Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD

N. Seiberg [a,b], E. Witten [b]

[a] *Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08855-0849, USA*
[b] *Institute for Advanced Study, Princeton, NJ 08540, USA*

Abstract

We study four dimensional $N = 2$ supersymmetric gauge theories with matter multiplets. For all such models for which the gauge group is SU(2), we derive the exact metric on the moduli space of quantum vacua and the exact spectrum of the stable massive states. A number of new physical phenomena occur, such as chiral symmetry breaking that is driven by the condensation of magnetic monopoles that carry global quantum numbers. For those cases in which conformal invariance is broken only by mass terms, the formalism automatically gives results that are invariant under electric-magnetic duality. In one instance, this duality is mixed in an interesting way with SO(8) triality.

1. Introduction

The holomorphic properties of supersymmetric field theories [1–3] can be a powerful tool in deriving exact results about them [3–6]. In $N = 1$ theories in four dimensions, the superpotential and the coefficients of the gauge kinetic terms are holomorphic and are constrained by these considerations. In four-dimensional $N = 2$ theories, the Kähler potential is also constrained by holomorphy [7] and therefore it can be analyzed similarly. This was used in the weak coupling analysis of Ref. [8] and in the exact treatment of Ref. [9]. The ability to make exact statements in these four-dimensional strongly coupled field theories makes them interesting laboratories where various ideas about quantum field theory can be tested, as has been seen in various $N = 1$ [4] and $N = 2$ [9] theories.

An important element in the analysis of these theories is the fact that supersymmetric field theories often have a continuous degeneracy of inequivalent ground states. Classically, they correspond to flat directions of the potential along which the squarks acquire

expectation values which break the gauge symmetry. The singularities in the moduli space of classical ground states are the points where the gauge symmetry is enhanced. Quantum mechanically, the vacuum degeneracy can be lifted by non-perturbative effects [10]. Alternatively, the vacuum degeneracy can persist and the theory then has a quantum moduli space of vacua. One then would like to know whether this space is singular and if so what is the physics of the singularities. This question has been studied in some $N = 1$ theories in Ref. [4]. The holomorphy of the superpotential enables one to determine the light degrees of freedom, and the quantum moduli space. In Ref. [9] the pure gauge $N = 2$ theory has been analyzed. In this case, the masses of the stable particles, the low energy effective interactions, and the metric on the quantum moduli space can be determined.

The classical moduli space of the $N = 2$ SU(2) gauge theory is parametrized by $u = \langle \mathrm{Tr}\, \phi^2 \rangle$ where ϕ is a complex scalar field in the adjoint representation of the gauge group. For $u \neq 0$ the gauge symmetry is broken to U(1). At $u = 0$ the space is singular and the gauge symmetry is unbroken. Our main goal is to determine – as quantitatively as possible – how this picture is modified quantum mechanically.

The quantum moduli space is described by the global supersymmetry version of special geometry. The Kähler potential

$$K = \mathrm{Im}\, a_D(u)\bar{a}(\bar{u}) \tag{1.1}$$

determines the metric or equivalently the kinetic terms. The pair (a_D, a) is a holomorphic section of an SL$(2, \mathbb{Z})$ bundle over the punctured complex u plane. They are related by $N = 2$ supersymmetry to a U(1) gauge multiplet. a is related by $N = 2$ to the semiclassical "photon" while a_D is related to its dual – "the magnetic photon."

For large $|u|$ the theory is semiclassical and

$$a \cong \sqrt{2u},$$
$$a_D \cong i\frac{2}{\pi} a \log a. \tag{1.2}$$

These expressions are modified by instanton corrections [8]. The exact expressions were determined in Ref. [9] as the periods on a torus

$$y^2 = (x^2 - \Lambda^4)(x - u) \tag{1.3}$$

of the meromorphic one-form

$$\lambda = \frac{\sqrt{2}}{2\pi} \frac{\mathrm{d}x\,(x - u)}{y}. \tag{1.4}$$

In (1.3), Λ is the dynamically generated mass scale of the theory.

The spectrum contains dyons labeled by various magnetic and electric charges. Stable states with magnetic and electric charges (n_m, n_e) have masses given by the BPS formula [11–13]

$$M^2 = 2|Z|^2 = 2|n_e a(u) + n_m a_D(u)|^2. \tag{1.5}$$

There are two singular points on the quantum moduli space at $u = \pm\Lambda^2$; they are points at which a magnetic monopole becomes massless. When an $N = 2$ breaking but $N = 1$

preserving mass term is added to the theory, these monopoles condense, leading to confinement [9].

Here, we extend our analysis to theories with additional $N = 2$ matter multiplets, known as hypermultiplets. As in Ref. [9], we limit ourselves to theories with gauge group SU(2). If matter multiplets are to be included while keeping the beta function zero or negative, there are the following possibilities. One can consider a single hypermultiplet in the adjoint representation. If the bare mass is zero, this actually gives a theory with $N = 4$ supersymmetry; it is possible to add a bare mass breaking the symmetry to $N = 2$. Or one can add N_f hypermultiplets in the spin one-half representation of SU(2), with an arbitrary bare mass for each multiplet; this preserves asymptotic freedom for $N_f \leqslant 3$, while the β function vanishes for $N_f = 4$. (It has been known for some time that the perturbative beta function vanishes for $N_f = 4$; our results make it clear that this is true nonperturbatively.) For all these theories, we will obtain the same sort of exact results that we found in Ref. [9] for the pure $N = 2$ gauge theory.

One motivation for studying these systems is that just as the pure gauge theory taught us something about confinement, the theory with matter may teach us about chiral symmetry breaking. In fact, we will find a new mechanism for chiral symmetry breaking – it arises in some of these models (those with $N_f = 2, 3, 4$) from the condensation of magnetic monopoles which carry global quantum numbers. These magnetic monopoles can be continuously transformed into elementary quanta as parameters are varied! That bizarre-sounding statement, which is possible because of the non-abelian monodromies, also means that one can interpolate continuously from the confining phase (triggered by condensation of monopoles) to the Higgs phase (triggered by condensation of elementary quanta). The fact that these two phases are in the same universality class is, of course, expected for $N_f > 0$ as elementary doublets are present [14].

In studying confinement in Ref. [9] and confinement and chiral symmetry breaking in the present paper, we are mainly working in a region close to a transition to a Higgs phase (in which the unbroken symmetry group is abelian) and apparently far from the usual strongly coupled gauge theories of gauge bosons and fermions only. However, one can reduce to a more usual situation by adding suitable $N = 1$-invariant perturbations of the superpotential. For instance, in Ref. [9] we exploited the possibility of perturbing the superpotential by $m\text{Tr}\,\Phi^2$, with m a complex parameter and Φ an $N = 1$ chiral multiplet related to the gauge bosons by $N = 2$ supersymmetry. (The absolute value of m is a bare mass, and its phase determines the parity violation in certain Yukawa couplings.) Our analysis of confinement was valid for small m, while a theory much more similar to ordinary QCD would emerge for large m.

But all experience indicates that as long as supersymmetry is unbroken, supersymmetric theories in four dimensions do not have phase transitions in the usual sense as a function of the complex parameters such as m; such transitions are more or less prevented by holomorphy. In the usual study of phase transitions, one meets singularities – phase boundaries – of real codimension one. By contrast, one gets in the supersymmetric case singularities of complex codimension $\geqslant 1$ in the space of vacua, permitting continuous interpolations from one regime to another. (The only situation in which one meets a singularity in interpolating from one regime to another is that in which the moduli space of vacua has several branches, which intersect somewhere, and one wishes to in-

terpolate from one branch to another.) Note that, as long as supersymmetry is unbroken, the energy vanishes and so the usual mechanism behind ordinary phase transitions – minimizing the energy – does not operate. The structures that have emerged for small m in Ref. [9] and the present paper are qualitatively similar to what one would guess (by analogy with QCD) for large m, and we do believe that the large m and small m theories are in the same universality class.

