac{1}{4t}\sum_i x_i^2 + R_0 t} \tag{1.151}$$

which is not periodic; the true solution will be [26]

$$\zeta(x_i, t) = \frac{1}{(4\pi t)^{N^2/2}} \sum_{n_i} \prod_{i<j} \frac{x_i - x_j + 2\pi(n_i - n_j)}{\sin(x_i - x_j + 2\pi(n_i - n_j)/2)} e^{-\frac{1}{4t}\sum_i (x_i + 2\pi n_i)^2 + R_0 t} \tag{1.152}$$

The partition function on the sphere can be obtained in the limit $x_i \to 0$; the terms with $n_i \neq 0$ are suppressed by exponentially small contributions $e^{-n_i^2 N/2A}$ (recall t=A/2N), so the leading term is that with $n_i = 0$, and we find

$$Z_{sphere} = e^{-N^2(logA/2 - A/24)} \tag{1.153}$$

up to exponentially small terms.

Note also that if we exclude these exponentially small terms we find the complete 1/N expansion which ends after the first two terms

$$F(A) = N^2 \frac{logA}{2} - \frac{N^2 - 1}{2} \frac{A}{24} \tag{1.154}$$

We give an alternative derivation for Menotti-Onofri formula, which makes clear its relation with the Poisson resummation. Our starting point will be formula (1.59).

We start from the cylinder kernel

$$\sum_R \frac{\psi_R(x_i)\psi_R(y_i)}{\Delta(e^{ix_i})\Delta(e^{iy_i})} e^{-\frac{A}{N}(\sum h_i^2 - (N-1)\sum h_i - E_0)} \tag{1.155}$$

The representations R are labelled by the decreasing integers h_i. Using the tricks explained at the end of section (1.3), we can write first the sum over all the integers, by dividing by N!, write the wavefunctions as determinants or sum over permutation

$$\psi_R(x_i) = \sum_P (-1)^P e^{ix_i h_{P(i)}} \tag{1.156}$$

eliminate one of the sum over permutation by using the symmetry in h_i, cancelling N!, obtaining the equivalent for U(N) of formula (1.59)

$$\frac{1}{\Delta(e^{ix_i})\Delta(e^{iy_i})} \sum_{P,h_i} (-1)^P e^{ih_i(x_i - y_{P(i)}) - t\sum h_i^2 + t(N-1)\sum h_i + tE_0} \tag{1.157}$$

performing a Poisson resummation (up to moltiplicative constants independent by x_i, t, as in the following)

$$\frac{1}{\Delta(e^{ix_i})\Delta(e^{iy_i})} \sum_{P,r_i} (-1)^P e^{-\frac{1}{4t}(2\pi r_i - x_i + y_{P(i)} + it(N-1))^2 + tE_0} \tag{1.158}$$

Collecting all the non vanishing terms, we find the general form of the cylinder kernel (see also [17])

$$\frac{1}{\Delta(e^{ix_i})\Delta(e^{iy_i})} \sum_{P,h_i} (-1)^P e^{A(N^2-1)/24 + i(N-1)(\sum x_i - \sum y_i)/2 - i\pi(N-1)\sum r_i}$$
$$e^{-\frac{1}{4t}(2\pi r_i - x_i + y_{P(i)})^2} \tag{1.159}$$

Now we rewrite the determinant in terms of the Itzkynson-Zuber integral [28], via the formula

$$\int dU e^{-\frac{1}{4t}tr(X_1 - UX_2U^+)^2} = (2t)^{N(N-1)/2} \frac{1}{N!} \prod_i^N p! \frac{det e^{-\frac{1}{4t}(\lambda_{1,i} - \lambda_{2,j})^2}}{\Delta(x_1)\Delta(x_2)} \tag{1.160}$$

for X_1, X_2 diagonal matrices with entries x_1, x_2. This formula is easily proved by considering a solution of the hermitian diffusion equation

$$g(M_1,t) = \int dM_2 e^{-\frac{1}{4t}tr(M_1 - M_2)^2} g(M_2, 0)$$
$$g(x_1,t) = \int dU \int dx_2 e^{-\frac{1}{4t}tr(X_1 - UX_2U^+)^2} g(x_2)\Delta^2(x_2) \tag{1.161}$$

Now, exactly as in the unitary case,

$$\zeta(x_1,t) = \Delta(x_1) g(x_1,t) \tag{1.162}$$

satisfies the heat equation

$$\frac{\partial}{\partial t}\zeta = \sum_i \partial_i^2 \zeta \tag{1.163}$$

and is requested to be antisymmetric, so has the kernel

$$\frac{1}{(4\pi t)^{N/2}}\frac{1}{N!}\sum_P(-1)^P e^{-\frac{1}{4t}\sum_i(x_{1,i}-x_{2,P(i)})^2} = \frac{1}{(4\pi t)^{N/2}}\frac{1}{N!}dete^{-\frac{1}{4t}(x_{1,i}-x_{2,j})^2} \tag{1.164}$$

By comparing with formula (1.161) we find (1.160).

Collecting all the terms, we can rewrite the cylinder kernel in the form

$$\sum_{r_i} t^{N^2/2}e^{A(N^2-1)/24+i(N-1)(\sum x_i-\sum y_i)/2-i\pi(N-1)\sum r_i}$$
$$\frac{\Delta(x_i+2\pi r_i)\Delta(y_i)}{\Delta(e^{ix_i})\Delta(e^{iy_i})}\int dU e^{-\frac{1}{4t}tr(2\pi R-X+UYU^+)^2} \tag{1.165}$$

where R is an unitary matrix with integer entries r_i. Now recalling

$$\Delta(e^{ix_i}) = (2i)^{N(N-1)/2}e^{(N-1)\sum ix_i/2}\Pi_{i<j}\sin\frac{(x_i-x_j)}{2} \tag{1.166}$$

and

$$e^{i\pi(N-1)\sum r_i}\prod_{i<j}\sin\left(\frac{x_i-x_j}{2}\right) = \sin\left(\frac{x_i-x_j+2\pi(r_i-r_j)}{2}\right) \tag{1.167}$$

we find

$$\sum_{r_i} e^{A(N^2-1)/24-N^2logA/2}\frac{\Delta(x_i+2\pi r_i)\Delta(y_i)}{\tilde{\Delta}(e^{ix_i+2\pi r_i})\tilde{\Delta}(e^{iy_i})}\int dU e^{-\frac{1}{4t}tr(2\pi R-X+UYU^+)^2} \tag{1.168}$$

In the limit $x_2 \longrightarrow 0$ the dependence on U disappears giving a the volume of the unitary group, the two Vandermonde cancel and we easily recover the Menotti-Onofri formula

$$\zeta(x_i,t) = e^{A(N^2-1)/24-N^2logA/2}\sum_{r_i}\prod_{i<j}\frac{x_i-x_j+2\pi(r_i-r_j)}{\sin(x_i-x_j+2\pi(r_i-r_j)/2)}$$
$$e^{-\frac{1}{4t}\sum_i(x_i+2\pi r_i)^2} \tag{1.169}$$

What's about the phase transition? As we have already observed, if we neglect exponentially small terms we find $e^{A(N^2-1)/24-N^2logA/2}$ and no trace for a singularity in A. But we can consistently neglect these exponentials? We have to take the limit $x_i \longrightarrow 0$ with care; if we set $x_i = 0$ in the exponential the symmetry in r_i cancel the dependence on r_i in the Vandermondes and we would find the incorrect result

$$e^{A(N^2-1)/24-N^2logA/2}\sum_{r_i}(-1)^{\sum r_i}e^{-\frac{1}{4t}\sum(2\pi r_i)^2} \tag{1.170}$$

We have to expand the exponential in power series of x_i, picking up further terms; one can work out the case of N=2 as an example.

$$\sum_{r_1,r_2}\left(1+2\pi\frac{r_1-r_2}{x_1-x-2+\ldots}\right)(-1)^{r_1+r_2}e^{-4\pi^2(r_1^2+r_2^2)/t}(1-4\pi(x_1r_1+x_2r_2)/t+\ldots) \tag{1.171}$$

The pole $1/x_1 - x_2$ must cancel; it happens due to the symmetry in r_i, but it leaves a finite contribution

$$\frac{-8\pi^2}{x_1 - x_2} \sum_{r_i} (r_1 - r_2)(-1)^{r_1+r_2} e^{-4\pi^2(r_1^2+r_2^2)/t}(x_1 r_1 + x_2 r_2)/t =$$

$$\frac{-8\pi^2}{(x_1 - x_2)t} \sum_{r_i} r_1^2 (-1)^{r_1+r_2} e^{-4\pi^2(r_1^2+r_2^2)/t} \tag{1.172}$$

and we find the result

$$e^{A(N^2-1)/24 - N^2 logA/2} \sum_{r_i} (1 - 8\pi^2 r_i^2/t)(-1)^{r_1+r_2} e^{-\frac{1}{4t}\sum(2\pi r_i)^2} \tag{1.173}$$

For general N we expect

$$e^{A(N^2-1)/24 - N^2 logA/2} \sum_{r_i} (-1)^{\sum r_i} P(r_i) e^{-\frac{1}{4t}\sum(2\pi r_i)^2} \tag{1.174}$$

where P is a polynomial in r_i with coefficients depending on A and N; now for large N we expect that the degree of the polynomial grows. It may happen that the polynomial resum to give an exponential contribution able to overcome the suppressing exponential. If we are not able to evaluate the complete expression, we can not garanteer that the exponentially small corrections will remain small in the large N limit, and so we can not say anything about phase transition from this formula. We will see in the next section that the corrections are indeed negligible only if $A \leq \pi$, the weak coupling phase where

$$Z = e^{A(N^2-1)/24 - N^2 logA/2} \tag{1.175}$$

For $A = \pi$ a phase transition occurs.

Before concluding this section, we want to make some remark about formula (1.141). At the beginning of the section we have computed the initial condition $F(x_i, 0) = \sum dimR\psi_R(x_i + A/2)$ in the case in which the representations are restricted to $h_i > 0$, the polynomial representations of U(N); with this restriction F is no longer suppose to enforce a delta function condition on the group, but reads

$$F(x_i, 0) = \sum_R dimR\psi_R(e^{ix_i + A/2}) = e^{-N(\sum log(1-e^{ix_i+A/2})} \Delta(e^{ix_i+A/2}) \tag{1.176}$$

We will diffusely consider this case in the following.

Taking into account all the U(N) representations, F becomes proportional to a delta function, since we have the ortonormality condition

$$\sum_R dimR\chi_R(U) = \delta(U - 1) \tag{1.177}$$

As we will see in section (1.9), a convenient way to enforce such a constrain is

$$F(x_i, 0) = \sum_R dimR\psi_R(e^{ix_i+A/2}) = e^{-N(\sum log(1-e^{ix_i+A/2}) + c(ix_i + A/2))} \Delta(e^{ix_i+A/2}) \tag{1.178}$$

in the limit $c \to \infty$.

1.7 QCD on a sphere: a large N phase transition

In this section we want to show how the phase transition occurs [11] [29].

The partition function for U(N) is more simply expressed in terms of the integers (semi-integers for N even) $h_i = n_i - i + (N+1)/2$,

$$Z_{G=0} = \sum_{h_1>h_2>\dots} \Pi_{i<j}(h_i - h_j)^2 e^{-\frac{A}{2N}\sum h_i^2 - \frac{A(N-1)(N+1)}{24}} \quad (1.179)$$

Note that these integers are not exactly the same we have encountered in the previous section; however, the expression of dimR does not change. By dividing by $N!$ we can enlarge the sum to arbitrary integers.

In the large N limit [11] we can use continuum variables $n(x) = n_i/N, x = i/N$ and write

$$Z_{G=0}(A) = \int Dh(x) e^{-N^2 S[h(x)]} \quad (1.180)$$

where

$$S[h(x)] = \frac{A}{2}\int_0^1 dx h^2(x) - \int_0^1 dx \int_0^1 dy log|h(x) - h(y)| - \frac{A}{24} \quad (1.181)$$

This procedure amounts to forget the discrete character of h variables in formula (1.179)

$$\int_{-\infty}^{\infty} dh_i \Pi_{i<j}(h_i - h_j)^2 e^{-\frac{A}{2N}\sum h_i^2 - \frac{A(N-1)(N+1)}{24}} \quad (1.182)$$

Recalling the expression of the measure on hermitian matrices in terms of eigenvalues and angular variables,

$$dM = dRdx_i\Delta(x_i)^2 \quad (1.183)$$

what we are computing is simply the gaussian hermitian matrix model,

$$Z_{G=0} = \int dM e^{-\frac{A}{2N}TrM^2} \quad (1.184)$$

The solution of such a matrix model is well known [19]. We introduce the density,

$$\rho(h) = \frac{\partial x(h)}{\partial h}$$
$$\int dh\rho(h) = 1 \quad (1.185)$$

Since we have a large parameter N^2 in front of the action, we can apply the saddle point approximation, finding for ρ the integral equation

$$\frac{A}{2}h = P\int ds \frac{\rho(s)}{h-s} \quad (1.186)$$

The standard procedure to solve this type of equation [19] gives the semicircle law of Wigner

$$\rho(h) = \frac{A}{2\pi}\sqrt{\left(\frac{4}{A} - h^2\right)} \tag{1.187}$$

For the derivative of the free energy we obtain

$$F\prime(A) =< \frac{1}{2N}\int h^2 > -\frac{1}{24} = \frac{1}{2}\int \rho(h)h^2 - \frac{1}{24} = \frac{1}{2A} - \frac{1}{24} \tag{1.188}$$

or

$$F(A) = \frac{A}{24} - \frac{1}{2}logA \tag{1.189}$$

as we could also have found by rescaling the field h in the action. This is the result of the previous section, when exponentially small terms are neglected.

In this way we have completely forgot the discrete nature of the variables of our problem. Just this discreteness leads to a complete different behaviour of the partition function for large A. Kazakov and Douglas [11] proposed to take into account the discreteness of h via,

$$h_i - h_j \geq i - j \tag{1.190}$$

which, in the continuum limit, becomes

$$\begin{aligned} &\frac{h(x) - h(y)}{x - y} \geq 1 \\ &\rho(h) \leq 1 \end{aligned} \tag{1.191}$$

Wigner solution satisfies this constrain only if

$$A \leq A_{crit} = \pi^2 \tag{1.192}$$

For $A \geq \pi^2$ we have to find another solution with $\rho \leq 1$. If we admit that $\rho = 1$ in a certain central interval around $h = 0$, which means that a certain number of central n_i is zero, we are led to search a solution in which ρ is defined in two separated intervals; it is a two-cut solution [30], more difficult to determine and expressed in terms of elliptic functions.

Kazakov and Douglas found in this way a solution for the strong phase which has a third order phase transition at $A = \pi^2$. Their solution has the correct behaviour for $A \to \infty$; the free energy has an expansion in terms of powers and exponentials of A, as proposed by Gross string interpretation.