Another motivation for the present work is quite different: we will gain new insights about electric-magnetic duality in strongly interacting gauge theories. We will see, for instance, that in those theories in which conformal invariance is broken only by mass terms, the formalism is inevitably invariant under electric-magnetic duality. There are two relevant examples. One is the $N = 4$ theory, which is the original arena of Olive-Montonen duality [15]. This duality was originally formulated as a $\mathbb{Z}_2$ symmetry, in terms of the coupling constant only, but when the θ angle is included it can be extended to an action of $SL(2, \mathbb{Z})$ on τ [16–18]. The other relevant case is the $N_f = 4$ theory. In both of these examples we will find a full $SL(2, \mathbb{Z})$ symmetry exchanging electric and magnetic charges. (However, because of a factor of two in the conventions that will be explained below, $SL(2, \mathbb{Z})$ is defined differently in the two cases.) In many ways, the richest behavior that we find is that of the $N_f = 4$ theory; it has an $SO(8)$ global symmetry, and it turns out that $SL(2, \mathbb{Z})$ duality is mixed with $SO(8)$ triality.

We begin our discussion in Section 2 with a warm-up example of an abelian theory with $N = 2$ supersymmetry. It exhibits some of the new elements which will be important later on. Then we turn to $N = 2$ QCD and discuss its classical properties in Section 3. In Section 4 we begin the analysis of the quantum theory. Section 5 is devoted to the stable particles in the theory – the BPS saturated states. We discuss their masses and quantum numbers. Section 6 deals with duality transformations. Here we mention only the differences compared with the situation in the pure gauge theory which we discussed in Ref. [9]. In Section 7 we motivate our suggestion for the qualitative structure of the moduli space for $N_f = 1, 2, 3$ – the number of singularities and their nature. Section 8 describes the low energy theory near the singularities and exhibits non-trivial consistency checks of our suggestion. In Section 9 we break $N = 2$ to $N = 1$ supersymmetry and recover the results in Ref. [4]. In Section 10 we suggest the qualitative structure on the moduli space for the theory with $N_f = 4$. Section 11 is an introduction to the more quantitative discussion of the metric in the remaining sections. In Sections 12, 13 and 14 we find the metric on the moduli space for $N_f = 1$, 2 and 3 respectively. Then, the masses of the particles and some consistency checks are discussed in Section 15. In Section 16 we analyze two scale invariant theories – the theory with $N = 4$ supersymmetry and the $N = 2$ theory with four flavors. In both cases we can turn on $N = 2$ preserving mass terms and solve for the metric on the moduli space. Our previous answers for the metric are obtained by taking appropriate scaling limits as some masses go to infinity. Duality invariance of the results is manifest. In Section 17 we show a highly non-trivial consistency check of our answers.

A note on conventions.

The $N_f > 0$ theories have fields in the two-dimensional representation of $SU(2)$, so they have particles of half-integral electric charge if we use the same normalization as

in Ref. [9]. Instead, we will multiply n_e by 2, to ensure that it is always integral, and compensate by dividing a by 2. The asymptotic behavior is thus

$$a \cong \tfrac{1}{2}\sqrt{2u}\,,$$
$$a_D \cong i\frac{4}{\pi}a\log a\,. \tag{1.6}$$

Because of this change of normalization, the effective coupling constant $\tau = \partial a_D / \partial a$ is also rescaled and now $\tau = (\theta/\pi) + (8\pi i/g^2)$. Correspondingly, the family of curves (1.3) is replaced by a different family that will be described later.

2. A warm-up example: QED with matter

2.1. QED

For background, we first consider abelian gauge theories with $N = 2$ supersymmetry and charged matter hypermultiplets – that is, the $N = 2$ analog of ordinary QED.

The "photon," A_μ is accompanied by its $N = 2$ superpartners – two neutral Weyl spinors λ and ψ that are often called "photinos," and a complex neutral scalar a. They form an irreducible $N = 2$ representation that can be decomposed as a sum of two $N = 1$ representations[1] : a and ψ are in a chiral representation, A, while A_μ and λ are in a vector representation, W_α.

We take the charged fields, the "electrons," to consist of k hypermultiplets of electric charge one. Each hypermultiplet, for $i = 1, \dots, k$, consists of two $N = 1$ chiral multiplets M^i and $\widetilde{M}_i$ with opposite electric charge; of course such an $N = 1$ chiral multiplet contains a Weyl fermion and a complex scalar.

The renormalizable $N = 2$ invariant Lagrangian is described in an $N = 1$ language by canonical kinetic terms and minimal gauge couplings for all the fields as well as a superpotential

$$W = \sqrt{2}AM^i\widetilde{M}_i + \sum_i m_i M^i \widetilde{M}_i\,. \tag{2.1}$$

The first term in related by $N = 2$ supersymmetry to the gauge coupling and the second one leads to $N = 2$ invariant mass terms.

Consider first the massless theory ($m_i = 0$). Its global symmetry is $SU(k) \times SU(2)_R \times U(1)_{\mathcal{R}}$. $SU(k)$ acts on the k hypermultiplets, while $SU(2)_R \times U(1)_{\mathcal{R}}$ is the R symmetry group. The two supercharges are in a doublet of $SU(2)_R$ and their chiral components have charge one under $U(1)_{\mathcal{R}}$. The $4k$ real scalars in M and $\widetilde{M}$ transform like $(\mathbf{k}, \mathbf{2}, 0) \oplus (\overline{\mathbf{k}}, \mathbf{2}, 0)$ and the scalar a transforms as $(\mathbf{1}, \mathbf{1}, 2)$. Here and throughout the paper we denote representations by their dimensions, except for representations of $U(1)$ which are labeled by their charge.

---

[1] We use the conventions and notation of Ref. [19].

2.2. The classical moduli space

The classical moduli space of this theory has a branch with non-zero a. Along this flat direction the gauge symmetry is unbroken and all the M particles acquire a mass. Since the photon is massless, we will refer to this branch of the moduli space as the Coulomb branch.

For $k \geqslant 2$ and $m_i = 0$ there are also flat directions where M is non-zero. Along these directions the U(1) gauge symmetry is broken and therefore we will refer to this branch of the moduli space as the Higgs branch. Up to gauge and global symmetry transformations these flat directions are

$$M = (B, 0, \ldots)$$
$$\widetilde{M} = (0, B, 0, \ldots). \tag{2.2}$$

(Vanishing of the D terms requires $|M| = |\widetilde{M}|$.) The global $U(1)_\mathcal{R}$ is unbroken and we will ignore it. The other global symmetry is broken as

$$SU(k) \times SU(2)_R \rightarrow \begin{cases} SU(2)_{R'} & \text{for } k = 2 \\ U(1) \times SU(2)_{R'} & \text{for } k = 3 \\ SU(k-2) \times U(1) \times SU(2)_{R'} & \text{for } k \geqslant 4 \end{cases} \tag{2.3}$$

where $SU(2)_{R'}$ is a diagonal subgroup of $SU(2)_R$ and an $SU(2) \subset SU(k)$. The light fields on the moduli space are in hypermultiplets of $N = 2$ and therefore, the moduli space is a hyper-Kähler manifold. They transform like

$$\begin{array}{ll} \mathbf{3} \oplus \mathbf{1} & \text{for } k = 2 \\ (\mathbf{3}, \mathbf{2}) \oplus (-\mathbf{3}, \mathbf{2}) \oplus (\mathbf{0}, \mathbf{3}) + (\mathbf{0}, \mathbf{1}) & \text{for } k = 3 \\ (\mathbf{k-2}, \mathbf{1}, \mathbf{2}) \oplus (\overline{\mathbf{k-2}}, -\mathbf{1}, \mathbf{2}) \oplus (\mathbf{1}, \mathbf{0}, \mathbf{3}) \oplus (\mathbf{1}, \mathbf{0}, \mathbf{1}) & \text{for } k \geqslant 4 \end{array} \tag{2.4}$$

The boson in the last representation labels inequivalent vacua and the other bosons are the Goldstone bosons.

2.3. The quantum moduli space

Quantum mechanically, this theory is probably not well-defined because it is not asymptotically free. However, it is often the case that such a theory is embedded in a larger theory which is asymptotically free. In Ref. [9], we have encountered such a theory with $k = 1$ as the low energy limit of an asymptotically free SU(2) gauge theory. Below we will see more examples.

The metric on the Higgs branch cannot be corrected quantum mechanically. The reason for that is that the space is an almost homogeneous space which has a unique hyper-Kähler metric invariant under the symmetries. The metric on the Coulomb branch can be corrected. Since the low energy theory at the generic points on the Coulomb branch includes only the gauge multiplet, the metric on this branch is of the special geometry type, i.e. it is determined by the Kähler potential

$$K = \text{Im } a_D(a)\overline{a}. \tag{2.5}$$

The gauge kinetic energy is proportional to

$$\int d^2\theta \, \frac{\partial a_D}{\partial a} \, W_\alpha^2. \tag{2.6}$$

In this $N = 2$ theory, the one-loop approximation to K is exact (there are no higher order perturbative corrections and there are no $U(1)$ instantons on $\mathbb{R}^4$) leading to

$$a_D = -\frac{ik}{2\pi} a \log(a/\Lambda). \tag{2.7}$$

The lack of asymptotic freedom appears here as a breakdown of the theory at $|a| = \Lambda/e$ where the metric on the moduli space Im $\partial a_D/\partial a$ vanishes and the effective gauge coupling is singular. This is the famous Landau pole.

When the masses in (2.1) are not zero the moduli space changes. The singularities on the Coulomb branch can move. Whenever $a = -m_i/\sqrt{2}$ one of the electrons becomes massless. Therefore

$$a_D = -\frac{i}{2\pi} \sum_i (a + m_i/\sqrt{2}) \log \left(\frac{a + m_i/\sqrt{2}}{\Lambda} \right). \tag{2.8}$$

If some of the masses are equal, the corresponding singularities on the Coulomb branch coincide and there are more massless particles there. In this case a Higgs branch with non-zero expectation values for these electrons touches the Coulomb branch at the singularity. When there is only one massless electron hypermultiplet, the $|D|^2$ term in the potential prevents a Higgs branch from developing.

2.4. BPS-saturated states

The $N = 2$ algebra has "large" representations with 16 states and "small" ones with only four states. As explained in Ref. [13], the masses of particles in small representations are determined by their quantum numbers. Indeed, in terms of the central extension Z in the $N = 2$ supersymmetry algebra, the mass of a particle in a small representation is $M = \sqrt{2}|Z|$.

The $N = 2$ algebra requires that Z is a linear combination of conserved charges. In Ref. [13] and also in Ref. [9], Z was a linear combination of the electric and magnetic quantum numbers n_e and n_m. The classical expression can be written $Z = n_e a + n_m a_D$ and this way of writing it is also valid quantum mechanically in the pure gauge theory, as was explained in Ref. [9]. When there are additional abelian conserved charges, they might conceivably appear in the formula for Z.

Such a modification is present in the version of QED described above. One can easily deduce this as follows. First, of all, by counting states one sees that the "electrons" are in "small" representations. (In fact, any multiplet in which all states are of spin $\leqslant 1/2$ is automatically a sum of "small" representations.) On the other hand, the masses of the electrons are not $\sqrt{2}|a|$, as would follow from the "old" formula for Z, but rather the mass of the ith hypermultiplet is $|\sqrt{2}a + m_i|$. To give such a result, the $U(1)$ charges S_i of the hypermultiplets must appear in Z as follows:

$$Z = n_e a + n_m a_D + \sum_i S_i m_i / \sqrt{2}. \tag{2.9}$$

This result can easily be verified by computing the Poisson brackets of the supercharges. The new term in Z will pervasively affect the analysis below.

2.5. Breaking $N = 2$ to $N = 1$

We will later need a deformation of this theory breaking $N = 2$ to $N = 1$ supersymmetry by adding a term μA to the superpotential (2.1). This breaks the global $SU(2)_R$ symmetry. Consider for simplicity the case where all the electrons are massless ($m_i = 0$). Using the equations of motion and the $U(1)$ D terms, it is easy to see that the Coulomb branch collapses to a point $a = 0$. For $k = 1$ there is a unique ground state with $M = -\widetilde{M} = (\mu/\sqrt{2})^{1/2}$ and the $U(1)$ gauge symmetry is spontaneously broken. For $k \geqslant 2$ there are also M flat directions. Up to symmetry transformations they have the form

$$M = (C, 0, \ldots)$$
$$\widetilde{M} = (-\frac{\mu}{\sqrt{2}C}, B, 0, \ldots) \tag{2.10}$$

with $|B|^2 + |\frac{\mu}{\sqrt{2}C}|^2 = |C|^2$. At the generic point the global symmetry is broken as

$$SU(k) \rightarrow \begin{cases} 1 & \text{for } k = 2 \\ U(1) & \text{for } k = 3 \\ SU(k-2) \times U(1) & \text{for } k \geqslant 4 \end{cases} \tag{2.11}$$

and at the special point $B = 0$ it is broken as

$$SU(k) \rightarrow \begin{cases} U(1) & \text{for } k = 2 \\ SU(2) \times U(1) & \text{for } k = 3 \\ SU(k-1) \times U(1) & \text{for } k \geqslant 4 \; . \end{cases} \tag{2.12}$$

3. Classical moduli space of QCD with matter

We now turn to QCD with an $SU(2)$ gauge group. The gluons are accompanied by Dirac fermions and complex scalars ϕ in the adjoint representation of the gauge group. We also add N_f hypermultiplets of quarks in the fundamental representation. (We will also consider the case of a single hypermultiplet in the adjoint representation, this being the $N = 4$ theory.) As in the previous section, each hypermultiplet contains a Dirac fermion and four real scalars. In terms of $N = 1$ superfields the hypermultiplets contain two chiral superfields $Q^i{}_a$ and $\widetilde{Q}_{ia}$ ($i = 1, \ldots, N_f$ is the flavor index and $a = 1, 2$ the color index) and the $N = 2$ gauge multiplets include $N = 1$ gauge multiplets and chiral multiplets Φ. The superpotential for these chiral superfields is

$$W = \sqrt{2}\widetilde{Q}_i \Phi Q^i + \sum_i m_i \widetilde{Q}_i Q^i \tag{3.1}$$

with color indices suppressed.

When the quarks are massless the global symmetry of the classical theory is a certain quotient of $O(2N_f) \times SU(2)_R \times U(1)_{\mathcal{R}}$. The reason for the $O(2N_f)$ symmetry (rather than $\widetilde{SU}(N_f) \times U(1)$) is that for an $SU(2)$ gauge theory the quarks Q and the antiquarks $\widetilde{Q}$ are in isomorphic representations of the gauge group. Therefore, we will also often denote these $N = 1$ chiral superfields by Q^r with $r = 1, \ldots, 2N_f$ labeling the components of an $SO(2N_f)$ vector. The squarks $(Q, \widetilde{Q})$ transform like $(\mathbf{2N_f}, \mathbf{2}, 0)$ and the scalar in the gauge multiplet as $(\mathbf{1}, \mathbf{1}, 2)$. It will be important that the symmetry of the hypermultiplets is $O(2N_f)$ and not $SO(2N_f)$; for instance there is a "parity" symmetry, $\mathbb{Z}_2 \subset O(2N_f)$, acting as

$$\rho : Q_1 \leftrightarrow \widetilde{Q}_1 \tag{3.2}$$

with all other squarks invariant.