An alternative way to understand this phase transition is via orthogonal polynomials [29]. We have found

$$Z_{G=0} = \frac{1}{N!}\sum_{h_i=-\infty}^{\infty} \Delta(h_i)^2 e^{-\frac{A}{2N}\sum h_i^2 - \frac{A(N^2-1)}{24}} \tag{1.193}$$

Introducing a complete set of discrete polynomials orthogonal with respect to the measure

$$\sum_{x=0}^{\infty} e^{-\alpha x^2} P_n(x)P_m(x) = \delta_{nm}h_n(\alpha) \tag{1.194}$$

with $P_n = x^n + ...$, it is a standard trick to write

$$\Delta(x) = det_{ij}(P_{j-1}(x_i)) \tag{1.195}$$

and, expanding the determinants, to find

$$Z_{G=0} = \frac{1}{N!} e^{-\frac{A}{24}(N^2-1)} \Pi_{j=0}^{N-1} h_j(A/2N) \tag{1.196}$$

From the standard recursion relation

$$xP_n(x) = P_{n+1}(x) + S_n P_n(x) + R_n P_{n-1} \tag{1.197}$$

Taking scalar products with P_k, we find, taking also into account the symmetry of measure, that lower degree polynomials do not enter the right hand side of the equation, and $S_n = 0$, $R_n = h_n/h_{n-1}$.

Differentiating with respect to α the equation

$$h_n(\alpha) = \sum e^{-\alpha x^2} P_n^2(x, \alpha) \tag{1.198}$$

we get

$$\begin{aligned} &\frac{dh_n}{d\alpha} = \sum e^{-\alpha x^2}[2P_n \frac{dP_n}{d\alpha} - x^2 P_n^2] = \\ &- < xP_n | xP_n >= - < P_{n+1} + R_n P_{n-1} | P_{n+1} + R_n P_{n-1} >= \\ &-h_n(R_{n+1} + R_n) \end{aligned} \tag{1.199}$$

where we have use the fact that $dP_n/d\alpha$ is a polynomial of order n-1 and so it is orthogonal to P_n. Finally,

$$R_{n+1}(\alpha) = -(R_n(\alpha) + \frac{d}{d\alpha} log h_n(\alpha)) \tag{1.200}$$

and, eliminating $h_n(\alpha)$,

$$\begin{aligned} R_{n+1} - R_{n-1} &= -\frac{d}{d\alpha} log R_n(\alpha) \\ R_0(\alpha) = 0, h_0(\alpha) &= \sum e^{-\alpha x^2} \end{aligned} \tag{1.201}$$

In the large N limit we find via Poisson resummation formula that

$$h_0(A/2N) = (\frac{2\pi N}{A})^{1/2} \sum_n e^{-2\pi^2 N n^2/A} = (\frac{2\pi N}{A})^{1/2} + ... \tag{1.202}$$

and we easily solve the recursion equation finding that

$$R_n^0(A/2N) = \frac{nN}{A} \tag{1.203}$$

We would have obtained the same result with continuous measure, where P_n would be the well-known Hermite polynomials.

From

$$F(A) = -\frac{A}{24}(N^2-1) + Nlogh_0 + \sum_{i=1}^{N-1}(N-i)logR_i \tag{1.204}$$

we find the known result

$$F(A) = -\frac{A}{24}(N^2-1) - \frac{N^2}{2}logA \tag{1.205}$$

The discreteness of our measure reveals itself in the exponential small corrections to h_0 we have neglected. It happens that they are negligible only when $A \le \pi^2$. $h_0(\alpha)$ can be written, via Poisson resummation formula, as a sum of $e^{-2\pi^2 n^2 N/A}$; this expansion is clearly related to the expansion in Menotti-Onofri formula of the previous section. These exponential corrections can grow in the iteration and to be not longer small. For example, the contribution of the first correction can be found expanding

$$R_n(A/2N) = \frac{nN}{A} + c_n(A)e^{-\frac{2\pi^2 N}{A}} \tag{1.206}$$

Defining

$$\eta = \frac{2\pi^2 N}{A}$$
$$c_n(A) = \frac{2}{\pi^2}\eta^2 G_n(\eta) \tag{1.207}$$

the recursion relation

$$G_{n+1}(\eta) - G_{n-1}(\eta) = -\frac{2}{n}[(\eta-1)G_n(\eta) - \eta G_n\prime(\eta)] \tag{1.208}$$

can be translated into a differential partial equation for the generating function $F(\eta, z) = \sum_{n=1}^{\infty} z^n G_n(\eta)$, with solution

$$F(\eta, z) = \frac{z}{(1-z)^2} e^{-\frac{2\eta z}{1-z}} \tag{1.209}$$

It is the generating function for generalized Laguerre polynomials,

$$G_n(\eta) = \int \frac{dz}{2\pi i}\frac{1}{z^{n+1}}F(\eta,z) = \int \frac{dt}{2\pi i}e^{-2\eta t}\left(1+\frac{1}{t}\right)^n = L^1_{n-1}(2\eta) \tag{1.210}$$

Combining all pieces,

$$R_n(A/2N) = \frac{nN}{A} - \frac{2}{\pi^2}\left(\frac{2\pi^2 N}{A}\right)^2 L^1_{n-1}\left(\frac{4\pi^2 N}{A}\right) e^{-2\pi^2 N/A} \tag{1.211}$$

To compute the partition function we need all the R_n up to $n = N$ (see (1.204)). When n is of order N, the n degree polynomials $L^1_n(cost.N)$ can contribute a factor N^N, able to overcome the exponentially small term in this formula.

The large N asymptotics of the Laguerre polynomials can be obtained by the integral formula (1.210) via a saddle point approximation. One easily finds

$$R_n = \frac{nN}{A} + (-1)^n \left(\frac{2n}{\pi^5}\right)^{1/2} n_{cr} \left(1 - \frac{n}{n_{cr}}\right)^{-1/4} e^{-\frac{2\pi^2 N}{A}\gamma(\frac{n}{n_{cr}})} \tag{1.212}$$

where

$$\gamma(x) = \sqrt{1-x} - \frac{x}{2} log \frac{1+\sqrt{1-x}}{1-\sqrt{1-x}} > 0 \tag{1.213}$$

and $n_{cr} = N\pi^2/A$.

The singularity for $n = n_{cr}$ is the signal of the phase transition. We see that the exponentially small term of the initial condition remains small if $n < n_{cr}$; for $n = n_{cr}$ the approximation is no longer valid. The partition function requires all the R_n for $n < N$. So, if $N < n_{cr}$ the system is in the weak coupling phase. The critical point corresponds to $A_{cr} = \pi^2$, exactly the value found by Kazakov and Douglas.

Using formula (1.204), with some manipulations we find

$$F(A,N) = -\frac{N^2}{2}\left(\frac{A}{12} + logA\right) + \frac{A}{24} + \\ \frac{(-1)^{N+1}}{\sqrt{2\pi N}} \frac{A}{2\pi^2} \left(1 - \frac{A}{\pi^2}\right)^{-1/4} e^{-\frac{2\pi^2 N}{N}\gamma(\frac{A}{\pi^2})} + \dots \tag{1.214}$$

We see that the 1/N expansion is confined to the first two terms $O(N^2), O(1)$, as we already known; non trivial effects are incorporated in the nonperturbative corrections to the free energy.

Let's see how this non-perturbative contribution can be reconduced to to an instanton contribution to the partition function. What is true in two-dimensional QCD is that the partition function can be computed as a sum over localized instantons [13]; the semiclassical approximation about all the classical solutions of the theory gives indeed the complete answer, even if the instantons are unstable. This was shown by Witten [13], using the link with topological Yang-Mills theory and the non-abelian Duistermaat-Heckman theorem.

On a two-dimensional sphere the classical instanton gauge configuration are diagonal matrices with entries proportional to the Dirac monopole potential

$$A_\theta(\theta,\phi) = 0, A_\phi(\theta,\phi) = \frac{1-\cos\theta}{2} \tag{1.215}$$

the proportionality coefficients are N arbitrary integers, and the corresponding action is easily computed to

$$S_{int}(n_1, n_2, \dots, n_N) = \frac{2\pi^2 N}{A} \sum_{i=1}^{N} n_i^2 \tag{1.216}$$

These instantons are unstable, but according to Duistermaat-Heckman theorem the exact partition function of two-dimensional QCD is given by the sum over all instanton contributions

$$Z_{G=0} = \sum_{n_1,n_2,...=-\infty}^{\infty} w(n_1,...,n_N)e^{-S(n_1,...,n_N)} \quad (1.217)$$

where $w(n_1,...,n_N)$ are the perturbative contributions about the instanton.

That QCD has indeed this expansion follows from what we have said about Menotti-Onofri formula in the previous section and, at the end of the story, from the Poisson resummation formula. There is a more convenient way (with respect to that used in the previous section) to perform the Poisson resummation [4].

To use Poisson resummation in the standard form, we have to compute the Fourier transform of

$$\Delta(p)^2 e^{-\alpha \sum p_i} \quad (1.218)$$

Since it is simple to compute

$$\begin{aligned} &\int dx_i e^{-ix_ip_i}\Delta(p)e^{-\alpha\sum p_i} = \delta(i\partial_{x_i})\int dx_i e^{-ix_ip_i}e^{-\alpha\sum p_i} \\ &P(x_i)\int dx_i e^{-ix_ip_i}\Delta(p)e^{-\alpha\sum p_i} = \\ &cost\Delta(x_i)\int dx_i e^{-ix_ip_i}\Delta(p)e^{-\alpha\sum p_i} \end{aligned} \quad (1.219)$$

(the Vandermonde is the only antisymmetric polynomial of order N(N-1)/2) we consider our expression as the product of two such simple expression, finding a convolution

$$\int dy_i \Delta(x_i+y_i)\Delta(x_i-y_i)e^{-\frac{1}{2\alpha}((x_i+y_i)^2+(x_i-y_i)^2)} \quad (1.220)$$

At the end of the computation we find,

$$w(n_1,...,n_N) = \int_{-\infty}^{\infty} \Pi_{i<j}((y_i-y_j)^2 - 4\pi^2(n_i-n_j)^2)e^{-\frac{N}{2A}\sum_{i=1}^N y_i^2} \quad (1.221)$$

It is a non trivial check that the computation of $w(1,0,...,0)$ performed by Gross [29] by a saddle point method gives exactly the result (1.214).

From

$$w(1,0,...,0) = \int dy_1 e^{-N/2A}\int \prod_{i=2}^N dy_i \Delta^2(y_2,...,y_N)e^{-\frac{N}{2A}\sum_{i=2}^N y_i^2}\prod_{i=2}^N(4\pi^2-(y_i-y_1)^2) \quad (1.222)$$

we see that we can consider the variable y_1 as immersed in a system of random gaussian variables $y_2,...,y_N$. Since the insertion in the integral is of exponentially subleading order O(N), $y_2,...,y_N$ distribute according to Wigner law,

$$<\prod_{i=2}^N(4\pi^2-(y_i-y_1)^2)> = e^{N\int dy\rho(y)ln(4\pi^2-(y-y_1)^2)} \quad (1.223)$$

where the average is taken and normalized with respect to a N-1 hermitian gaussian system. The integration over y_1 is dominated by $y_1 = 0$ and w(0) is the partition function of a N hermitian gaussian system. Putting all together, we evaluate

$$\frac{w(1,0,...)}{w(0,0,...)} = (-1)^{N-1} e^{\frac{2\pi^2}{A}N - \frac{2\pi^2}{A}N\gamma(A/\pi^2)} \tag{1.224}$$

which reproduce (1.214).

So the exponentially correction to the free energy is the contribution of the one instanton sector.

Summarizing what we have seen: the partition function of QCD, expressed as a sum with genus dependent coefficients of gaussian exponential $e^{-An^2/N}$ can be consistently translate by Poisson resummation in a sum of $e^{-2\pi^2 n^2 N/A}$, interpreted as the contribution of the instanton of sector n. The contribution of the zero instanton sector is the weak coupling result, corresponding to forget the discrete character of the variables in our problem. Their discreteness can be taken into account by summing over the non trivial instanton sectors. Such contribution is exponentially small if $A \leq \pi^2$; when $A = \pi^2$ the instantons condensate to give a third order phase transition.

1.8 Coulomb gas on an half line

In this section we want to solve a related model [31], obtained by restricting the sum over integers h_i to the positive ones, i.e. restricting to the polynomial representation of U(N)

$$\sum_{h_1 > h_2 > ... > 0} \Pi_{i<j}(h_i - h_j)^2 e^{-\frac{A}{2N}\sum h_i^2} \tag{1.225}$$

See in the alternative way,

$$\sum_{h_i} e^{-[\frac{A}{2N}\sum_i h_i^2 - \sum_{ij} log|h_i - h_j|)]} \tag{1.226}$$

QCD on the sphere can be interpreted as the statistical partition function of N particles on the discrete line subjected to a Coulomb force and an harmonic potential. Our model corresponds to the same system on the half line with a barrier in h=0.

In formula (1.225) we can extend the sum over all the positive h by dividing by $N!$. In the form given we can use the same trick of the previous sections to compute it.

The stationary Coulomb gas [32] is a gas of N point charges with positions $h_1, h_2, ..., h_N$ free to move on the infinite straight line $-\infty \leq h \leq \infty$, immersed in a plane. The particles feel a harmonic potential which attracts them towards $h = 0$ and a logarithmic electrostatic repulsion.

The partition function

$$Z = \int_{-\infty}^{\infty} dh_i e^{-t\sum h_i^2 + \beta\sum_{i<j} ln|h_i - h_j|} \tag{1.227}$$

can be computed exactly for any β [32]. For specific values of β other physical quantities can be computed by the method of random matrices (unitary for $\beta = 2$, symplectic for $\beta = 4$, etc..).

It is possible to compute the partition function also in the case of the addition of an anharmonic interaction [19].

QCD_2 describes a stationary coulomb gas with $\beta = 2$ on a lattice line. We know that the model has two phases in the large N limit. In the weak coupling phase the result is the same of that of the gas on the continuum line up to exponentially small instanton correction.

If we want to study the gas within two walls, we have to enforce the constrain

$$|h_i| \leq cN \tag{1.228}$$

where c is a constant. The presence of N in the constrain is due to the fact that, due to the exclusion principle, the n particles cannot be compressed in an interval shorter than N, or, from the point of view of QCD, that the h_i were originally non decreasing integers.

We will study the stationary gas on a half-line, $h \geq 0$. In the next section we will see how to extend the result for the general line

$$h_i \geq -cN \tag{1.229}$$

For analogy with QCD, we solve the problem with

$$c(U(N)) = nN + \sum n_i(n_i - 2i + 1) = \sum h_i^2 - (N-1)\sum h_i - E_0 \tag{1.230}$$

with $E_0 = -N(N-1)(N-2)/6$.

$$Z = \frac{1}{\Delta(\alpha)\Delta(\beta)} \sum_{h_i} \psi_{h_i}(\alpha)\psi_{h_i}(\bar{\beta}) e^{-t(\sum h_i^2 - (N-1)\sum h_i - E_0)} \tag{1.231}$$

where we have to set $\alpha = \beta = 1$ at the end of the calculation.

In analogy with QCD, we expect the existence of a weak coupling phase, where the discreteness of model is lost up to instanton corrections and reveals only at a certain value of the area as a phase transition. We will compute quantities for the weak coupling phase.

Acting on $\psi_{k_i}(\alpha = e^{ix_i})$, H can be expressed in terms of derivatives with respect to x_i,

$$H = -\sum \partial_i^2 + i(N-1)\sum \partial_i - E_0 \tag{1.232}$$

obtaining

$$Z = \frac{1}{\Delta(\alpha)\Delta(\beta)} e^{-t(H-E_0)} \sum_{h_i} \psi_{h_i}(\alpha)\psi_{h_i}(\bar{\beta}) \tag{1.233}$$

Now we use the same trick of previous section (equation (1.110)) to sum over representations,

$$Z = \frac{1}{\Delta(\alpha)\Delta(\beta)} e^{-t(H-E_0)} \Delta(\alpha)\Delta(\bar{\beta}) e^{\sum \frac{1}{k} S_k(\alpha) S_k(\bar{\beta})} \tag{1.234}$$

H acts as a derivative operator on α. We can put $\beta = 1$ by the beginning and we will have to put $\alpha = 1$ at the end of the calculation. Moreover, we can take into account the linear term in the casimir by a dilatation of α, as in the previous sections. We find

$$Z = \frac{1}{\Delta(\alpha)} e^{-t(-\sum \partial_i^2 - E_0)} \Delta(\alpha) e^{N \sum \frac{e^{kA/2}}{k} S_k(\alpha)} \tag{1.235}$$

where, in the large N limit, we can set $(N-1)t = A/2$.

The presence of the quadratic operator H means that we are dealing with a heat kernel problem (see section (1.6)): by a double Fourier transform, we can write

$$e^{-t(-\sum \partial_i^2)} f(x_i) = \int dy e^{-\frac{1}{4t}\sum (x_i - y_i)^2} f(y_i) \tag{1.236}$$

Our function f is $e^{N\sum \frac{e^{kA+ikx_i}}{k}} = e^{-N\sum_i log(1-e^{iy_i+A/2})}$. Putting together all the terms,

$$Z = e^{-N^2/12} \frac{1}{\Delta(e^{ix_i})} \int \frac{dy_i}{t^{N/2}} e^{-N[\sum_i \frac{1}{2A}(x_i - y_i)^2 + log(1-e^{iy_i+A/2})]} \Delta(e^{iy_i}) \tag{1.237}$$

At the end of the computation we have to set $x_i = 0$. The integral over y_i is zero for symmetry when $x_i = 0$, and the leading term when $x_i \to 0$ must cancel the singularity $1/\Delta$ in front of the integral.

We show now how to treat with the Vandermonde insertion in the integral. Without such an insertion, the integral is over N decoupled variables, and we could compute it by a saddle point approximation, since we have the large N factor in front of the exponent.

If we are dealing with

$$e^{F(x)} = \int \frac{dy}{t^{N/2}} e^{-\frac{1}{4t}(x-y)^2 + F_0(y)} = e^{-\frac{1}{4t}x^2} \int \frac{dy}{t^{N/2}} e^{-\frac{1}{4t}y^2 + F_0(y) + \frac{x}{2t}y} \tag{1.238}$$

we have

$$\int \frac{dy}{t^{N/2}} e^{-\frac{1}{4t}(x-y)^2 + F_0(y)} \Delta(e^{iy_i}) = e^{-\frac{1}{4t}x^2} \Delta\left(e^{i2t\frac{\partial}{\partial x_i}}\right) e^{\frac{1}{4t}x^2 + F(y)} \tag{1.239}$$

Now, (recalling $h_i^0 = N - i$)

$$\Delta\left(e^{i2t\frac{\partial}{\partial x_i}}\right) f(x_i) = \sum_P (-1)^P e^{\sum_i h^0_{P(i)}\partial_i} f(x_i) = \sum_P (-1)^P f(x_i + ih^0_{P(i)}) \tag{1.240}$$

we find

$$\sum_P (-1)^P e^{ix_i h^0_{P(i)} - t(h^0_{P(i)})^2 + F(x_i + 2tih^0_{P(i)})} \tag{1.241}$$

In the limit $x_i \to 0$, we can expand the exponential to first order in x_i,

$$e^{\sum_i (-t(h_i^0)^2 + F(2ith_i^0))} \frac{\sum_P (-1)^P e^{ix_i(h^0_{P(i)} - i\partial_i F(2ith^0_{P(i)})}}{\sum_P (-1)^P e^{ix_i h^0_{P(i)}}} \tag{1.242}$$

now, the standard Weyl trick for the dimension of a representation, in the case in which F is the sum of identical function $F = \sum_i F(x_i)$, leads to

$$\Pi_{i<j}\frac{(h_i^0 - i\partial_i F(2ith_i^0) - h_j^0 + i\partial_j F(2ith_j^0))}{(h_i - h_j)} \tag{1.243}$$

Finally, in the large N limit, we can compute the sum by passing to continuous variables $x = i/N$, obtaining the definitive formula

$$e^{N^2[\int_0^1 dxF(iAx)+\frac{1}{2}\int_0^1 dx\int_0^1 dylog\left(1-i\frac{\partial F(ixA)-\partial F(iyA)}{x-y}\right)]} \tag{1.244}$$

Note that the factorization of F in a sum of identical functions of the various x_i is crucial; fortunately, this is indeed our case.

After having extracted the boring vandermonde factor, we can compute the integral over y_i as N identical decoupled integrals whose leading term can be found via saddle point method.

$$e^{\sum_{i=1}^N F(x_i)} = \Pi_i \int dy_i e^{-N[\frac{1}{2A}(x_i-y_i)^2+log(1-e^{iy_i+A/2})]} \tag{1.245}$$

The saddle point solution is

$$ix_i = iy - A\frac{1}{1-e^{-iy-A/2}} \tag{1.246}$$

which expresses the saddle point solution y in function of x_i via an implicit trascendent equation.

We find,

$$e^{F(x_i;A)} = e^{N[\frac{A}{2(1-e^{-i\tilde{y}-A/2})^2}-log(1-e^{i\tilde{y}+A/2})]} \tag{1.247}$$

where $\tilde{y}$ is the solution of equation (1.246).

Denoting $g(ix_i) = iy + A/2$, g is the inverse of the trascendent function

$$t = ix_i = g - \frac{A}{(1-e^{-g})} - A/2 \tag{1.248}$$

We choose the determination of the function which makes the logarithm well-defined.

To use formula (1.244) we have to substitute $t = ix_i = -xA$, so

$$F(x) = \frac{A}{2(1-e^{-g(-xA)})^2} - log(1-e^{g(-xA)}) \tag{1.249}$$

From the implicit equation,

$$\partial_{x_i} g = i\frac{dg}{dt} = i(1+\frac{Ae^{-g}}{(1-e^{-g})^2})^{-1} \tag{1.250}$$

we find

$$-i\partial_{x_i}F = -\frac{1}{1-e^{-g(-xA)}} \tag{1.251}$$

and, collecting all terms in formula (1.244),

$$F(A) = \frac{1}{A}\int_{-A}^{0} dx \left(\frac{A}{2(1-e^{-g(x)})^2} - log(1-e^{g(x)}) \right) +$$
$$\frac{1}{2A^2}\int_{-A}^{0} dx \int_{-A}^{0} dylog \left(1 + A\frac{\frac{1}{1-e^{-g(x)}} - \frac{1}{1-e^{-g(y)}}}{x-y} \right) + cost \quad (1.252)$$

It is a boring exercise to compute the small A limit of the partition function; first of all,

$$g(x) = -\sqrt{A} + \frac{A+x}{2} - \sqrt{A}\left(\frac{x}{4} + \frac{x^2}{4A} + \frac{A}{6}\right) + o(A^{3/2}) \qquad (1.253)$$

where we have taken into account that the integration range in formula (1.252) forces x to be of order A. Finally,

$$F(A) = -\frac{1}{2}logA + \frac{1}{2} + \frac{\sqrt{A}}{8} + ... \qquad (1.254)$$

The small A limit is probably the only significative quantity we can consistently consider, since we cannot control the point in which the phase transition, if exists, takes place. The ortogonal polynomials give equations so far intractable, and the form of the partition function is not convenient for a Poisson resummation. On the other hand, we might consider the continuum theory with the constrain $h(x) \geq 0$. The only solution of the equation (1.186) without poles (and cuts: weak phase!) is the Wigner semicircle and is necessarily symmetric about the origin and so ruled out. Solutions with a pole at the barrier $h = 0$ seem physically significative; since they contradict the constrain $\rho \leq 1$, we find no reasonable solution without cuts. It might happen that the model have only the strong phase.

If we move the barrier to the left ($c \neq 0$) when $c \leq -2/\sqrt{A}$, we can recover solutions for the weak phase. The case $c \neq 0$ will be considered in the next section.

1.9 An application to QCD

How to use the results of previous section for QCD? What we have computed is a hybrid partition function with the casimir corresponding to U(N) gauge theory and the sum over the SU(N) representations.

There is a simple trick to incorporate the lacking representations. In section (1.4), we have seen that

$$Z^{U(N)} = \sum_{[n]=0}^{N-1} \left(\sum_{\lambda} e^{-\frac{A}{2}(\lambda+\frac{n}{N})^2} \right) T_{[n]}^{SU(N)} \qquad (1.255)$$

We see that the λ of order greater than N, are exponentially suppressed by $O(e^{-N^3})$ or worse.

This suggests to bound the sum over λ in the following way

$$\begin{aligned} &\lambda \geq -cN \\ &c \to \infty \end{aligned} \tag{1.256}$$

For what regard the leading ($O(N^2)$) term in U(N) partition function, it is sufficient to compute

$$lim_{c\to\infty} \sum_{h_1>...>h_N=\lambda>-cN} \chi_{h_i}\chi_{h_i} e^{-\frac{A}{2N}c_h} \tag{1.257}$$

The main point is that

$$\sum_{h_i\geq -cN} \chi_h(\alpha)\chi_h(\bar{\beta}) = \frac{1}{(\Pi_i \alpha_i \bar{\beta}_i)^{cN}} \sum_{h\geq 0} \chi_h(\alpha)\chi_h(\bar{\beta}) \tag{1.258}$$

leads to the factor

$$e^{N\sum \frac{S_k(\alpha)}{k} - cN\sum x_i} \tag{1.259}$$

We have to compute a saddle point for a modified action,

$$e^{-N[\frac{1}{2A}(y_i-x_i)^2 + log(1-e^{iy_i+A/2}) + c(iy_i+A/2)]} \tag{1.260}$$

Looking at the new equation

$$A(c-x) = g - \frac{A}{1-e^{-g}} - \frac{A}{2} \tag{1.261}$$

we see that, for $c \to \infty$, $g \to 0$. Expanding in power series of $1/c$, we find

$$g = -\frac{1}{c} + \frac{1-x}{c^2} + \frac{1/2A - x^2 + 2x - 13/12}{c^3} + ... \tag{1.262}$$

and substituting in formula (1.252), we find with a little algebra that all the divergent powers and logarithms of c cancel, giving

$$F(A) = \frac{1}{2} logA - \frac{A}{24} + cost \tag{1.263}$$

the correct result for the weak coupling phase. With a saddle point method we have not taken into account possible non perturbative contributions and the phase transition.

We can compute with the same method the partition function for a disk. According to Migdal's formula, we have to take x_i different from zero, being the eigenvalues of the holonomy around the circle. Consequently, we can no longer use the trick to extract the Vandermonde and formula (1.244).

The saddle point equation

$$g - \frac{A}{1-e^{-g}} - \frac{A}{2} - Ac = ix_i \tag{1.264}$$

can again be expanded in powers of 1/c.

$$g_i = -\frac{1}{c} + \frac{a_i}{c^2} + \frac{b_i}{c^3} + \ldots$$
$$a_i = \frac{ix_i}{A} + 1$$
$$b_i = \frac{1}{A} + \frac{(ix_i + A/2)^2}{A^2} - \frac{ix_i + A/2}{A} - \frac{1}{3} \tag{1.265}$$

It is again true that

$$\frac{e^{-N\sum \frac{x_i^2}{2A}}}{\Delta(e^{ix_i})} \sum_P (-1)^P e^{-N(F(x_i+2tih^0_{P(i)})-(x_i+2tih^0_{P(i)})^2/2A)} \tag{1.266}$$

where now

$$F(x_i, A) = \frac{(x_i - iA/2)^2}{2A} - 1 - logc - \frac{1}{c}\left(\frac{ix_i}{A} + 1\right) + \frac{1}{c^2}\left(-\frac{x_i^2}{2A^2} + dx_i + e\right) + \ldots \tag{1.267}$$

The first term cancels exactly. All the other terms are linear or quadratic and can be extracted from the sum over P, except for the quadratic contribution $x_i h^0_{P(i)}$. Once extracted from the sum, all terms with powers of 1/c vanish; under the sign of sum it remains a term of order $1/c^2$. To leading order,

$$Z(a, x_i) = e^{-N^2\left(\frac{1}{N}\sum \frac{x_i^2}{2A} + \frac{i}{N}\sum \frac{x_i}{2} - \frac{A}{24} - 1 - logc\right)} \frac{\Delta(e^{\frac{ix_i}{Ac^2}})}{\Delta(e^{ix_i})} \tag{1.268}$$

But for $c \rightarrow \infty$,

$$\Delta(e^{\frac{ix_i}{Ac^2}}) = e^{-N^2 logAc^2/2}\Pi_{i<j}(ix_i - ix_j) \tag{1.269}$$

and, writing

$$\Delta(e^{ix_i}) = (2i)^{N(N-1)/2} e^{(N-1)\sum ix_i/2} \Pi_{i<j} \sin\frac{(x_i - x_j)}{2} \tag{1.270}$$

we find the kernel for the disk in the large N limit,

$$Z(x_i, A) = 2^{-N^2/2}\Pi_{i<j}\frac{x_i - x_j}{\sin(x_i - x_j)/2} e^{-\frac{N}{2A}\sum x_i^2} e^{-N^2\left(\frac{1}{2}logA - \frac{A}{24} - 1\right)} \tag{1.271}$$

To take into account the periodicity of the kernel we make the substitution $x_i \rightarrow x_i + 2\pi n_i$, with n_i arbitrary integer.

This is the standard formula we have meet a lot of times. The fact that the large N limit gives the correct result also for finite N is another sign that the N expansion stops after the first term (see section (1.6)).

Chapter 2

Four dimensional N=4 QCD

String theories compactified on a circle are known to possess duality invariances: their spectrum and interactions are invariant under a transformation that maps the radius of the circle, R, into α'/R (α' is the string tension). This symmetry holds at *each order* in the string loop expansion, and, in general, interchanges light states with ultra-massive ones. It is therefore essentially non perturbative in α' and it is sometimes called target-space duality.

Recently another type of duality symmetry has been conjectured. It is called strong-weak coupling duality, or S-duality [33]. It is a symmetry of four-dimensional strings, which acts by transforming the four-dimensional dilaton field S as

$$iS \rightarrow \frac{aiS+b}{ciS+d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,Z). \tag{2.1}$$

Since the real part of S multiplies the gauge kinetic term in the 4-D effective action of strings, this symmetry contains a duality $S \rightarrow 1/S$, which transforms, for instance, a weakly coupled theory ($\mathrm{Re}\, S \equiv 4\pi/g^2 \gg 1$) into a strongly coupled one. Thus S-duality, unlike the previously mentioned $R \rightarrow \alpha'/R$ duality, is essentially non-perturbative in the gauge coupling constant: it cannot hold order by order in the loop expansion.