Globally, the symmetry group is not quite the product of $O(2N_F) \times SU(2)_R \times U(1)_{\mathcal{R}}$ with the Lorentz group. A $\mathbb{Z}_2 \subset U(1)_{\mathcal{R}}$ is isomorphic to $(-1)^F$ which is in the Lorentz group. Also, when combined with the center of the $SU(2)_R$, this $\mathbb{Z}_2$ acts the same as the $\mathbb{Z}_2$ in the center of $O(2N_f)$.

As in the abelian example, there is always a flat direction with non zero ϕ. Along this direction the gauge symmetry is broken to $U(1)$ and all the quarks are massive. Only $U(1)_{\mathcal{R}}$ is spontaneously broken there. We will refer to this branch of the moduli space as the Coulomb branch. For $N_f = 0, 1$ there are no other flat directions, but such directions appear (when $m_i = 0$) for $N_f \geqslant 2$. Since the gauge symmetry is completely broken along these flat directions, we will refer to them as the Higgs branches. The Higgs branches can be analyzed as follows.

First, it follows from requiring that the superpotential be stationary and the D terms vanish that on the Higgs branches, Φ must be zero. The flat directions in Q space are found by setting to zero the D terms, dividing by the gauge group $SU(2)$, and asking for the superpotential to be stationary. The combined operation of setting the D terms to zero and dividing by $SU(2)$ is equivalent to dividing by $SL(2, \mathbb{C})$. The quotient by $SL(2, \mathbb{C})$ of the space of squarks can be parametrized by the $SL(2, \mathbb{C})$-invariant functions $V^{rs} = Q^r{}_a Q^{sa}$ of the squarks; here $r, s = 1, \ldots, 2N_f$ and a are the flavor and color indices, and $V^{rs} = -V^{sr}$. V generates the ring of $SL(2, \mathbb{C})$-invariant polynomials in the Q's. The V's are not independent but obey certain quadratic equations stating that V is of rank two. For $N_f = 2$, there is a single such equation

$$\epsilon_{rstu} V^{rs} V^{tu} = 0. \tag{3.3}$$

We still must impose the condition that the superpotential should be stationary. Since the superpotential is linear in Φ, and $\Phi = 0$, the only non-trivial condition is $\partial W / \partial \Phi = 0$, or $X_{ab} = 0$ with $X_{ab} = \sum_r Q^r{}_a Q^r{}_b$. This is equivalent to

$$0 = Q^{ra} Q^{sb} X_{ab} = -V^{rt} V^{ts} \tag{3.4}$$

for arbitrary r, s.

For example, the above equations can be analyzed as follows for $N_f = 2$. The symmetry group $O(4)$ is locally $SU(2) \times SU(2)$, and the antisymmetric tensor V^{rs}

decomposes as $(\mathbf{3},\mathbf{1}) \oplus (\mathbf{1},\mathbf{3})$; we will call the two pieces V_L and V_R. The symmetric tensor in (3.4) transforms as $(\mathbf{1},\mathbf{1}) \oplus (\mathbf{3},\mathbf{3})$. The $(\mathbf{3},\mathbf{3})$ piece is bilinear in V_L and V_R and so vanishes if and only if $V_L = 0$ or $V_R = 0$. There are thus two Higgs branches, with V being self-dual or anti-self-dual. The $(\mathbf{1},\mathbf{1})$ part of (3.4) gives $V_L{}^2 = 0$ (or $V_R{}^2 = 0$, for the other branch), which actually duplicates the content of (3.3). By $V_L{}^2$ we mean of course $v_1{}^2 + v_2{}^2 + v_3{}^2$ where $v_r = V^{r4}$. The manifold given by the equation $v_1{}^2 + v_2{}^2 + v_3{}^2 = 0$ for three complex variables v_i is equivalent to the quotient $\mathbb{C}^2/\mathbb{Z}_2$,[2] and so admits a flat hyperkahler metric with a $\mathbb{Z}_2$ orbifold singularity at the origin.

Thus in particular, for $N_f = 2$, there are two Higgs branches which meet each other and the Coulomb branch at the origin. The two branches are exchanged by the "parity" symmetry generated by ρ of (3.2). For $N_f > 3$, instead, there is a single irreducible Higgs branch (this follows, for instance, from the symmetries) which meets the Coulomb branch at the origin.

Along the Higgs branches the $U(1)_\mathcal{R}$ symmetry is unbroken and the other global symmetry is broken as

$$O(2N_f) \times SU(2)_R \rightarrow \begin{cases} SU(2) \times SU(2)_{R'} & \text{for } N_f = 2 \\ O(2N_f - 4) \times SU(2) \times SU(2)_{R'} & \text{for } N_f \geqslant 3 \end{cases} \quad (3.5)$$

where the first $SU(2)$ is a diagonal subgroup of an $SU(2) \subset SO(4) \subset SO(2N_f)$ and the $SU(2)$ gauge symmetry. $SU(2)_{R'}$ is a diagonal subgroup of the other $SU(2) \subset SO(4) \subset SO(2N_f)$ and the original $SU(2)_R$ symmetry. The massless fields transform like

$$\begin{cases} (\mathbf{1},\mathbf{3}) \oplus (\mathbf{1},\mathbf{1}) & \text{for } N_f = 2 \\ (\mathbf{2N_f} - \mathbf{4},\mathbf{2},\mathbf{2}) \oplus (\mathbf{1},\mathbf{1},\mathbf{3}) \oplus (\mathbf{1},\mathbf{1},\mathbf{1}) & \text{for } N_f \geqslant 3 \end{cases} \quad (3.6)$$

The boson in the last representation labels inequivalent ground states. The other bosons are the Goldstone bosons.

4. A first look at the quantum theory

4.1. Symmetries of the quantum theory

Now we consider the quantum modifications to the symmetry structure. Since the one-loop beta function of the theory is proportional to $4 - N_f$ (higher order perturbative corrections to it vanish), we limit ourselves to $N_f = 0, 1, 2, 3$ where the theory is asymptotically free and to $N_f = 4$ where the theory will turn out to be scale invariant.

The global $U(1)_\mathcal{R}$ and the "parity" $Z_2 \subset O(2N_f)$ are anomalous. For $N_f > 0$, a discrete $\mathbb{Z}_{4(4-N_f)}$ anomaly free subgroup is generated by

$$W_\alpha \rightarrow \exp\left[\frac{i\pi}{2(4 - N_f)}\right] W_\alpha \left(\exp\left[\frac{-i\pi}{2(4 - N_f)}\right] \theta\right)$$

---

[2] Introduce new variables a, b, defined up to an overall sign, by $v_1 + iv_2 = a^2$, $v_1 - iv_2 = b^2$, $v_3 = iab$.

$$\Phi \to \exp\left[\frac{i\pi}{4 - N_f}\right] \Phi \left(\exp\left[\frac{-i\pi}{2(4 - N_f)}\right] \theta\right)$$

$$Q^1 \to \widetilde{Q}_1 \left(\exp\left[\frac{-i\pi}{2(4 - N_f)}\right] \theta\right)$$

$$\widetilde{Q}_1 \to Q^1 \left(\exp\left[\frac{-i\pi}{2(4 - N_f)}\right] \theta\right) \qquad\qquad (4.1)$$

with all other squarks invariant. (For $N_f = 0$, the Q's are absent and cannot be used to cancel an anomaly; the anomaly-free global symmetry is $\mathbb{Z}_8$ rather than $\mathbb{Z}_{16}$, and is generated by the square of the above.) A $\mathbb{Z}_2 \subset \mathbb{Z}_{4(4-N_f)}$ is equal to $(-1)^F$. We can combine this symmetry with an $SU(2)_R$ transformation to find a $\mathbb{Z}_{4(4-N_f)}$ symmetry which commutes with $N = 1$ supersymmetry

$$\Phi \to \exp\left[\frac{i\pi}{4 - N_f}\right] \Phi$$

$$Q^1 \to \exp\left[\frac{-i\pi}{2(4 - N_f)}\right] \widetilde{Q}_1$$

$$\widetilde{Q}_1 \to \exp\left[\frac{-i\pi}{2(4 - N_f)}\right] Q^1$$

$$Q^i \to \exp\left[\frac{-i\pi}{2(4 - N_f)}\right] Q^i$$

$$\widetilde{Q}_i \to \exp\left[\frac{-i\pi}{2(4 - N_f)}\right] \widetilde{Q}_i \qquad\qquad (4.2)$$

for $i \geqslant 2$. In this form it is clear that a $\mathbb{Z}_2$ subgroup of this group acts the same as a $\mathbb{Z}_2$ in the center of $SO(2N_f)$.