The most important phenomenological applications of S-duality regard N=1 4-D string vacua; on the other hand, the best chance to *prove* S-duality lies in toroidal compactifications of the heterotic string. These compactifications possess an unbroken N=4 supersymmetry. Arguments in favor of S-duality have been given almost exclusively for N=4 compactifications. They fall into two groups. One of them conjectures a map between strings and five-branes [34]; under this map the role of S-duality and the target-space duality are reversed, in particular, S-duality holds order by order in the five-brane perturbation expansion. The other deals with the low-energy effective theory of the heterotic string, obtained in the limit $\alpha' \rightarrow 0$, $g=$ constant (the flat limit). If S-duality exists, it must holds order by order in the α' expansion, in particular it must be a symmetry of the flat limit. Thus the study of this limit provides us with *necessary* conditions for the existence of S-duality. The flat limit of a heterotic N=4 compactification is a N=4 supersymmetric gauge theory. Arguments exist about S-duality of this flat limit when the gauge group ($U(1)^6\times$ SO(44)) is fully broken to $U(1)^{28}$. In this latter, Abelian case S-duality is closely related to the charge-monopole duality [35].

In this chapter, we would like to bring other arguments in favor of S-duality, always in the flat limit $\alpha' \rightarrow 0$, by proposing a way to study this symmetry even when the unbroken gauge group is non-Abelian. We think that this study of non-Abelian S-duality is important *per se*, even without reference to strings, in that it may shed light on the wider problem of finding the dynamical realization of gauge symmetry in the case of gauge theories with extended supersymmetry and, in particular, of finite theories.

A N=4 super Yang-Mills theory is completely determined by the gauge group. The Lagrangian fields form a supermultiplet $(A^a_\mu, \lambda^a_I, \phi^a_{IJ})$ made of 1 spin-one, 4 spin-1/2, and 6 spin-zero fields, all in the adjoint of a gauge group G. The index a labels the adjoint representation, whereas the indices $I, J = 1,2,3,4$ are the so-called extension indices; finally the λ^a_I are Weyl spinors and the scalars obey a reality condition $2\phi^a_{IJ} = \epsilon_{IJKL}(\phi^a_{KL})^*$. The non-constant fluctuations of the dilaton S are $O(\alpha'^{1/2})$ since its kinetic term is $O(\alpha'^{-1})$. Thus, in the flat limit, S becomes indistinguishable from a bona-fide coupling constant. The Lagrangian reads

$$\begin{aligned} L &= \frac{1}{16\pi}\mathrm{Re}\, S[F^a_{\mu\nu}F^{a\,\mu\nu} + +\bar{\lambda}^{a\,I}\not{D}\lambda^a_I + f_{abc}\bar{\lambda}^{a\,I}\phi^b_{IJ}\lambda^{c\,J} + D_\mu\phi^{a\,IJ}D^\mu\phi^a_{IJ} \\ &\quad + f_{abc}f_{ade}\phi^b_{IJ}\phi^{c\,JK}\phi^d_{KL}\phi^{e\,LI}] + \frac{1}{16\pi}\mathrm{Im}\, S F^a_{\mu\nu}\tilde{F}^{a\,\mu\nu}. \end{aligned} \tag{2.2}$$

Here $\phi^{a\,IJ} \equiv (\phi^a_{IJ})^*$, and f_{abc} are the structure constants of G. The Cartan-Killing metric is δ_{ab}. By Eq. (2.2), one reads off the relation between S, the coupling constant g, and theta angle θ:

$$\mathrm{Re}\, S = \frac{4\pi}{g^2}, \quad \mathrm{Im}\, S = \frac{\theta}{2\pi}. \tag{2.3}$$

Thus, S-duality implies, among other things, that N=4 super Yang-Mills is invariant under $g \rightarrow 4\pi/g$. It may seem surprising that a non-Abelian gauge theory may possess a strong-weak coupling symmetry. This is less surprising if one considers the case in which the non-Abelian symmetry is fully broken to its Cartan subgroup $U(1)^r$, ($r=$ rank G). This is achieved by giving a generic VEV to the scalars present in the theory, along one of the flat directions of the scalar potential in Eq. (2.2) This broken gauge theory possesses stable, finite-energy monopoles of magnetic charge $4\pi/g$ [36, 37, 38]. Their mass saturates a bound due solely to the supersymmetry algebra, and therefore receive no corrections in perturbation theory [39]. Thus one may expect the theory to be invariant under the duality $S \rightarrow 1/S$, at least for $\mathrm{Im}\, S = 0$, and when accompanied by the interchange of electric charges with magnetic monopoles. A wealth of other arguments supporting full S-duality were given in refs. [5, 40]. In those references, the fact that the unbroken gauge group is Abelian allows for a unambiguous definition of the electric and magnetic fluxes (and charges). The problem of studying N=4 theories when the unbroken gauge group is non-Abelian remains open. Besides, another problem has to be addressed. As mentioned before, in arguing for SL(2,Z) S-duality one makes use of the *Abelian* definition of electric and magnetic charges. This definition is singular when the gauge group is unbroken [8, 36]. A non-singular, gauge invariant definition of the fluxes has been given by 't Hooft in [41]. The 't Hooft proposal applies to any gauge theory which contains only elementary fields in the adjoint representation of the gauge group and, in particular, to N=4. Let us recall the main features of this proposal. This will be the content of the first section. The chapter will continue with a brief review of the evidences for S-duality in supersymmetric Yang-Mills. In section 4 we come to the main content of this chapter, a proposal for the partition function on a box [42]. The canonical quantization of the theory at first order in the coupling constant, reveals that the relevant modes live on the toron valley, the Cartan torus divided by the Weyl group. We show that the ansatz is compatible with t'Hooft duality, a physical non trivial check. The last section deals with the extension to SU(N).

2.1 t'Hooft duality

For simplicity, in this section, we will consider the case of pure non supersymmetric gauge theory.

If we consider only fields in the adjoint representation, they are invariant under the center elements of the gauge group and we actually have an $SU(N)/Z_N$ gauge theory.

Putting the theory in a box of sides a_i, we can impose twisted boundary conditions on the fields [41]. Considering firstly the two directions 1,2, the most general boundary condition is

$$\begin{aligned} A_\mu(a_1, x_2) &= \Omega_1(x_2)A_\mu(0, x_2) \\ A_\mu(x_1, a_2) &= \Omega_2(x_1)A_\mu(x_1, 0) \end{aligned} \tag{2.4}$$

where Ω are gauge transformations, not necessarily periodic. To get no contradictions in the corner we must have

$$\Omega_1(a_2)\Omega_2(0) = \Omega_2(a_1)\Omega_1(0)Z \tag{2.5}$$

where Z is an element of the center of SU(N).

We can perform general periodic gauge transformations which alter the boundary conditions (2.4); for example, we can obtain the simpler condition

$$\begin{aligned} A_\mu(a_1, x_2) &= \Omega(x_2)A_\mu(0, x_2) \\ A_\mu(x_1, a_2) &= A_\mu(x_1, 0) \end{aligned} \tag{2.6}$$

with only one Ω left, which now satisfies

$$\Omega(a_2) = \Omega(0)Z \tag{2.7}$$

Further transformations are possible that change Ω, but there is no way to eliminate Z from the formulas. The boundary condition is so linked to the choice of an element in the center of SU(N).

We can repeat the reasoning for all the planes (μ, ν) in the euclidean space; we have to specified an element of the center $e^{2\pi i n/N}$ (an integer modulo N, $n_{\mu\nu}$) for each of these 6 independent planes.

We can label

$$\begin{aligned} k_i &= n_{0i} \\ m_i &= \epsilon_{ijk} n_{jk} \end{aligned} \tag{2.8}$$

and compute functional integrals with these twisted boundary conditions

$$W(k,m) = \int_{[k,m]} [DA] e^{S(A)} \tag{2.9}$$

choosing with care an euclidean invariant gauge fixing procedure.

To elucidate the physical meaning of the topological number $n_{\mu\nu}$, we consider the canonical quantization of the theory in the $A_0 = 0$ gauge. For each curve C at fixed time t we construct the Wilson loop

$$A(C,t) = TrPe^{\int igA_i(x_i,t)dx_i} \tag{2.10}$$

called also order parameter. In analogy with ordinary electromagnetism, we say that $A(C,t)$ measures magnetic flux through C, or electric flux around C; these are gauge invariant quantities.

There exists a dual operator, the disorder parameter, $B(C,t)$ which measures electric flux through C or magnetic flux around C, defined by postulating the equal time commutation relations

$$\begin{aligned} &[A(C,t), A(C',t)] = 0 \\ &[B(C,t), B(C',t)] = 0 \\ &A(C)B(C') = B(C')A(C)e^{2\pi i n/N} \end{aligned} \tag{2.11}$$

where n is the number of time C' winds around C in a certain direction.

An explicit definition of B(C) can be given as follows. Define a pseudo-gauge transformation, well defined everywhere except on C; the singularity is such that for every closed curve that surrounds C n times

$$\Omega^C(2\pi) = e^{2\pi i n/N} \Omega^C(0) \tag{2.12}$$

The singularity can be, for example, a string; its position is not physical since we can change it with a gauge transformation. This discontinuity is not felt locally by the gauge fields which are invariant under the center of the group. The right definition of B(C) supposes also a smoothening of the singularity with a sort of regularization. This means that B(C) is not a gauge transformation; however, it leaves invariant the Hilbert space.

The attempt to define the magnetic flux in the, say, 3 direction with the operator A(C) around the 12 plane, faces the immediate difficulty that such quantity is not conserved. So we consider instead a curve C in the 3 direction which winds the box exactly once. B(C) on such a curve creates one magnetic flux line in the 3 direction. But now the action of B(C) changes the boundary conditions of the gauge field A_μ in the direction 3, since Ω^C has a jump of $e^{2\pi i/N}$.

$$m_3 \rightarrow m_3 + 1 \tag{2.13}$$

So, m_i clearly counts the number of times the operator B(C) has acted, i.e. the number of magnetic fluxes created in the direction i.

To elucidate the meaning of the electric flux, we continue to consider the theory quantized in the $A_0 = 0$ gauge. We have again the invariance under time independent gauge transformations. For such transformations which are topologically trivial (continuously connected to the identity) we choose, as usual, the identity representation on physical states,

$$\Omega|\psi >= |\psi > \tag{2.14}$$

the only one consistent in the large a_i (flat) limit.

For non trivial transformations, however, we have two type of topological effects. One is related to the 2nd Chern class of the map (instanton effects) and have monodimensional representations labelled by the θ angle. A second effect arises in the box and is related to the boundary conditions of these gauge transformations. If we allowed these transformations to be periodic only up to an element of the center Z_i, we would still be considering an invariance of the hamiltonian and, moreover, we have not changed the boundary conditions on the fields A_μ, defining the path integral W(k,m). This means that they commute with the magnetic flux, which, as we have seen, is simply the boundary condition labelled by m.

Such gauge transformations form homotopy classes labelled by three integers (k_1, k_2, k_3). The partition function for canonical quantization Tre^{-tH} supposes periodic boundary conditions in the time direction; in the spatial direction, however, we are allowed to take twisted conditions corresponding, for example, to fixed magnetic flux m. When inserted in the partition function

$$Tr\Omega(k)e^{-tH} \tag{2.15}$$

the operator Ω, leaving invariant H, has the only effect to twist the temporal boundary conditions by the integer k, according to (2.4). So, we identify,

$$W(k,m) = Tr_m\Omega(k)e^{-tH} \tag{2.16}$$

Since Ω is a invariance of the hamiltonian, it must have an one-dimensional representation on physical states,

$$\omega(k)|\psi> = e^{i\omega(k)}|\psi> \tag{2.17}$$

and since the operator $\Omega(k)$ realizes the center Z_N, $e^{i\omega}$ must be a representation of Z_N

$$\omega(k) = \frac{2\pi}{N}\sum_i e_i k_i \tag{2.18}$$

with e_i integers defined modulo N. Let's see what is their physical interpretation.

The quantities A(C) are gauge invariant by definition, but are not requested to be invariant also under such a not periodic transformation. Indeed, if we consider C as the closed curve that winds once around the direction i

$$A(C) \rightarrow Tr\Omega(0)Pe^{ig\int A_i dx_i}\Omega^{-1}(2\pi) = e^{-2\pi i k_i/N}A(C) \tag{2.19}$$

If we apply A(C) on the hilbert space, taking C in the 3 direction for simplicity

$$|\psi'> = A(C)|\psi> \tag{2.20}$$

since

$$A(C)\Omega(k)|\psi> = \Omega(k)e^{-2\pi i k_3/N}A(C)|\psi> \tag{2.21}$$

we find

$$\Omega(k)|\psi'> = e^{i\omega(k)+2\pi i k_3/N}|\psi'> \tag{2.22}$$

so $e_3 \rightarrow e_3 + 1$. Clearly, the integers e_i count the number of times the operators A(C) have acted, i.e. the number of electric fluxes in the i direction.

So we have given a non abelian definition of electric and magnetic flux via the twisted boundary conditions of the fields; in this way, such fluxes are defined only modulo N.

Both the non abelian electric and magnetic fluxes defined by the integers e_i, m_i reduce to the ordinary electromagnetic quantities in the abelian case.

How can we compute the free energy with fixed electric and magnetic flux? For what regards the magnetic flux we already know that it is sufficient to fix the spatial boundary conditions. For the electric flux, we have to take a suitable combination of path integrals W(k,m) to obtain the sector of the hilbert space corresponding to a given eigenvalue of the electric flux. In hamiltonian formalism, we can interpret

$$P(e) = \frac{1}{N^3}\sum_k e^{-2\pi i k e/N}\Omega(k) \tag{2.23}$$

as the projector on states with fixed electric flux e.

Therefore we have for the free energy with fixed fluxes,

$$e^{-\beta F(e,m)} = \frac{1}{N^3}\sum_k e^{-2\pi i k e/N}Tr_m\Omega(k)e^{-\beta H} \tag{2.24}$$

In this way, we find that the partition function with fixed electric flux is the discrete Fourier transform of the path integrals with fixed boundary conditions.

$$e^{-\beta F(e,m)} = \frac{1}{N^3} \sum_k e^{-2\pi i k e/N} W(k,m) \tag{2.25}$$

Summarizing, while the magnetic flux is directly connected with the boundary conditions of the path integral in the spatial directions, the electric flux is not diagonal on the space with given boundary conditions, but it is rather a Fourier transform of it.

The t'Hooft duality is simply the statement of euclidean invariance. W must be invariant under joint rotations of the sides a_i and the integers $n_{\mu\nu}$. In particular, if we perform the euclidean rotation in R^4, $(i\sigma_1, -i\sigma_1)$, indicating with $\tilde{k}$ the first two component of the vector k, and with $\tilde{k}'$ the same with the two components interchanged, we must have

$$W(\tilde{k}, k_3, \tilde{m}, m_3; \tilde{a}, a_3, \beta) = W(\tilde{m}, k_3, \tilde{k}, m_3; \tilde{a}', \beta, a_3) \tag{2.26}$$

which implies

$$e^{-\beta F(\tilde{e}, e_3, \tilde{m}, m_3; \tilde{a}, a_3, \beta)} = \frac{1}{N^2} \sum_{\tilde{k}, \tilde{l}} e^{2\pi i/N[-(\tilde{k}\tilde{e})+(\tilde{l}\tilde{m})] - a_3 F(\tilde{l}, e_3, \tilde{k}, m_3; \tilde{a}', \beta, a_3)} \tag{2.27}$$

Let's note that, being a simple consequence of euclidean invariance, this is an exact statement in gauge theories.