Since $u = \mathrm{Tr}\, \phi^2$ transforms as $u \to e^{2i\pi/(4-N_f)}$, the global symmetry acting on the u plane is $\mathbb{Z}_{4-N_f}$ for $N_f > 0$, or $\mathbb{Z}_2$ for $N_f = 0$.

4.2. A first look at the quantum moduli space

We now begin the analysis of the quantum moduli space. The first basic fact is that for large fields, the theory is weakly coupled and the quantum moduli space is well approximated by the classical moduli space.

Consider first the Higgs branches. The $SO(2N_f) \times SU(2)/SO(2N_f - 4) \times SU(2) \times SU(2)$ structure should persist quantum mechanically. This structure admits a unique hyper-Kähler metric up to a constant multiple (and the multiple is fixed by the behavior for large fields). For instance, for $N_f = 2$, the metric is the orbifold metric on $\mathbb{R}^4/\mathbb{Z}_2$. Therefore, there are no quantum corrections to the metric, and the singularity cannot be removed.

If these manifolds continue to touch the Coulomb branch in the quantum theory (as we will claim), one might be tempted to guess that the $SU(2)$ gauge symmetry should be restored there. The reason for that is that on the Coulomb branch there is always a massless photon while on the Higgs branches the three gauge bosons are degenerate. However, this assumes that the three massive gauge bosons always exist

as stable particles. If this is not so, it could be that (as we will eventually argue) the photon of the Coulomb branch is the only massless gauge boson at the point where the branches meet.

We now turn to discuss the Coulomb branch. We parametrize it by the gauge invariant coordinate $u = \langle \mathrm{Tr}\, \phi^2 \rangle$.

For $N_f = 0$ the anomaly free discrete symmetry (4.1) acts on u as $\mathbb{Z}_2$, as explained at the end of the last section. In analyzing instanton corrections to the metric on the u plane, we can treat the $U(1)_{\mathcal{R}}$ symmetry as unbroken by assigning charge 4 to u and charge 8 to the single instanton factor Λ_0^4.

For $N_f = 1, 2, 3$ the anomaly free discrete symmetry $\mathbb{Z}_{4(4-N_f)}$ described in (4.1) acts on u as $\mathbb{Z}_{4-N_f}$. The expectation value of u breaks the discrete symmetry to the $\mathbb{Z}_4$ that acts trivially on u. We can treat the $U(1)_{\mathcal{R}} \times \mathbb{Z}_2$ (this $\mathbb{Z}_2$ is the ρ symmetry (3.2)) as unbroken by assigning charge 4 and even parity to u and charge $2(4 - N_f)$ and odd parity to the instanton factor $\Lambda_{N_f}^{4-N_f}$.

For $N_f = 4$ the $U(1)_{\mathcal{R}}$ symmetry is anomaly free and u has charge 4. The "parity" $\mathbb{Z}_2 \subset O(8)$ is still anomalous. We can still treat it as unbroken by assigning odd parity to the instanton factor $q^{1/2} = e^{i\pi\tau} = e^{-(8\pi^2/g^2)+i\theta}$.

As for $N_f = 0$, the metric and the dyon masses are determined by a holomorphic section of an $SL(2, \mathbb{Z})$ bundle:

$$
\begin{aligned}
a &= \frac{1}{2}\sqrt{2u} + \ldots \\
a_D &= i\frac{4 - N_f}{2\pi} a(u) \log \frac{u}{\Lambda_{N_f}^2} + \ldots
\end{aligned}
\tag{4.3}
$$

where the ellipses represent instanton corrections and Λ_{N_f} is the dynamically generated scale of the theory with N_f flavors (we will later rescale it to a convenient value). The metric is $ds^2 = \mathrm{Im}\,(a_D'\bar{a}')du\,d\bar{u}$ and the dyon masses $M^2 = 2|Z^2|$ are expressed in terms of $Z = n_e a + n_m a_D$ where (n_m, n_e) are the electric and magnetic charges. As we said above, we use a normalization such that all electric charges are integers.

For $N_f \neq 0$, the contributions to (4.3) from terms with an odd number of instantons vanish: this follows from the anomalous $\mathbb{Z}_2$ in $O(2N_f)$. The amplitudes with odd instanton number are odd under this $\mathbb{Z}_2$, and so cannot generate contributions to the metric for u, which is even. However, there is no reason why the even instanton contributions of the form $(\Lambda_{N_f}^2/u)^{n(4-N_f)}$ should vanish. We therefore expect that

$$
\begin{aligned}
a &= \frac{1}{2}\sqrt{2u}\left(1 + \sum_{n=1}^{\infty} a_n(N_f)\left(\frac{\Lambda_{N_f}^2}{u}\right)^{n(4-N_f)}\right), \\
a_D &= i\frac{(4 - N_f)}{2\pi} a(u) \log \frac{u}{\Lambda_{N_f}^2} + \sqrt{u}\sum_{n=0}^{\infty} a_{Dn}(N_f)\left(\frac{\Lambda_{N_f}^2}{u}\right)^{n(4-N_f)}.
\end{aligned}
\tag{4.4}
$$

Under the $\mathbb{Z}_2$ or $\mathbb{Z}_{4-N_f}$ symmetry of the u plane, a transforms linearly and a_D picks up a multiple of a; the latter fact means that the symmetry shifts the electric charges of magnetic monopoles.

Use of a gauge invariant order parameter u is convenient because one can, for instance, determine the unbroken global symmetry without worrying about the possibility that a broken global symmetry becomes unbroken when combined with a gauge symmetry. However, since the SU(2) gauge symmetry is not completely broken, there are massive charged states in the spectrum and we should be careful in determining the way the unbroken global symmetry acts on them. In particular, the unbroken $\mathbb{Z}_4$ is generated by the operator in (4.2) raised to the $4 - N_f$ power. This transformation changes the sign of ϕ and so acts as charge conjugation on the charged fields. For $N_f = 1, 3$ this transformation acts as parity in $O(2N_f)$ (because an odd power of ρ appears), so the parity transformation is realized on the spectrum but reverses all electric and magnetic charges; the states of given charge are in representations of SO(2) or SO(6) only. For $N_f = 2, 4$, the parity symmetry is altogether spontaneously broken (since the unbroken symmetries all contain the parity raised to an even power), so the states are only in SO(4) or SO(8) representations.

5. BPS-saturated states

As we explained in our discussion of QED, a special role is played by the BPS-saturated states which are in "small" representations of the $N = 2$ algebra and are frequently the only stable states in the spectrum.

Since on the Higgs branches the gauge group is completely broken, there are no electric and magnetic charges that can appear as central charges in the $N = 2$ algebra. The only central charge could be a U(1) charge of a hypermultiplet. A Higgs branch exists only when there are at least two degenerate hypermultiplets (not necessarily with zero bare mass). Then, if there is an abelian symmetry acting on these hypermultiplets, the massless fields carry the corresponding charge, and it cannot appear in the central extension. Therefore, a contribution to the central extension exists only when there is also another hypermultiplet, which is not degenerate with the others (its bare mass might be zero). We do not know of striking physical phenomena associated with these states, mainly because in the absence of electric and magnetic charge there do not appear to be monodromies.

On the Coulomb branch, the simplest BPS-saturated states are the elementary quarks whose mass (if the bare masses of the hypermultiplets are set to zero, as we do until further notice) is $M = \sqrt{2}|a|$. They are in the vector representation of SO(2N_f).

Since the SU(2) gauge symmetry is spontaneously broken to U(1) along the Coulomb branch, there are also magnetic monopoles in the spectrum. For $N_f \neq 0$, the quark fields have fermion zero modes in the background of the monopoles [20–22]. To be precise, each SU(2) doublet of fermions has a single zero mode. With N_f hypermultiplets and therefore $2N_f$ doublets, there are $2N_f$ zero modes transforming in the vector representation of SO($2N_f$). Rather as in the quantization of the Ramond sector of superstrings, the quantization of these fermions zero modes turns the monopoles into spinors of SO($2N_f$). The states of definite charge furnish a representation of SO($2N_f$), not of O($2N_f$), for a reason noted at the end of the last section. Note that the occurrence of spinors (in addition to the hypermultiplets, which are vectors) means that at the

quantum level the symmetry group is really the universal cover of $SO(2N_f)$.

There is, however, an important subtlety here. Monopoles can carry electric charge because the classical monopole solution is not invariant under electric charge rotations. There is a collective coordinate associated with electric charge rotations, and quantizing it gives a spectrum of states of various electric charge. A 2π rotation by the electric charge operator does not give the identity; in a state of $n_m = 1$, it gives a topologically non-trivial gauge transformation, whose eigenvalue is $e^{i\theta}(-1)^H$; here θ is the usual theta angle and $(-1)^H$ is the center of the $SU(2)$ gauge group (it acts as -1 on the elementary hypermultiplets and 1 on the vector multiplet). This is the effect described in Ref. [23] (where the pure gauge theory was considered, so $(-1)^H$ was equivalent to 1). If the electric charge operator, which we will temporarily call Q_0, is normalized so that a W boson has unit charge, the operator statement is

$$e^{2\pi i Q_0} = e^{i n_m \theta}(-1)^H. \tag{5.1}$$

In the present context, we want to normalize the charge operator so that the eigenvalues for the hypermultiplets are ± 1 (so W bosons have charge ± 2). The normalized charge operator is thus $Q = 2Q_0$, and formula (5.1) becomes

$$e^{i\pi Q} = e^{i n_m \theta}(-1)^H. \tag{5.2}$$

It has the following significance: if we write the charge as $Q = n_e + n_m \theta/\pi$ with $n_e \in \mathbb{Z}$, then the states of even n_e have $(-1)^H$ even, and the states of odd n_e have $(-1)^H$ odd.

$(-1)^H$ is the "chirality" operator in the spinor representation of $SO(2N_f)$, so the above statement means that the monopoles of even n_e are in one spinor representation, and the monopoles of odd n_e are in the other spinor representation. This result ensures the following: if M is a monopole with $n_e = q$ and M' is a monopole with $n_e = q+1$, then the state of $(n_m = 0, n_e = 1)$ produced in $M'\overline{M}$ annihilation has $(-1)^H = -1$ and can be formed from the elementary fields. If there were no correlation between electric charge and $SO(2N_f)$ chirality, then monopole-antimonopole annihilation would produce states that do not in fact exist.

For $N_f = 1, 3$, the internal "parity" ensures that a dyon transforming as a positive chirality spinor representation of $SO(2N_f)$ is degenerate with a particle with opposite electric and magnetic charge and opposite $SO(2N_f)$ chirality. There is no such relation for $N_f = 2, 4$ where the internal parity is simply spontaneously broken.

As in Ref. [18], the spectrum may also include states with magnetic charge $n_m \geqslant 2$. Certain general restrictions on the quantum numbers of these states (some of which can be deduced from (5.2)) ensure that they can be interpreted as bound states of already known particles and that particle–antiparticle annihilation gives consistent results. For $N_f \geqslant 2$, these restrictions are conveniently stated in terms of the quantum numbers of the states under the center of the universal cover of $SO(2N_f)$. For $N_f = 2$, the universal cover is $SU(2) \times SU(2)$ and the center is $\mathbb{Z}_2 \times \mathbb{Z}_2$. We write a representation of $\mathbb{Z}_2 \times \mathbb{Z}_2$ as (ϵ, ϵ'), where $\epsilon = 0$ for the trivial representation of $\mathbb{Z}_2$ and $\epsilon = 1$ for the non-trivial representation. Then for $N_f = 2$, states of arbitrary (n_m, n_e) transform under the center as $((n_e + n_m) \bmod 2, n_e \bmod 2)$. For $N_f = 3$, the universal cover of $SO(6)$ is $SU(4)$, and its center is $\mathbb{Z}_4$. $\mathbb{Z}_4$ acts by $\exp(\frac{1}{2}i\pi(n_m + 2n_e))$. For $N_f = 4$, the universal cover is

Spin(8), with center $\mathbb{Z}_2 \times \mathbb{Z}_2$. The four representations of $\mathbb{Z}_2 \times \mathbb{Z}_2$ are conveniently labeled by representations of Spin(8) that realize them; we will call them o (associated with the trivial representation of Spin(8)), v (associated with the vector), and s and c (the two spinors). Spin(8) has a "triality" group of outer automorphisms which is isomorphic to the permutation group S_3 of the three objects v, s, and c. The quantum numbers of all particles under the center of Spin(8) are determined by ($n_m \bmod 2, n_e \bmod 2$): $(0, 1)$ corresponds to v, e.g. the elementary quark; $(1, 0)$ corresponds to s, e.g. the fundamental neutral monopole; $(1, 1)$ corresponds to c, e.g. the first excited dyon with magnetic charge 1; $(0, 0)$ corresponds to o, e.g. the elementary gauge fields.

We now turn on an $N = 2$ invariant mass term, m_{N_f}, for one of the quarks. The global $SO(2N_f)$ symmetry is explicitly broken to $SO(2N_f - 2) \times SO(2)$. Since the global symmetry includes an abelian continuous symmetry, it can contribute to the central extension in the algebra. As in QED (see Section 2), the mass of BPS-saturated states is given by $M = \sqrt{2}|Z|$ with

$$Z = n_e a + n_m a_D + S \frac{m_{N_f}}{\sqrt{2}}, \tag{5.3}$$

where S is the $SO(2)$ charge. Just as in QED, the appearance of the extra term can be deduced from the fact that the hypermultiplets are in "small" representations. It follows from (5.3) that for $a = \pm m_{N_f}/\sqrt{2}$ one of the elementary quarks is massless. This fact can be easily verified using the classical Lagrangian.

6. Duality

As in the pure gauge theory [9], we can perform $SL(2, \mathbb{Z})$ duality transformation on the low energy fields. Although they are non-local on the photon field A_μ, they act simply on (a_D, a). Several new issues appear when matter fields are present.

First, consider the situation of one massive quark with mass m_{N_f} and examine what happens when a approaches $m_{N_f}/\sqrt{2}$ where one of the elementary quarks becomes massless. As in the discussion in Section 2, loop diagrams in which this quark propagates make a logarithmic contribution to a_D. The behavior near $a = m_{N_f}/\sqrt{2}$ is thus

$$a \approx a_0,$$
$$a_D \approx c - \frac{i}{2\pi}(a - a_0)\ln(a - a_0) \tag{6.1}$$

with $a_0 = m_{N_f}/\sqrt{2}$ and c a constant. The monodromy around $a = a_0$ is thus

$$a \to a,$$
$$a_D \to a_D + a - a_0 = a_D + a - \frac{m_{N_f}}{\sqrt{2}}. \tag{6.2}$$

Thus, under monodromy, the pair (a_D, a) are not simply transformed by $SL(2, \mathbb{Z})$; they also pick up additive constants. It was explained in Section 3.1 of Ref. [9] that the duality symmetry of the low energy theory permits such constants to appear; but it

was also shown in Ref. [9], Section 4, that this possibility is not realized for the pure $N = 2$ gauge theory. The above simple consideration of a massless quark shows that this possibility does enter for $N_f > 0$.