The t'Hooft duality can be used to derive statements about the phases of a gauge theory. We briefly review some of these conclusions [41]. Let us assume that the theory has a mass gap and so no massless particle is present. Then the asymptotic region $a_i \to \infty$ will be approached exponentially. It is so excluded that

$$F(e,m) \to 0, \tag{2.28}$$

exponentially when $a_i \to \infty$ for all e,m , since it is clearly inconsistent with the duality equation.

It happens that only some of the F(e,m) are light, and necessarily (otherwise the duality equation would be inconsistent) some others go to infinity. For example, for SU(2) or SU(3) either all electric or all magnetic fluxes are light; just for SU(4) the situation is more complicated. What it is however true is that there exist necessarily heavy fluxes. This is different from abelian case, where the energy of a given flux behaves as

$$E = \frac{C a_i}{a_j a_k} \tag{2.29}$$

spreaded out in space, and contributing a constant free energy in the limit in which the box becomes large. Non abelian fluxes, with F large and large in the thermodynamic limit, probably create strings with energy proportional to their length

$$F \to \rho a \tag{2.30}$$

Duality will never enable us to determine whether it is the electric or magnetic flux that is heavy. In QCD it is usually assumed that the magnetic fluxes are light while the electric ones are heavy and behave like strings (confinement mode). In the Higgs mode, however, we have the complementary situation of confined magnetic fluxes [43].

To extract further informations one usually makes the assumption of factorizability in the thermodynamic limit

$$\begin{aligned} F(e,m) &= F(e) + F(m), \\ a_i, \beta &\to \infty \end{aligned} \tag{2.31}$$

Probably the main result which follows from the consistence duality equation [41] is that if we suppose confinement in the electric domain

$$F(e) \to \rho a_i \tag{2.32}$$

where ρ is the fundamental string tension, we are able to have informations about the magnetic flux, via duality equation and there will be no magnetic confinement. In fact, the magnetic free energy behaves as

$$e^{-\rho a_i a_j} \tag{2.33}$$

and causes a rapid decrease of the magnetic flux.

We could equally have started from the assumption of magnetic flux confinement, and conclude the impossibility of the contemporaneous electric confinement.

An example of what happens in the case of presence of massless particles, we can consider the Georgi-Glashow model [44], where SU(2) is spontaneously broken to an abelian theory U(1) via the vacuum expectation value of the Higgs field. There are electrically charged particles $W^{+,-}$ with charge e and monopol solution of magnetic charge

$$g = \frac{4\pi}{e} \tag{2.34}$$

Montonen and Olive [35] observed a duality between magnetic monopols and dual charged vector particles, but we will return on this point in the next sections.

In the thermodynamic limit, since massive particles are exponentially suppressed, only the U(1) Maxwell fields contribute to the free energy. However massive particles can indirectly contribute with quantum configuration in which two opposite charge vector bosons are created, go in opposite direction, wind around the torus and annihilate together, leaving a configuration in which the electric U(1) flux in this direction has increased by two units. This is another way of saying that the electric flux is defined only modulo 2. We have to sum over all these configurations with their Boltzmann factor.

In this abelian model factorization clearly takes place. The energy of a k electric flux in the, say, 1 direction is

$$F_e(k) = \frac{g^2 a_1 k^2}{8 a_2 a_3} \tag{2.35}$$

where k can take the values 0,1. Other values of k are reached via pair quantum creation with two unity of flux. For example, the e=1 free energy is

$$e^{-\beta F(e=1)} = \sum_{k \epsilon Z} e^{-\beta \frac{g^2 a_1}{2 a_2 a_3}(k+1/2)^2} \tag{2.36}$$

for F(e=0) free energy we substitute $k + 1/2 \longrightarrow k$. The magnetic free energy can be obtained replacing $g \longrightarrow 4\pi/g$, the charge of the monopole.

$$e^{-\beta F(m=1)} = \sum_{k \epsilon Z} e^{-\beta \frac{8\pi^2 a_1}{g^2 a_2 a_3}(k+1/2)^2} \tag{2.37}$$

With these theta functions as free energy the verification of duality equation is a straightforward exercise in Poisson resummation.

We conclude that the Georgi-Glashow model is in a self-dual phase, or Coulomb phase. We will see that N=4 supersymmetric Yang-Mills resemble this abelian model more that the full non abelian QCD theory.

2.2 A perturbative calculation

In the case of non abelian Yang-Mills theory, it is also possible to undergo a perturbative calculation of an effective hamiltonian [45].

We put the system in the usual box, with periodic boundary conditions, and use the $A_0 = 0$ gauge. The hamiltonian H acts on functionals of the fields A, invariant under time independent gauge transformations. As in the previous section, taking the transform of the field A with gauge transformations periodic up to elements of the center, we find the twisted sectors; by diagonalizing this set of quasi-periodic transformations we find the electric flux sectors.

In the electric and magnetic fields,

$$\begin{aligned} E_i^a(x) &= \frac{1}{i}\frac{\delta}{\delta A_k^a(x)} \\ B_i^a(x) &= \frac{1}{2}\epsilon_{ijk}(\partial_l A_i^a(x) - \partial_i A_l^a(x) + f^{abc} A_l^b(x) A_i^c(x)) \end{aligned} \tag{2.38}$$

the hamiltonian reads

$$H = \int_0^a d^3x \left(\frac{1}{2} g^2 E_k^a(x) E_k^a(x) + \frac{1}{2g^2} B_k^a(x) B_k^a(x) \right) \tag{2.39}$$

In order to keep the magnetic energy bounded in the small coupling limit, the wave functionals $\psi[A]$ of the low-lying states have to be supported essentially around the potentials with $B_k = 0$.

Locally, a solution of $B_k = 0$ is a pure gauge; in general, the Cartan (diagonal) constant gauge potentials

$$A_k^{ij}(\phi) = \delta^{ij}\frac{1}{ia_k}\phi_k^i \tag{2.40}$$

defined by N angles with the constraint

$$\sum_i \phi_k^i = 0 \tag{2.41}$$

called torons, is the general solution of $B_k = 0$, up to a gauge transformation. The remain of the gauge group, as we have stressed several times in the previous chapter, is the Weyl group acting on the angles ϕ_k as a permutation. The toron manifold, describing the flat directions of the magnetic energy, is therefore T/W, a compact manifold with boundary (as we have seen in section (1.3).

In the small coupling limit, at least heuristically, the low energy wave functions are localized around the toron manifold, where the magnetic energy vanishes. The kinetic energy for the motion on this manifold (the electric contribution in the hamiltonian) is of order g^2, but the fluctuations around the manifold can reach terms of order O(1). What actually happens is that the energy of fluctuations is not equal in every point of the toron valley, and only around a finite number of points the potential energy is minimized; all other point on the manifold correspond to unstable vacua.

Parametrizing the gauge fields A near the toron manifold,

$$A_k(x) = \Lambda(x)[A_k(x) + gq_k(x)]\Lambda^{-1} + \Lambda(x)\partial_k\Lambda(x)^{-1} \tag{2.42}$$

where q_k is a fluctuation orthogonal to the toron manifold, and the gauge orbit

$$\begin{aligned} &\int dx^3 q_k^{aa}(x) = 0 \\ &D_k q_k = \partial_k q + AdA_k[\phi]q_k \end{aligned} \tag{2.43}$$

The fundamental fact is that the fluctuations contribute an O(1) term to the hamiltonian

$$H = \int dx^3[\frac{1}{2}p_k^a(x)p_k^a(x) + q_k^a(x)(\Omega[\phi]q)_k^a(x)] \tag{2.44}$$

where

$$(\Omega[\phi]q)_k = -D_l D_l q_k + D_l D_k q_l \tag{2.45}$$

The kinetic part for toron fields ϕ does not contribute to the order O(1), their contribution being of order g^2. H can be therefore diagonalized at fixed values of angles ϕ_k^i. In particular, the lowest energy value is

$$\frac{1}{2}Tr(\Omega[\phi])^{1/2} \tag{2.46}$$

Note that this zero point energy depends on the point of the toron manifold, as previously said. The true ground state can be obtained minimizing $Tr\Omega^{1/2}$. It can be done [45] with the result

$$E[\phi] = E[0] + \frac{1}{\pi^2 a} \sum_{a\neq b,\nu\neq 0} \frac{1}{(\nu^2)^2}(1 - \cos\nu(\phi^a - \phi^b)) \tag{2.47}$$

The ground state is obtained when

$$\phi^a = \phi^b(mod 2\pi) \tag{2.48}$$

Since $\sum \phi_k = 0$ this means

$$\phi_k^a = \frac{2\pi}{N} n_k (mod 2\pi) \tag{2.49}$$

where n_k are integers modulo N. It is not difficult to recognize in these states the true vacuum and their central conjugate. Note that the value of the energy is always the same constant E[0].

We find two fundamental results.

Firstly, all the toron vacua are quantum instable except for the true vacuum and its central conjugate.

Secondly, since the pseudo gauge transformations which connect different central sectors commute with the hamiltonian and the energy of the central ground state is the same, there is an exact degeneracy of the spectrum for all the electric flux sectors; this is completely different from abelian gauge theories where the sectors are separated by an energy gap of order g^2/L.

Note that this result is true only for ordinary gauge theories; in the case of supersymmetric Yang-Mills theories, the harmonic oscillators of bosons and fermions exactly cancel, giving an exact degeneration on all the toron manifold. The low energy states can be found by quantizing the modes living on such manifold.

We come to the qualitative features of the perturbative expansion. We isolate the constant part of the gauge fields

$$A_k^a(x) = g^{2/3} L^{-1} c_k^a + g q_k^a(x) \tag{2.50}$$

where c_k^a are coordinates on the toron manifold and q are transverse modes $\int q = 0$. The canonical momenta are similarly decomposed

$$\pi_k^a(x) = g^{2/3} L^{-2} e_k^a + g^{-1} p_k^a(x) \tag{2.51}$$

with

$$e_k^a = -i\partial_{c_k^a}, \int dx p_k^a(x) = 0 \tag{2.52}$$

One finds [45] the following expansion

$$\begin{aligned} H &= H_0 + \sum_{i=1}^{\infty} g^{2i/3} H_i \\ H_0 &= \frac{1}{2}\int dx[p_i^a p_i^a + \partial_l q_i^a \partial_l q_i^a] \end{aligned} \tag{2.53}$$

Note that the O(1) order hamiltonian is harmonic, of obvious spectrum, which is purely discrete and with an energy gap of order $2\pi/L$ between the ground state and the first excited state. At this level, there is a complete degeneration of the constant modes, lifted by the perturbation theory in g.

The low energy states are the perturbation around the ground state of the fluctuation q. The main results on this argument [45], are that the perturbative expansion is actually an expansion in $\lambda = g^{2/3}$, and that we can take it into account by the effective hamiltonian

$$\begin{aligned} H &= \frac{\lambda}{L}\sum_{n=0}^{\infty} \lambda^n H_n \\ H_0 &= \frac{1}{2} e_i^a e_i^a + \frac{1}{4}(f^{abc} c_k^b c_l^c)(f^{ade} c_k^d c_l^f) \\ H_1 &= -cost c_k^a c_k^a \end{aligned} \tag{2.54}$$

The lowest order hamiltonian is the one we would obtain if we restrict the Yang-Mills action to fields depending only on time. Such fields are the slow modes of the theory. The fast modes are systematically integrated out to give the effective hamiltonian indicated.

2.3 Evidence for S-duality

It is an old conjecture [35] that there is duality between charged particles and monopoles. This duality exchanges strong and weak coupling regime, exchanging at the same time electric and magnetic quantities. Montonen and Olive proposed a symmetry with the above properties that exchanges the gauge group with the dual one (the group whose weight lattice is the dual of that of the originary gauge group) [46]. According to this conjecture, the strong coupling limit of the theory is equivalent to the weak coupling limit with ordinary particles and monopoles exchanged.

It was soon evident that for pure gauge theories elementary particles and monopoles do not have the same quantum and Lorentz number, violating the duality statement. For supersymmetric Yang-Mills theories things go better, but we have to increase considerably the number of extended supersymmetries; N=2 is again not enough. It happens that the natural arena for Montonen and Olive duality is N=4 supersymmetric Yang-Mills, where particles and monopoles have the same spin content.

Moreover N=4 Yang-Mills, even if it is actually less interesting has a certain number of features which simplify things considerably. First of all, in N=2 the U(1)-R symmetry is anomalous, preventing the existence of a microscopic theta angle. For N=4 the anomaly cancel so the theory possesses a natural theta angle. It has also a natural dimensionless gauge coupling g, since the N=4 beta function vanishes at all orders in perturbation theory [47, 48, 49]; this will allows us to formulate a reliable perturbation theory in a box, without fear that the coupling runs when we change size to the box. Moreover, the N=4 symmetry forces the effective hamiltonian to have the same form of the original one; conjectures based on the vanishing of the beta function [38] and explicit computations [50] seem to confirm this fact.

Recently, more evidence have been provided for the existence of S-duality in N=4 Yang-Mills theories [5] [6] and, to some extent, also in N=2 theories.

Besides the existence of monopoles, also solitons with both 1 electric and 1 magnetic charge are easily found (dyons). A crucial simplification comes from the fact that the mass of these particles, satisfying the Bogomonly bound, are semiclassically exact. [39]

The N=4 theory, as already noted, has a further parameter, the theta angle. It is convenient to to combine the two parameters, g and θ, in a single complex variable

$$\lambda = \frac{\theta}{2\pi} + \frac{4\pi}{g^2} \tag{2.55}$$

Then, the originary proposed Montonen-Olive symmetry $g \longrightarrow 4\pi/g$ can be extended to a full symmetry $SL(2,Z)$

$$\lambda \longrightarrow \frac{p\lambda + q}{r\lambda + s} \tag{2.56}$$

The S transformation at $\theta = 0$ is the celebrated $g \longrightarrow 4\pi/g$ proposed by Montonen and Olive, while the T transformation simply means that the theory is periodic in θ with period 2π.

The mass of a state in the sector with electric-magnetic quantum numbers k=(e,m) is given by the Bogomonly bound saturating equation

$$cost k^T M k \tag{2.57}$$

where the matrix M is

$$M(\lambda) = \frac{1}{\lambda_2} \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & |\lambda|^2 \end{pmatrix} \tag{2.58}$$

If, besides λ, also the vector k=(e,m) transforms under SL(2,Z)

$$\begin{pmatrix} e \\ m \end{pmatrix} \rightarrow \begin{pmatrix} pe - qm \\ -re + sm \end{pmatrix} \tag{2.59}$$

this energy is SL(2,Z) invariant.

Now, given the existence of a state of quantum number (1,0) (the charged particles, the elementary fields in the lagrangian), SL(2,Z) predicts also the existence of states of numbers (p,q). States with (0,1) numbers are realized as classical solution of the theory, the usual BPS monopoles. Also states with m=1 and e generic are easily found (dyons). A non trivial check of S-duality has been done by Sen which showed the existence of bound states with two units of magnetic charge, with the topological interpretation of self-dual harmonic forms on the moduli space of BPS multimonopole configurations. The topological interpretation of the states of arbitrary magnetic charge has been given by Segal [51].