If one arranges a_D, a, and the bare mass m as a three dimensional column vector $(m/\sqrt{2}, a_D, a)$, then the monodromy in (6.2) can be written in the general form

$$\mathcal{M} = \begin{pmatrix} 1 & 0 & 0 \\ r & k & l \\ q & n & p \end{pmatrix} \tag{6.3}$$

with $\det \mathcal{M} = kp - nl = 1$. This is the most general form permitted by the low energy analysis of Ref. [9]. The specific form of the first row in (6.3) means that m is monodromy invariant; intuitively this reflects the fact that m is a "constant," not a "field."

Since the central charge in (5.3) must be monodromy-invariant, one can deduce at once how the charges transform. If one arranges the charges as a row vector $W = (S, n_{\mathrm{m}}, n_{\mathrm{e}})$, then W transforms by $W \rightarrow W\mathcal{M}^{-1}$. Explicitly,

$$\mathcal{M}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ lq - pr & p & -l \\ nr - kq & -n & k \end{pmatrix}. \tag{6.4}$$

Thus, the electric and magnetic charges n_{e} and n_{m} mix among themselves but do not get contributions proportional to the global symmetry charge S. On the other hand, the S charge can get contributions proportional to gauge charges n_{e} or n_{m}. Equivalently, the global symmetry can be transformed to a linear combination of itself and a gauge symmetry but not the other way around. Notice that the monodromy matrix mixing the charges in this way survives even if the bare mass m vanishes.

Here we get an elementary example of the situation suggested in Section 4 of Ref. [9] – the spectrum of BPS-saturated one-particle states is not transformed in the expected way under monodromy. As is clear from (5.3), one of the elementary quarks is massless at $a = m_{N_f}/\sqrt{2}$. For large mass this happens at large $|u|$ where semiclassical techniques are reliable. The monodromy around that point shifts the S value of a magnetic monopole by an amount proportional to the magnetic charge; this follows upon using the monodromies in (6.2) to determine the matrices $\mathcal{M}$ and hence $\mathcal{M}^{-1}$. However, in this regime the spectrum of magnetic monopoles can be worked out explicitly using semiclassical methods and in particular the values of S are bounded. (In semiclassical quantization of the monopole, the only zero modes carrying S are the fermion zero modes, and there are only finitely many of them. The boson zero modes, which could carry an arbitrary charge, are S-invariant.) This means that a phenomenon first considered in two dimensions by Cecotti et al. [24] must be operative: as we circle around the singularity, the monopole crosses a point of neutral stability where it can decay to two other states. Then monodromy indeed changes the quantum numbers in the expected fashion but what starts as a one particle state comes back as a multiparticle state; meanwhile, the spectrum of BPS-saturated one-particle states jumps. In this example, since everything is happening in the semiclassical regime, it must be possible to exhibit the jumping very explicitly.

Self-duality for $N_f = 4$?

$N = 4$ super Yang-Mills theory has a spectrum of BPS-saturated states that seems to be invariant under $SL(2, \mathbb{Z})$ acting on (n_m, n_e). This is one of the main pieces of evidence for Olive-Montonen duality in that theory.

For the pure $N = 2$ theory, the spectrum is not $SL(2, \mathbb{Z})$-invariant, though it was seen in Ref. [9] that an $SL(2, \mathbb{Z})$-invariant formalism arises naturally in determining many of its properties. The same situation will prevail, as we will see, for the $N_f = 1, 2, 3$ theories.

But as we will now explain, the $N_f = 4$ theory is a candidate as another theory that may possess manifest $SL(2, \mathbb{Z})$ symmetry in the spectrum, though with the details rather different from the case of $N = 4$.

The states with $(n_m, n_e) = (0, 1)$ are the elementary hypermultiplets, which transform in the vector representation v of Spin(8). The states with $(n_m, n_e) = (1, 0)$ transform as one spinor representation s, and the states with $(n_m, n_e) = (1, 1)$ transform as the other spinor representation c. All of these representations are eight dimensional, and they are permuted by the S_3 group of outer automorphisms of Spin(8). One could think of these states as being permuted by certain $SL(2, \mathbb{Z})$ transformations together with triality. If one is willing to optimistically assume that suitable multi-monopole bound states exist, generalizing the one found in Ref. [18], for every relatively prime pair of integers (p, q), then an $SL(2, \mathbb{Z})$-invariant spectrum is possible. One wants for each such (p, q) to have eight states of $(n_m, n_e) = (p, q)$, transforming according to an eight dimensional representation of Spin(8) that depends on the reduction of (n_m, n_e) modulo 2. In this way, the theory could have an $SL(2, \mathbb{Z})$ symmetry, mixed with Spin(8) triality.

In fact, in the latter part of this paper, when we solve quantitatively for the low energy structure of the $N_f = 4$ theory with arbitrary bare masses, we will get a triality and $SL(2, \mathbb{Z})$-invariant answer, strongly indicating that this possibility is realized.

Clearly, since $SL(2, \mathbb{Z})$ permutes the various Spin(8) representations, it does not commute with Spin(8). In fact, the four classes of Spin(8) representations are permuted under $SL(2, \mathbb{Z})$ like the four spin structures on the torus – o is like the odd spin structure while the other three are like the even ones. The full group is a semidirect product $\text{Spin}(8) \ltimes SL(2, \mathbb{Z})$.

An explicit description of this semidirect product is as follows. The outer automorphism group (triality) of Spin(8) is the group S_3 of permutations of three objects. It permutes the three eight dimensional representations. This group can be regarded as the group of 2×2 matrices of determinant one with entries that are integers mod 2. Therefore (by reduction mod 2) there is a homomorphism from $SL(2, \mathbb{Z}) \to S_3$. The kernel consists of matrices congruent to 1 mod 2. $SL(2, \mathbb{Z})$ acts on Spin(8) by mapping to S_3 which then acts on Spin(8); using this action of $SL(2, \mathbb{Z})$ on Spin(8), one constructs the semidirect product $\text{Spin}(8) \ltimes SL(2, \mathbb{Z})$.

It should be noted that for $N_f \neq 0$ the elementary massive gluons are only neutrally stable against decay to the massive quarks. Apparently, they are analogous to bound states at threshold in nonrelativistic quantum mechanics (which can exist as discrete states). If this is the right interpretation and the theory is indeed dual, there should also be massive Spin(8)-invariant BPS-saturated states of spins $\leqslant 1$ and charges $(n_m, n_e) = (2k, 2l)$ with arbitrary relatively prime k and l.

7. The singularities for $N_f = 1, 2, 3$

In this section, we will begin our study of the singularities of the quantum moduli spaces, using a method that is natural for the asymptotically free theories of $N_f \leqslant 3$. We first consider the case of very large bare masses compared to the dynamical mass scale Λ, where the theory reduces to the $N_f = 0$ theory and the vacuum structure is known from Ref. [9]. Then we extrapolate to small masses. The conformally invariant $N_f = 4$ theory involves somewhat different issues and its structure will be determined later.

We start with the $N_f = 3$ theory with three equal mass quarks $m_r = m \gg \Lambda$. The mass terms break the global $\text{Spin}(6) = \text{SU}(4)$ flavor symmetry to $\text{SU}(3) \times \text{U}(1)$. Classically there is a singularity at $a = m/\sqrt{2}$ where some of the elementary quarks are massless. The massless fields there are electrically charged and they form a triplet of $\text{SU}(3)$. Since we consider the case $m \gg \Lambda$, this singularity is in the semiclassical region $u \approx 2a^2 = m^2 \gg \Lambda^2$ and it persists quantum mechanically. For $u \ll m^2$ the three quarks are massive and can be integrated out semiclassically. The low energy theory is the pure gauge ($N_f = 0$) theory. The scale Λ_0 of the low energy theory can be determined at the one-loop approximation in terms of the masses and the scale of the high energy theory Λ_3 to be $\Lambda_0^4 = m^3 \Lambda_3$. Therefore, the moduli space at small u is given approximately by that of the pure gauge theory with scale Λ_0. It has two singular points where, respectively, monopoles of $(n_{\mathrm{m}}, n_{\mathrm{e}}) = (1, 0)$ and $(n_{\mathrm{m}}, n_{\mathrm{e}}) = (1, 1)$ are massless. These two monopoles are $\text{SU}(3)$ invariant.