The existence of these bound states is certainly a non trivial check that S-duality is in fact realizes in N=4 Yang-Mills theories. Further evidence comes from a recent work of Vafa and Witten [6]. They compute the super-partition function of the theory on various manifolds, finding a fully duality invariant result. Actually, they define a twisted version of N=4 Yang-Mills on an arbitrary manifold (with all periodic boundary conditions even for fermions); this twisted version coincides with the original theory on Hyper-kaeler manifolds, like K3 or the torus (where such partition function is the Witten index). Standard arguments imply that the theory localizes around the instanton solutions, giving a holomorphic power serie in λ. In fact, the lagrangian of the form

$$\frac{4\pi}{g^2}\int F^2 + \frac{\theta}{2\pi}\int F\tilde{F} \tag{2.60}$$

evaluated on an instanton $F = \tilde{F}$, contributes the topological number n multiplyied by λ. So we have

$$Z = \sum a_n \lambda^n \tag{2.61}$$

where the coefficient a_n can be computed as the Euler characteristic of some instanton moduli space.

What is surprising is that when we insert the right coefficient a_n for K3 and some other manifold, the resulting sum becomes a modular function, giving a full of evidence for S-duality.

To conclude, we note that in N=2 we have excluded the S-duality for the wrong quantum number of the monopoles. What really happens is that, the theta angle which is anomalous can be consistently defined in the low energy effective theory and that such theory has an SL(2,Z) symmetry [7].

2.4 The partition function

Now we come to the discussion of the N=4 Yang-Mills partition function for SU(2), proposing a full consistent S-duality result. The main points we will use are the finiteness of the theory, which allows to perform a reliable perturbation expansion for small g, and the somewhat supported claim that the effective action for monopoles is identical with the original theory.

Let's summarize what we have learned about the electric and magnetic structure of the gauge theory; the presence of scalars and fermions will be a minor complication.

First of all, we note that all the fields in the lagrangian are in the adjont representation, so we are actually working with an $SU(N)/Z_N$ gauge theory to which the t'Hooft proposal for an unambiguous electric and magnetic flux definition applies.

One begins by evaluating the Euclidean functional integral in a box of sides (a_1, a_2, a_3, a_4).

$$W[\vec{k},\vec{m}] = \int [dA^a_\mu d\lambda^a_I d\phi^a_{IJ}] \exp(-\int d^4x L), \quad k_i \equiv n_{4i}, \quad n_{ij} \equiv \epsilon_{ijk} m_k, \tag{2.62}$$

where $i, j, .. = 1, 2, 3$ label the spatial coordinates. The Lagrangian L is given in Eq. (2.2). The boundary conditions for all fields are periodic up to a gauge transformation

$$\Phi(x + a_\nu e_\nu) = \Phi^{\Omega_\nu(x)}(x), \tag{2.63}$$

where e_ν denotes a canonical versor of R^4 and repeated indices are not summed. Φ and Φ^Ω denote generically a field of the supermultiplet and its gauge transform under Ω, respectively. The only exception to these boundaries conditions is for fermions; they are antiperiodic in the temporal direction.

If we allow for gauge rotations periodic only up to elements of the center, we have to impose the corner condition (2.1) and so the integers $n_{\mu\nu}$ in Eq. (2.62) are defined modulo two by the consistency condition [41]

$$\Omega_\mu(x + a_\nu e_\nu) = (-1)^{n_{\mu\nu}} \Omega_\mu(x), \quad n_{\mu\nu} = -n_{\nu\mu}. \tag{2.64}$$

where (-1) is the generator of the center of SU(2).

As we have seen, the twist in the space directions n_{ij} corresponds to a magnetic flux $2m_i = \epsilon_{ijk} n_{jk}$. However, the twist n_{i4} is linked to the electric flux only by the Fourier transform

$$\exp\{-\beta F[\vec{e},\vec{m}]\} = \frac{1}{8} \sum_{\vec{k}\in(Z_2)^3} \exp(\pi i \vec{e}\cdot\vec{k}) W[\vec{k},\vec{m}]. \tag{2.65}$$

Here $a_4 \equiv \beta$ and $F[\vec{e},\vec{m}]$ is the free energy of a configuration with electric flux $\vec{e} = (e_1, e_2, e_3)$ and magnetic flux $\vec{m}$. Notice that the fluxes are defined modulo two, that is, modulo the order of the center (Z_2) of SU(2).

The behavior of these free energies in the thermodynamical limit $a_i \to \infty$ determines the phases of a gauge theory.

The $F[\vec{e},\vec{m}]$ can be computed also in the Hamiltonian formalism, for instance, in the gauge $A_0 = 0$. The formulas of section (2.1) now read,

$$e^{-\beta F[\vec{e},\vec{m}]} = \mathrm{Tr}_{\vec{m}} \mathcal{P}[\vec{e}] e^{-\beta H}, \quad \mathcal{P}[\vec{e}] = \frac{1}{8} \sum_{\vec{k}\in(Z_2)^3} e^{-\pi i \vec{e}\cdot\vec{k}} \prod_{i=1}^{3} T_i^{k_i}. \tag{2.66}$$

The symbol $\mathrm{Tr}_{\vec{m}}$ denotes the trace over gauge fields obeying boundary conditions twisted by $\vec{m}$. $\mathcal{P}[\vec{e}]$ is the projector onto states with definite value of the electric flux; this projector can be generated by the gauge transformations T_i, which are periodic up to the nontrivial element of the center of SU(2). The T_i act non-trivially on physical states, but trivially on *local* physical states. They obey the following boundary conditions:

$$T_i(\vec{x} + a_j e_j) = (-1)^{\delta_{ij}} T_i(\vec{x}) \tag{2.67}$$

A suitable choice for T_i is

$$T_i = e^{2\pi i \frac{x}{a_i} \tau^3} \tag{2.68}$$

Now we come to the evaluation of W(e,m). In general, it is impossible to evaluate the free energies in a closed form. This computation becomes possible in two limiting cases:

a) $\vec{m} = 0$, $g \ll 1$, and $\theta = 0$;

b) $\vec{e} = 0$, $g \gg 1$, and $\theta = 0$.

Let us compute at first case a). Since the coupling constant is small, we can compute $F[\vec{e}, 0]$ by using perturbation theory. This possibility exists since N=4 super Yang-Mills has vanishing beta function [47, 48, 49]. Thus, the renormalized coupling constant does not depend on the size of the box. Therefore a theory with $g \ll 1$ is weakly interacting at *any* length scale[1]. We had to set $\vec{m}$ to zero since in this regime the magnetic flux is expected to interact strongly, with coupling constant $O(1/g) \gg 1$. At $\vec{m} = 0$ all fields in the box obey periodic boundary conditions, up to a periodic, globally defined gauge transformation; therefore, in a given gauge they may be expanded in Fourier series

$$\Phi(\vec{x}) = \frac{1}{V} \sum_p e^{i\vec{p}\cdot\vec{x}} \Phi_{\vec{p}}, \quad p^i = \frac{2\pi}{a^i} n^i, \; n^i \in Z, \quad V \equiv a_1 a_2 a_3. \tag{2.69}$$

In the box, as we have discussed at length in section (2.2), all non-zero Fourier modes have energies $O(V^{-1/3})$. Moreover, the potential energy scales as $1/g^2$. For instance, the gauge fields possess a "magnetic" energy $\int d^3x g^{-2} \vec{B}^2$. The zero modes belonging to the Cartan subalgebra of SU(2), instead, have energies $O(g^2 V^{-1/3})$. At $g \ll 1$ only this latter set gives a significant contribution to the free energy [45, 52].

The zero modes are the general gauge transformations of the modes belonging to the Cartan subalgebra of SU(2) ($\Phi \equiv \sum_a \Phi_a \tau_a$, $\tau_a = \sigma_a/2$):

$$\begin{aligned} \vec{A}(\vec{x}) &= \Omega(\vec{x})^\dagger [\nabla + \tau_3 \vec{c}] \Omega(\vec{x}), \quad \Omega(\vec{x}) = e^{i\tau_a \omega_a(\vec{x})}, \\ \lambda^I_\alpha(\vec{x}) &= \Omega(\vec{x})^\dagger \tau_3 a^I_\alpha \Omega(\vec{x}), \quad [a^{I\dagger}_\alpha, a^J_\beta]_+ = \delta^{IJ}\delta_{\alpha\beta}, \quad [a^I_\alpha, a^J_\beta]_+ = 0, \\ \phi^{IJ}(\vec{x}) &= \Omega(\vec{x})^\dagger \tau_3 \phi^{IJ} \Omega(\vec{x}). \end{aligned} \tag{2.70}$$

[1]This argument shows that a perturbative calculation of $F[\vec{e}, 0]$ is unreliable in any asymptotically free theory, as for example pure Yang-Mills.

Here $\alpha, \beta = 1, 2$ are Weyl indices. Notice that the constant gauge configurations $\vec{c}$ (the torons of section (2.2) are defined only up to periodic gauge transformations. Thus, in a box of sides $\vec{a}$, they parametrize a compact space with boundary obtained by the identifications

$$c_i \approx c_i + \frac{4\pi}{a_i}, \quad c_i \approx -c_i. \tag{2.71}$$

The first identification is due to the gauge transformation

$$\Omega(\vec{x}) = \exp(4\pi i \tau_3 x_i / a_i), \tag{2.72}$$

the second is brought about by the gauge transformation $\Omega = 2\tau_2$, generating G-parity (which is the Weyl group of SU(2)). It is not difficult to recognize our old friend, the maximal torus divided by the Weyl group, T/W.

The Lagrangian of N=4 super Yang-Mills reduced to the zero modes reads, in our normalizations

$$L = \frac{V}{2g^2} \int dt \left[\left(\frac{d\vec{c}}{dt} \right)^2 + a^{\dagger}_{\alpha\, I} \frac{d a^I_{\alpha}}{dt} + \frac{1}{4} \frac{d\phi_{IJ}}{dt} \frac{d\phi^{IJ}}{dt} \right]. \tag{2.73}$$

By denoting with $\vec{\pi}$ and Π_{IJ} the canonically conjugate momenta of $\vec{c}$ and ϕ^{IJ}, respectively, one finds the Hamiltonian

$$H = \frac{g^2}{2V} \left[\vec{\pi}^2 + \frac{1}{4} \Pi_{IJ} \Pi^{IJ} \right]. \tag{2.74}$$

Notice that the fermions do not give any contribution to the energy. The physical wave functions Ψ are gauge invariant. When reduced to the toron manifold they become periodic functions of $\vec{c}$, even under G-parity.

$$\Psi(c_i + 4\pi/a_i) = \Psi(c_i), \quad \Psi(-\vec{c}) = \Psi(\vec{c}). \tag{2.75}$$

Thus, the spectrum of the momenta π_i is quantized, and the eigenvalues are $a_i k_i/2$, $k_i \in Z$. To find the eigenstates of the electric flux we recall that the projector over these eigenstates, $\mathcal{P}[\vec{e}]$, was defined in Eq. (2.66). It reduces, on the eigenstates of π_i, to

$$\mathcal{P}[\vec{e}] \exp(i k_i c_i a_i/2) = \begin{cases} \exp(i k_i c_i a_i/2) & \text{if } k_i = e_i \text{ modulo } 2, \\ 0 & \text{otherwise} \end{cases} \tag{2.76}$$

From this formula we learn that the sector of zero flux correspond to even momenta k, while the sector of flux 1 correspond to odd k.

We ought to deal with two problems before being able to write down the partition function.

The first one is that true physical states are not eigenstates of $\vec{\pi}$, but rather G-invariant linear combinations of them. For what regard the gauge bosons, this implies that G-invariant states are

$$\psi = \frac{1}{2} [e^{\frac{1}{2} k^i c_i a_i} + e^{-\frac{1}{2} k^i c_i a_i}] \tag{2.77}$$

while G-odd combinations are

$$\psi = \frac{1}{2}[e^{\frac{1}{2}k^i c_i a_i} - e^{-\frac{1}{2}k^i c_i a_i}] \tag{2.78}$$

Note that when $k_i = 0$ we have only the G-even function, the identity. To properly count the multiplicity of the physical states we must recall that the fermionic wave functions are obtained by applying the fermion creation operators $a_\alpha^{I\dagger}$ to the Fock vacuum $|0\rangle$. Thus G-parity even fermionic functions contain an even number of fermions, and G-odd ones an odd number. When $k_i \neq 0$ one finds physical G-invariant states by combining the G-odd bosonic wave functions with all G-odd fermionic wave functions, and the G-even bosonic wave functions with the G-even fermionic ones. Fermi statistics allows to find the multiplicity of the fermionic wave functions of given G-parity. It turns out to be the same for both parities, namely, 128. When $k_i = 0$ only the even bosonic wave functions exist, thus, the ratio of physical states with $k_i \neq 0$ to states with $k_i = 0$ is 1:2. This result is valid for any component of the momentum k_i separately. Note that the zero modes of the scalars, labelled by a continuous momentum p, contribute an even and odd function for every value of momentum, except for the zero measure set p=0, washed out by the phase space integration over p; thus they do not affect the previous conclusion.

The second problem involves the zero modes of the scalars ϕ^{IJ}. Different values of ϕ^{IJ} are not related by 4-D gauge transformations, thus the spectrum of their momenta is continuous, and their contribution to the statistical sum Eq. (2.66) is divergent. One may regularize this divergence either by hand, constraining the range of integration of ϕ^{IJ} to a finite volume, or by recalling that N=4 super Yang-Mills comes from the dimentional reduction of N=1 super Yang Mills in 10 dimensions [53]. The scalars are thus the four-dimensional relics of ten-dimensional gauge fields. If the compactification radius R of the extra dimensions is kept finite one finds that the range of integration of the scalars becomes finite, by the same mechanism which makes the range of $\vec{c}$ finite. The scalar contribution to the free energy thus is well defined and finite. In the $R \to 0$ limit it becomes

$$\mathrm{Tr}\, e^{-\beta H_{\text{scalars}}} = \int d\phi_{IJ} d\phi^{IJ} d\Pi_{IJ} d\Pi^{IJ} e^{-\beta g^2 \Pi_{IJ}\Pi^{IJ}/8V} = \text{const}\, R^{-6} \left(\frac{\beta g^2}{8V}\right)^{-3}. \tag{2.79}$$

It is crucial to note that since T_i may be written as $\exp(2\pi i x \tau_3/a_i)$, it acts trivially on ϕ^{IJ} (and a_α^I). Thus the contribution of the scalars to the free energy is an additive constant *independent of the electric flux*; it may be eliminated by a shift in the definition of the free energy, common to all flux sectors.

Now, it is immediate to evaluate the statistical sum in Eq. (2.66) and find

$$\exp\{-\beta F[\vec{e}, 0]\} = 128 \frac{\beta R^2}{V} \sum_{\vec{k}\in Z^3} \prod_{i=1}^{3} \exp\{(-\pi g^2 \beta a_i^2/2V)(k_i + e_i/2)^2\}. \tag{2.80}$$

The $\vec{e}$-independent normalization factor $\beta R^2/V$ has been chosen for convenience. The compactification radius R appears explicitly in this formula. This is not surprising since R regularizes an infrared divergence.

This formula is already interesting since it means that the free energy of a nonzero electric flux behaves as a power. By denoting with L the size of the box this energy is $O(g^2/L)$. This fact suggests that N=4 super Yang-Mills is always in the Coulomb phase [41]. A splitting $O(g^2/L)$ between the zero-flux sector and the nonzero-flux ones is peculiar to supersymmetric theories. In non-supersymmetric models, as pure Yang-Mills for instance, the classical degeneracy of the toron configurations is lifted at the quantum level,as we have seen. As a result the only true vacua are the trivial one $\vec{c} = 0$ and its 2^3 central conjugates. In this latter case the energy of an electric flux is zero to all orders in perturbation theory. Indeed, perturbation theory itself is different, in that the perturbative wave function is localized around the (discrete) true vacua. On the other hand in N=4 supersymmetry the vacuum degeneracy persists at least to all orders in perturbation theory.

The free energy $F[0,\vec{m}]$ at $\theta = 0$ and $g \gg 1$ (case b) above) can be found as follows. Strong arguments in favor of the $g \to 4\pi/g$ duality at $\theta = 0$ were given in ref. [38]. The core of the reasoning there is that magnetic monopoles in N=4 can be arbitrarily light: their mass depends on the scale of gauge symmetry breaking, which is arbitrary, being associated with a flat direction. The effective Lagrangian of light composite fields is expected to be renormalizable (this is a consequence of the so-called "Veltman's theorem" [54]). A N=4 renormalizable Lagrangian is completely fixed by the gauge group, and the gauge group that acts on magnetic monopoles is isomorphic to the original one, which acts on elementary fields. Moreover, the fact that N=4 supersymmetry gives vanishing beta functions means that the monopole effective Lagrangian, besides being an infrared fixed point, it is also an *ultraviolet* fixed point. Therefore this Lagrangian describes correctly monopole dynamics at all scales. This Lagrangian has the same form of the original one, Eq. (2.2), with coupling constant $g_M \equiv 4\pi/g$. We must also remark that, independently from this argument, the Lagrangian for the monopole zero modes (which are the relevant ones for our computation) has been explicitly found in ref. [50], and it coincides with Eq. (2.73). Therefore, one may evaluate $F[0,\vec{m}]$ in complete analogy with the electric case. The resulting free energy is given by Eq. (2.80), after the substitutions $g \to 4\pi/g$, $\vec{e} \to \vec{m}$.

Our aim now is to find a S-duality invariant formula for $F[\vec{e},\vec{m}]$, which reduces to $F[\vec{e},0]$ ($F[0,\vec{m}]$) in the weak (strong) coupling limit. The very existence of such an extension is far from obvious; moreover, $F[\vec{e},\vec{m}]$ must satisfy three non-trivial constraints.

a) It should factorize at $\theta = 0$ [41]: $F[\vec{e},\vec{m}] = F[\vec{e},0] + F[0,\vec{m}]$.

b) It should account for Witten's phenomenon [55]; namely, when $\theta \to \theta + 2\pi$ the electric flux $\vec{e}$ must transform into $\vec{e} + \vec{m}$.

c) It must obey 't Hooft's duality equations.

We are going to present an SL(2, Z)-invariant free energy obeying all these constraints. We believe that the existence of such a free energy is another strong argument in favor of S-duality. Besides, $F[\vec{e},\vec{m}]$ is a physical quantity since it is fully gauge-invariant. It is one of the simplest observables that can be used to investigate a non-Abelian gauge theory.

The S-duality in Eq. (2.1) transforms $M(\lambda)$,introduced in the previous section, as follows

$$M\left(\frac{a\lambda+b}{c\lambda+d}\right)=AM(\lambda)A^t,\quad A=\begin{pmatrix} d & c \\ b & a \end{pmatrix}. \tag{2.81}$$

Thus, an S-duality invariant generalization of Eq. (2.80) (and the corresponding one for $\vec{e}=0$, $\vec{m}\neq 0$) is

$$\exp\{-\beta F[\vec{e},\vec{m},\lambda]\} = 128\frac{\beta R^2}{V}\prod_{i=1}^{3}\sum_{p_i\in Z^2}\exp\left[-2\pi\beta\frac{a_i^2}{V}(p_i+f_i)^t M(\lambda)(p_i+f_i)\right].$$
$$p_i=\begin{pmatrix} k_i \\ l_i \end{pmatrix};\ k_i,\ l_i\in Z,\quad f_i=\begin{pmatrix} e_i/2 \\ m_i/2 \end{pmatrix}. \tag{2.82}$$

Eq. (2.82) is manifestly invariant under the transformation

$$\lambda\longrightarrow\frac{a\lambda+b}{c\lambda+d},\quad \begin{pmatrix} e_i/2 \\ m_i/2 \end{pmatrix}\longrightarrow\begin{pmatrix} a & -b \\ -c & d \end{pmatrix}\begin{pmatrix} e_i/2 \\ m_i/2 \end{pmatrix}. \tag{2.83}$$

Notice that electric and magnetic fluxes transform under duality, as expected. In particular, under $\lambda\longrightarrow 1/\lambda$, $b=c=-1$, $a=d=0$, and $\vec{e}\leftrightarrow\vec{m}$.

Eq. (2.82) obeys the constraints a), b) and c).

a) Factorization at $\theta=0$ is manifest.

b) Witten's phenomenon is correctly reproduced: under $\lambda\longrightarrow\lambda+1$ the matrix $M(\lambda)$ transforms into

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}M(\lambda)\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{2.84}$$

and our ansatz Eq. (2.82) becomes, after a redefinition of the dummy variable p,

$$\begin{aligned}\exp\{-\beta\ \ F\ \ [\vec{e},\vec{m},\lambda+1]\} &= \\ &= 128\frac{\beta R^2}{V}\prod_{i=1}^{3}\sum_{p_i\in Z^2}\exp\left[-2\pi\beta\frac{a_i^2}{V}(p_i+f_i')^t M(\lambda)(p_i+f_i')\right] \\ &= \exp\{-\beta F[\vec{e}+\vec{m},\vec{m},\lambda]\},\end{aligned} \tag{2.85}$$

with $f'^t_i=(e_i/2+m_i/2,m_i/2)$.

c) The least evident property that must be obeyed by $F[\vec{e},\vec{m},\lambda]$ is 't Hooft duality. Since the antiperiodic time boundary condition for the fermions violate t'Hooft duality, the left hand side of eq. (2.27) must be computed with the only antiperiodic fermion in the 3 direction. This is a very different object, but the euclidean invariance can be used to obtain its behaviour for g very small or g very large at $\theta = 0$ and the same trick to extend it to a full S-dual result. We find no difference in the result for different boundary conditions. To prove Eq. (2.27) for $\theta \neq 0$ for all the intermediate values of g, one may take, for simplicity, $e_3, m_3 = 0$. The right hand side of Eq. (2.27) then becomes

$$128\frac{1}{4}\sum_{q_i}\sum_{k\in Z^6}\exp\left\{2\pi i(q_1^t B f_1 + q_2^t B f_2) - 2\pi\frac{a_3 a_2}{\beta a_1}(k_1+q_1/2)^t M(\lambda)(k_1+q_1/2)\right.$$
$$\left. -2\pi\frac{a_3 a_1}{\beta a_2}(k_2+q_2/2)^t M(\lambda)(k_2+q_2/2) - 2\pi\frac{a_3\beta}{a_1 a_2}(k_3+f_3)^t M(\lambda)(k_3+f_3)\right\}. \tag{2.86}$$

The notations here are as follows

$$B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad q_i = \begin{pmatrix} r_i \\ s_i \end{pmatrix}, \quad r_i,\ s_i \in \{0,1\}. \tag{2.87}$$

Let us perform a Poisson resummation on the directions $i = 1, 2$

$$\sum_{q_i}\sum_{k\in Z^4}\exp\left\{2\pi i(q_1^t B f_1 + q_2^t B f_2) - 2\pi\frac{a_3 a_2}{\beta a_1}(k_1+q_1/2)^t M(\lambda)(k_1+q_1/2)\right.$$
$$\left. -2\pi\frac{a_3 a_1}{\beta a_2}(k_2+q_2/2)^t M(\lambda)(k_2+q_2/2)\right\} = \frac{\beta^2}{4a_3^2}\sum_{q_i}\sum_{k\in Z^4}\exp\left\{2\pi i(q_1^t B f_1 + \right.$$
$$+ q_2^t B f_2 + k_1^t q_1/2 + k_2^t q_2/2) - \frac{\pi}{2}\frac{\beta a_1}{a_2 a_3}k_1^t M(\lambda)^{-1}k_1$$
$$\left. -\frac{\pi}{2}\frac{\beta a_2}{a_1 a_3}k_2^t M(\lambda)^{-1}k_2\right\}. \tag{2.88}$$

Using the fact that $M^{-1}(\lambda) = BM(\lambda)B^t$, and that

$$\frac{1}{16}\sum_{q_i}\exp[2\pi i(q_1^t B f_1 + q_2^t B f_2 + k_1^t q_1/2 + k_2^t q_2/2)] = \begin{cases} 1 & \text{if } k_i = -2Bf_i \bmod 2 \\ 0 & \text{otherwise} \end{cases}, \tag{2.89}$$

one finds that, by substituting Eq. (2.88) into Eq. (2.86), this latter expression equals the left hand side of 't Hooft duality equation (2.27).

This completes the demonstration that our ansatz for the free energy does satisfy the three relevant physical constraints mentioned above.

We would like to stress again that in our ansatz the N=4 supersymmetric gauge theory lies in a Coulomb phase, without mass gap. In this phase the theory realizes the symmetry in Eq. (2.1) in a self-dual way. For instance, the theory can be smoothly deformed from a weak-coupling region ($g \ll 1$, $\theta = 0$) to a strong-coupling one ($g \gg 1$, $\theta = 0$) without undergoing a phase transition. This property is peculiar to the Coulomb phase, as remarked firstly by 't Hooft [41].

2.5 The case of SU(N)

The major difficulty in extending the results of previous section to the case of SU(N) or other gauge groups, is the proper counting of the physical states, which are not simply the plane waves on the toron manifold, but the Weyl invariant combinations that we can obtain from discrete gauge plane waves, fermionic zero modes and scalar plane wave zero modes.

The Lie algebra of SU(N) is the vector space of $N \times N$ hermitian traceless matrices. A convenient Cartan basis is given by the matrices

$$T^a = E^{aa} - E^{a+1,a+1} \tag{2.90}$$

for which the Cartan matrix has the standard form with 2 on the principal diagonal and -1 on the subdiagonals.

The lagrangian restricted to the modes constant in space reads

$$L = \frac{V}{g^2} \int dt (\sum \alpha_{ab} \dot{c}_i^a \dot{c}_i^b + \alpha_{ab} a_I^a a_I^b + \alpha_{ab} \phi_{IJ}^a \phi_{IJ}^b) \tag{2.91}$$

When we introduce conjugate momenta,

$$\pi_i^a = \frac{\delta}{\delta c_i^a} \tag{2.92}$$

the hamiltonian, for what regards the gauge degrees of freedom, is

$$H = \frac{g^2}{4V} \sum \alpha_{ab}^{-1} \pi_i^a \pi_i^a + scalars \tag{2.93}$$

The element

$$U_{(x)}^a(x,y,z) = e^{2\pi i \frac{x}{L} T^a} \tag{2.94}$$

to fix the given direction x, is well-defined and periodic (U(L)=1) in the box in which we have defined the system; thus it generates a gauge transformation, which acts on the zero modes as

$$c_i^a \rightarrow c_i^a + \frac{2\pi}{L} \tag{2.95}$$

Therefore the zero modes c_i^a are periodic of period $2\pi/L$.

The Hilbert space is constructed with the wavefunctions

$$e^{ik_i^a c_i^a a_i} \times fermions \times scalars \tag{2.96}$$

But the physical states are requested to be invariant under the Weyl group, which is composed by constant gauge transformations. The energy of a gauge bosons state can be written as

$$E = \frac{g^2}{4V} \sum a_i^2 \alpha_{ab}^{-1} k_i^a k_i^b = \frac{g^2}{4V} \sum a_i^2 k_i^2 \tag{2.97}$$

if $k_i = k_i^a f^a$ in the notations of section (1.4), and the integer momentum k can be interpreted as an element of the weight lattice Λ_W.

The partition function is not simply the sum of the statistical weights associated with these energies over all the weight lattice, i.e. is not the theta function

$$\sum_{k\epsilon\Lambda_W} e^{-costk^2} \tag{2.98}$$

since we have again to divide by the Weyl group. The energy term k^2, being the natural Cartan scalar product on the weight lattice is invariant, but the multiplicity with which we must count the single vector k depends on the number of invariant under Weyl group we can construct with the given vawefunctions.

If the problem were purely bosonic, the correct partition function would not be a theta function. This phenomena happens already for SU(2), where the state with momentum k and that with momentum -k become physically significative only in the symmetric combination, while the state with k=0 is already trivially invariant. So the result for the partition function would be

$$Z = 1 + \sum_{k>0} e^{-costk^2} \tag{2.99}$$

which is different from the theta function,

$$\theta = \sum_{k\epsilon Z} e^{-costk^2} \tag{2.100}$$

In SU(N) the situation is again more complicated. The problem, exactly as for SU(2), comes from the walls of the Weyl chambers. W acts freely on states in the interior of the chambers, while for every state on the walls there exists at least an element of the Weyl group which leaves it invariant; this causes a different molteplicity factor for interior, walls, walls of walls, and so on.

The introduction of fermions and scalars changes again the multiplicity, since now we have to combine gauge wavefunctions, fermionic zero modes and scalar zero modes, and only the combination of the three is needed to be invariant. The exact computation of the multiplicities is a complicated group theory problem; fortunately, the existence of the divergence dur to the scalars simplifies considerably the problem.

We face the problem to compute the partition function [56]

$$TrPe^{-\beta H} \tag{2.101}$$

on the Hilbert space constructed out of the vacuum with the 8 fermionic oscillators a_I^{a*}, the gauge exponential wavefunctions $e^{ik^a c_a}$ and the 6 scalar zero modes $e^{ip^a_{IJ} x_a}$. The states $|k, a, p>, k\epsilon\Lambda_W, p\epsilon R$ is a suitable Fock basis of eigenvalues of the hamiltonian with energy $k^2 + p^2$. $P = (1/|W|)\sum_{g\epsilon W} g$

is the projector over Weyl invariant states. We already know how the Weyl group acts on the states k; it is generated by the reflection about simple roots

$$k \to k - (f_a k) f_a \tag{2.102}$$

It takes the state $|k>= e^{ikc}$ into $|k^g>$, another state in the Fock space. It leaves invariant the states with $f_a k = 0, i.e. k_a = 0$, the walls of the Weyl chamber. It is the N-1 dimensional representation of the symmetric group S_N. On the fermionic zero modes a_I, W is realized linearly with the same $N-1 \times N-1$ matrices. Finally, W transforms the continuous momentum p exactly as k.

The partition function reads

$$TrPe^{-\beta H} = \frac{1}{|W|} \sum_{g\epsilon W} \sum_{k,a} \int dp < k,a,p|g|k,a,p > e^{-\beta k^2 - \beta p^2} \tag{2.103}$$

Now, the matrix element $< k|g|k >$ is different from zero only if k is left invariant by g; so we can restrict the sum over g to the little group W_k of k (the set of elements of W which leaves k fixed).

$$\frac{1}{|W|} \sum_{k\epsilon\Lambda_W} e^{-\beta k^2} \sum_{g\epsilon W_k} \sum_{a} \int dp < a,p|g|a,p > e^{-\beta p^2} \tag{2.104}$$

Since W acts freely in the interior of the Weyl chamber, W_k is trivial for such elements, becoming more and more large on the various walls of walls and being the complete Weyl group only on k=0.

Let us compute the matrix element $< a,p|g|a,p >$.

On the fermionic Fock space, g acts as a linear transformation; as $N-1 \times N-1$ matrices, the elements of W in this fundamental N-1 dimensional representation form a discrete subgroup of SO(N-1); so g can be diagonalized. On the vector subspace corresponding to the eigenvalue g_i of the matrix g, we have that the trace receives contributions from the only two states existing in the Fock subspace $|0>, a^*|0>$

$$\sum_a < a|g|a >= 1 + g_i \tag{2.105}$$

On the complete fermionic Fock space, taking into account also the spinorial (8) degrees of freedom, we obtain the result

$$det^8(1+g) \tag{2.106}$$

For what regards the scalars contribution, we have

$$< p|g|p >= \int dp \int_0^R dx e^{-\beta p^2 + ip(x - x^g)} \tag{2.107}$$

where we have put a cut-off R to regularize the configuration space divergence of the scalars; where possible we will send $R \to \infty$. We find

$$\int_0^R dx e^{-\frac{1}{4\beta} x(1-g)^2 x} = \frac{R^{n_g}}{det'^6(1-g)} \tag{2.108}$$

where we have take into account that there are 6 scalars; n_g is the number of eigenvalues of g equal to 1, and the prime means the esclusion of zero eigenvalues.

Summarizing, the total partition function reads

$$\sum_{k\epsilon\Lambda_W} e^{-\beta k^2}\frac{1}{|W|}\sum_{g\epsilon W_k} R^{n_g}\frac{det^8(1+g)}{det'^6(1-g)} \tag{2.109}$$

Now, it is clearly difficult to evaluate exactly the coefficients. However, we need to take in this sum only the most divergent contribution in R, since the subleading terms disappear when we normalize to get physical quantities. For each W_k there is only one element with the maximum power n_g, and so the maximum divergence, the identity. All the other elements contribute subleading terms. We see that the most divergent coefficient is always the same and henceforth the exact theta function gets reconstructed, up to a moltiplicative factor.

Now we have a candidate for the full partition function of N=4 Yang-Mills theory. Doubling the indices to take into account the existence of magnetic degrees of freedom, we find now that $\vec{k}_i^a$ has three indices: i=1,2,3 denoting the spatial direction, a=1,...,N-1 the Cartan index and $\rightarrow$ indicating the electric or magnetic nature of the state.

The plane wave energy is now

$$E=\sum_i \frac{\pi a_i^2}{V}\vec{k}_i^T M(\lambda)\vec{k}_i, \quad k_i\epsilon\Lambda_W \tag{2.110}$$

where M is the Sen matrix previously defined.

Our ansatz for the N=4 partition function for SU(N) is therefore

$$e^{-\beta F}=\sum_{\vec{k}_i\epsilon\Lambda_W} e^{-\pi\frac{\beta a_1}{a_2 a_3}\vec{k}_1 M\vec{k}_1-\pi\frac{\beta a_2}{a_1 a_3}\vec{k}_2 M\vec{k}_2-\pi\frac{\beta a_3}{a_1 a_2}\vec{k}_3 M\vec{k}_3} \tag{2.111}$$

Conditions a) and b) of previous section are trivially satisfy also by this generalization. What is a non trivial check is t'Hooft duality.

We have again to understand how the partition function splits into flux sectors according to the center Z_N. The different sectors are generated by gauge transformations periodic up to elements of the center.

For example,

$$U(x)=e^{2\pi i\frac{x}{L}\frac{m}{N}T} \tag{2.112}$$

where T is diagonal a matrix with all element 1 except for the last -N+1 (to make the matrix traceless) satisfies

$$U(L)=U(0)e^{2\pi i\frac{m}{N}} \tag{2.113}$$

thus generating m electric elementary flux in the direction x. Now, $T=\sum aT^a$; defining the generator of the center Z_N

$$\Omega=e^{2\pi i\frac{x}{L}\frac{1}{N}\sum_{a=1}^{N-1} aT^a} \tag{2.114}$$

we see that it acts on wavefunctions as

$$\psi = e^{i a_i c_i^a k_i^a} \longrightarrow e^{2\pi i \frac{1}{N} \sum_{a=1}^{N-1} a k^a} \psi \qquad (2.115)$$

The projector on states of electric flux e (mod N) being,

$$P[e] = \frac{1}{N} \sum_{m=0}^{N-1} e^{-2\pi i \frac{em}{N}} \Omega^m \qquad (2.116)$$

we find that a state k is invariant if and only if

$$\frac{1}{N} \sum_{m=0}^{N-1} e^{-2\pi i \left(\frac{me}{N} - m \frac{\sum a k^a}{N} \right)} = 1 \qquad (2.117)$$

i.e. if

$$\sum_{a=1}^{N-1} a k^a = e(modN) \qquad (2.118)$$

But we already know (see section (1.4)) that this condition is exactly that which divides Λ_W in N copies of Λ_R. The partition function of fixed flux are the old functions

$$T_\mu(t) = \sum_{\Lambda_R} e^{-\beta t (r + \mu f^1)^2} \qquad (2.119)$$

of which we know the nice modular properties, crucial for verifying t'Hooft duality,

$$T_\mu(t) = \frac{1}{\sqrt{N} t^{N-1/2}} \sum_{\nu=0}^{N-1} e^{2\pi i \frac{\mu\nu}{N}} T_\nu(\frac{1}{t}) \qquad (2.120)$$

The free energy at fixed electric and magnetic flux is

$$e^{-\beta F(\vec{e_i}, a_i, \beta)} = \sum_{\vec{r_i} \epsilon \Lambda_R} e^{-\sum_i \pi \frac{\beta a_i}{a_j a_k} (\vec{r} + \vec{e} f^1)_i^T M(\lambda) (\vec{r} + \vec{e} f^1)_i} \qquad (2.121)$$

Let us verify t'Hooft duality up to a normalization factor.

The generalization of (2.120) we need is

$$T_{\vec{e}} \left(\frac{\beta a_1}{a_2 a_3} M \right) = cost \sum_{\vec{k}=0}^{N-1} e^{2\pi i \frac{\vec{e}\vec{k}}{N}} T_{\vec{k}} \left(\frac{a_2 a_3}{\beta a_1} M^{-1} \right) \qquad (2.122)$$

We need to verify

$$e^{-\beta F(\vec{e_i}, a_i, \beta)} = \frac{1}{N^2} \sum_{\vec{k_1}, \vec{k_2}} e^{\frac{2\pi i}{N} (\vec{k_1} \Omega \vec{e_1} + \vec{k_2} \Omega \vec{e_2}) - a_3 F(\vec{k_1}, \vec{k_2}, \vec{e_3}; a_2, a_1, \beta, a_3)} \qquad (2.123)$$

where $\Omega = -i\sigma_1$.

The factor corresponding to direction 3 is invariant in both sides. The factor in direction 1, for example, produces

$$\sum_{\vec{k_1}} e^{\frac{2\pi i}{N} \vec{k_1} \Omega \vec{e_1}} T_{\vec{k_1}} \left(\frac{a_3 a_2}{a_1 \beta} M \right) = cost T_{\Omega \vec{e_1}} \left(\frac{\beta a_1}{a_2 a_3} M^{-1} \right) \qquad (2.124)$$

Now in the exponential of the function in the right hand side, since $M^{-1} = -\Omega M \Omega$, $\Omega^T = -\Omega$, $\Omega^2 = -1$, we have

$$(\Omega \vec{r} - \vec{e_1} f^1) M (\Omega \vec{r} - \vec{e_1} f^1) \tag{2.125}$$

and, changing variables, $\vec{r} \rightarrow -\Omega \vec{r}$, we obtain for the right hand side

$$T_{\vec{e}} \left(\frac{\beta a_1}{a_2 a_3} M \right) \tag{2.126}$$

which is exactly the factor in the left hand side. The same happens for direction 2. So, t'Hooft duality is verified.

Bibliography

[1] A. Migdal, Zh. Eksp. Teor. Fiz. 69, 810 (1975)(Sov. Phys. Jetp. 42, 413)

[2] D. Gross, Nucl. Phys. B400 (1993) 161; D. Gross and W. Taylor, Nucl. Phys. B400 (1993) 181, B403 (1993) 395.

[3] S. Cordes, G. Moore and S. Ramgoolam, YCTP-P23-93, RU-94-20, hep-th/9402107

[4] J. A. Minahan and A. P. Polychronakos, Phys. Lett. B312 (1993) 155; CERN-TH-7016-93, hep-th/9309119

[5] A. Sen, Nucl. Phys. 304B (1993) 109; Phys. Lett. 303B (1993); Int. Jou. Mod. Phys. A8 (1993) 5079; Mod. Phys. Lett. A8 (1993) 2023; preprints TIFR/TH/94-03, hep-th/9402002; TIFR/TH/94-08, hep-th/9402032.

[6] C. Vafa and E. witten, hep-th/9408074, IASSNS-HEP-94-54, HUPT-94/A017.

[7] N. Seiberg and E. Witten, RU-94-52, IAS-94-43, hep-th/9407087; N. Seiberg and E. Witten, hep-th/9408099, RU-94-60, IASSNS-HEP-94/55.

[8] G. 't Hooft, Nucl. Phys. 79B (1974) 276.

[9] V. Kazakov and I. Kostov, Phys. Lett. B128 (1983) 316; I. Kostov, Phys. Lett. B138 (1984) 191; D. Weingarten, Phys. Lett. B90 (1990) 285

[10] D. Gross and E. Witten, Phys. Rev. D21 (1980) 446

[11] M. Douglas and V. A. Kazakov, LPTENS-93120, RU-93-17, hep-th/9305047

[12] E. Witten, J. Phys. G9 (1992) 303

[13] E. Witten, Comm. Math. Phys. 141 (1991) 153

[14] M. Bershadsky, S. Cecotti, H. Ooguri and C. Vafa, HUTP-93-A025 hep-th/9309140; P. Horava, EFI-93-66, hep-th/9311156

[15] C. Itzykson, Int. Journ. Mod. Phys. A1 (1986) 65

[16] M. R. Douglas, hep-th/9303159, 9311130

[17] M. Caselle, A. D'Adda, L. Magnea, S. Panzeri, DFTT-50-93 hep-th/9309107

[18] H. Weyl, The classical group, Princeton Univ. Press, Princeton, N. Y. 1946.

[19] E. Brezin, C. Itzykson, G. Parisi and J. B. Zuber, Comm. Math. Phys. 50 (1978) 35

[20] D. Gross and I. Klebanov, Nucl. Phys. B344 (1990) 475

[21] M. R. Douglas, K. Li and M. Staudacher, Nucl. Phys. B240 (1994) 18.

[22] R. Rudd, RU-94-58, hep-th/9407156

[23] A. C. Pipkin, A course in integral equation, Berlin, Springer, 1991.

[24] G. Paffuti and P. Rossi, Phys. Lett. B92 (1980) 321; D. Friedan, Comm. Math. Phys. 78, (1981) 353

[25] P. Rossi, Ann. of Phys. 132, (1981) 463

[26] P. Menotti and E. Onofri, Nucl. Phys. B190 (1981) 288.

[27] M. L. Mehta, Comm. Math. Phys. 79 (1981) 327.

[28] C. Itzykson and J.-B. Zuber, J. Math. Phys. 21(3) (1980) 411.

[29] D. Gross and A. Matytsin, PUPT-1459, hep-th/9404004

[30] G. M. Cicuta, L. Molinari and E. Montaldi, Mod. Phys. Lett. A1 (1986) 125

[31] L. girardello, M. Pernici and A. Zaffaroni, in preparation

[32] M. L. Mehta, Random Matrices, Boston, Academic press 1991

[33] A. Font, L. Ibanez, D. Lust and F. Quevedo, Phys. Lett. 249B (1990) 35; S.J. Rey, Phys. Rev. D43 (1991) 526.

[34] A. Strominger, Nucl. Phys. 343B (1990) 167; C. Callan, J. Harvey and A. Strominger, Nucl. Phys. 359B (1991) 40; S.J. Rey, Phys. Rev. D43 (1991) 526; M. Duff, Class. Quant. Grav. 5 (1988) 189; M. Duff and J. Lu, Nucl. Phys. 354B (1991) 129; 357B (1991) 354; Phys. Rev. Lett. 66 (1991) 1402; Class. Quant. Grav. 9 (1991) 1; M. Duff, R. Khuri and J. Lu, Nucl. Phys. 377B (1992) 281; J. Dixon, M. Duff and J. Plefka, preprint CTP-TAMU-60/92, hep-th/9208055; I. Pesando and A. Tollsten, Phys. Lett. 274B (1992) 374.

[35] C. Montonen and D. Olive, Phys. Lett. 72B (1977) 213.

[36] M.K. Prasad and C.M. Sommerfeld, Phys. Rev. Lett. 35 (1975) 760.

[37] H. Osborn, Phys. Lett. 83B (1979) 321.

[38] P. Rossi, Phys. Lett. 99B (1981) 229.

[39] E. Witten and D. Olive, Phys. Lett. 78B (1978) 99.

[40] J.H. Schwarz and A. Sen, Nucl. Phys. 411B (1994) 35; Phys. Lett. 312B (1993) 105; J.P. Gauntlett and J.A. Harvey, preprint EFI-94-11, hep-th/9403072.

[41] G. 't Hooft, Nucl. Phys. 153B (1979) 141.

[42] L. Girardello, A. Giveon, M. Porrati and A. Zaffaroni, Phys. Lett. B334 (1994) 331

[43] H. B. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45

[44] M. Georgi and S. L. Glashow, Phys. Rev. Lett. 28 (1972) 1494

[45] M. Lüscher, Nucl. Phys. 219B (1983) 261.

[46] P. Goddard, J. Nyuts and D. Olive, Nucl. Phys. B125 (1977) 1

[47] M. Grisaru, M. Roček and W. Siegel, Phys. Rev. Lett. 45 (1980) 94; L.V. Avdeev, D.V. Tarasov and A.A. Vladimirov, Phys. Lett. 96B (1980) 94; W.E. Caswell and D. Zanon, Phys. Lett. 100B (1981) 152.

[48] S. Mandelstam, Nucl. Phys. 213B (1983) 149; P.S. Howe, K.S. Stelle and P.K. Townsend, Nucl. Phys. 214B (1983) 519; 236B (1984) 125; L. Brink, O. Lindgreen and B. Nilsson, Nucl. Phys. 212B (1983) 401; Phys. Lett. 123B (1983) 328; P.S. Howe, K.S. Stelle and P.C. West, Phys. Lett. 124B (1983) 55; P.C. White, Class. Quant. Grav. 9 (1992) 413.

[49] M.A. Shifman, A.I. Vainshtein and M.I. Vysotsky, Nucl. Phys. 254B (1985) 619; N. Seiberg, Phys. Lett. 206B (1988) 75.

[50] J.P. Gauntlett, Nucl. Phys. 400B (1993) 103; preprint EFI-93-09, hep-th/9305068; J. Blum, preprint EFI-94-04, hep-th/9401133.

[51] G. Segal, to appear

[52] E. Witten, Nucl. Phys. 202B (1982) 253.

[53] M. Sohnius and P. West, Phys. Lett. 100B (1981) 245.

[54] M. Veltman, Acta Phys. Polon. B12 (1981) 437.

[55] E. Witten, Phys. Lett. 86B (1979) 283.

[56] C. Imbimbo and S. Mukhi, to appear