As the mass m of the quarks is reduced, the singular point at large u moves toward the origin and the location of the two other singular points can change. As we discussed in the previous section, the values of $(n_{\mathrm{m}}, n_{\mathrm{e}})$ and the charges under the abelian symmetries of the massless particles at the singularities can change. However, their non-abelian global charges cannot change. The states massless at the various singularities transform, respectively, as $\mathbf{3}$, $\mathbf{1}$, and $\mathbf{1}$ of the global $\text{SU}(3)$ symmetry. For $m = 0$ the global symmetry is enlarged from $\text{SU}(3) \times \text{U}(1)$ to $\text{SU}(4)$, Therefore, the massless particles at the different singularities must be in representations of $\text{SU}(4)$. The only way for this to happen is that two of the singularities with a massless $\mathbf{3}$ and $\mathbf{1}$ of $\text{SU}(3)$ must combine into a singularity with a massless $\mathbf{4}$ of $\text{SU}(4)$ while the third singularity goes elsewhere [3]. Therefore, the $N_f = 3$, $m = 0$ theory has precisely two singularities in the u plane, with massless particles that are respectively a $\mathbf{4}$ and $\mathbf{1}$ of $\text{SU}(4)$.

The $\text{SU}(4)$ quantum numbers of the particles at the singularities can be used to constrain their electric and magnetic charges. As explained in Section 5, the smallest choice of $(n_{\mathrm{m}}, n_{\mathrm{e}})$ for a state in the $\mathbf{4}$ of $\text{SU}(4)$ is $(1, 0)$ and for an $\text{SU}(4)$ singlet it is $(2, 1)$. As we will discuss at the end of Section 14, it is possible to show that if our picture is correct, then the states that become massless at the singularities are

---

[3] The three singularities cannot all combine together, since the fundamental group of the once-punctured u plane is abelian, and an abelian representation of the fundamental group, when combined with the known behavior at infinity, will lead (as we saw in Section 5.2 of Ref. [9]) to an indefinite metric on the quantum moduli space. Nor can any of the singularities go to infinity without changing the coefficient of the logarithm $a_D \sim a \ln a$; this coefficient cannot be changed as it is determined by the one-loop beta function.

continuously connected to BPS-saturated states with the same global quantum numbers that exist in the semiclassical region of large u. The $(1,0)$ in the **4** of SU(4) certainly exists semiclassically. It is not obvious whether an SU(4) singlet bound state of two monopoles exists semiclassically (a somewhat similar state was found in Ref. [18] for $N = 4$), but we conjecture that such a state must exist.

The same procedure can be used to determine the singularities of the massless $N_f = 1, 2$ theories. One starts with large equal masses for the quarks and analyzes the singularities at large u. As the masses get smaller the singularities move toward the origin. Finally, the global symmetries determine the full structure.

An alternate procedure, which we will use here, is to follow the singularities of the massless $N_f = 3$ theory as some quarks become heavy. First we give a mass m_3 to only one of the hypermultiplets. It breaks SO(6) $\rightarrow$ SO(2) $\times$ SO(4) (or equivalently SU(4) $\rightarrow$ SU(2) $\times$ SU(2) $\times$ U(1)). Of the six fermion zero modes, two for each hypermultiplet, two are lifted, say η_1 and η_2, giving a perturbation $im_3\eta_1\eta_2/2$ to the monopole Hamiltonian. As the η_i act as gamma matrices upon quantization, the perturbation has eigenvalues $\pm m_3/2$, with equal multiplicities; it can be diagonalized to give

$$\frac{1}{2}\begin{pmatrix} m_3 & & & \\ & m_3 & & \\ & & -m_3 & \\ & & & -m_3. \end{pmatrix} \tag{7.1}$$

This breaks SU(4) $\rightarrow$ SU(2) $\times$ SU(2) $\times$ U(1) which is of course the right pattern. The four monopoles are split to two pairs. One pair transforms as $(\mathbf{2}, \mathbf{1}, \frac{1}{2})$ and the other as $(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$ under the unbroken SU(2) $\times$ SU(2) $\times$ U(1).

Given the U(1) charges, the mass formula (5.3) for BPS-saturated states shows that one pair becomes massless at $a_D + m_3/2\sqrt{2} = 0$, and one at $a_D - m_3/2\sqrt{2} = 0$. Since, in the $m_3 = 0$ limit, a_D is a good local coordinate near its zero, there is one nearby point obeying the first of these equations and one nearby point obeying the second.

Therefore, for small non-zero m_3 there are three singularities. At two of them there are two massless particles transforming as one or the other spinor of SO(4), while at the third singularity there is a massless SO(4) singlet state.

Now we increase m_3. For large m_3, we expect one singularity at large u where the elementary quark becomes massless. This state is an SO(4) singlet, so we identify the singularity it generates with the continuation to large m_3 of the singularity of the small m_3 theory that is generated by a massless singlet.

We can now integrate out the heavy quark, eliminating the singularity just described (since it goes to infinity for $m_3 \rightarrow \infty$) and leaving the other singularities. The low energy theory is the massless $N_f = 2$ theory. Its scale Λ_2 is determined by a one-loop calculation to be $\Lambda_2^2 = m_3\Lambda_3$. As we take m_3 to infinity holding Λ_2 fixed we are left with the two other singularities of the $(m_1 = m_2 = 0, m_3 \neq 0)$ $N_f = 3$ theory. Each of those singularities has two massless states in one or the other spinor representation of SO(4). Given these SO(4) quantum numbers and assuming that these states are continuously connected to states that exist semiclassically for large u, the minimal choices for their electric and magnetic charges are $(n_m, n_e) = (1, 0)$ and $(n_m, n_e) = (1, 1)$.

As we explained above, the $\mathbb{Z}_2$ symmetry which changes the sign of u does not commute with SO(4). It exchanges the two spinors. Therefore, the two singularities at finite u are related by the $\mathbb{Z}_2$ symmetry. Their different electric charges can also be determined by rotating u continuously by π. This amounts to changing θ by π, an operation which adds one unit of electric charge to the monopole; that is why they have equal n_m and have n_e differing by 1.

We now repeat this analysis for the flow from $N_f = 2$ to $N_f = 1$. Consider one of the pairs of massless monopoles (in $(\mathbf{2}, \mathbf{1})$ of SO(4) $\cong$ SU(2) × SU(2)). A mass term for the second quark m_2 breaks SO(4) → SO(2) × SO(2). For small m_2 the mass term in the monopole theory is

$$\frac{1}{2} \begin{pmatrix} m_2 & \\ & -m_2 \end{pmatrix} \tag{7.2}$$

and it splits the pair. Similarly, the other pair in $(\mathbf{1}, \mathbf{2})$ of SO(4) is also split and the massive theory has four singularities. As m_2 becomes large, precisely one of these singularities moves to large u (since the quark of large m_2 has a component that is massless at one value $u \approx m_2{}^2$) where a semiclassical analysis is reliable. Three singularities are left behind. Now we integrate out the massive quark. The low energy theory is the massless $N_f = 1$ theory with scale $\Lambda_1^3 = m_2 \Lambda_2^2$. This theory has the three singularities that do not go to infinity as $m_2 \to \infty$. They are related by the discrete $\mathbb{Z}_3$ global symmetry of the $N_f = 1$ theory. As we move from one singularity to the other, the monopole acquires one unit of electric charge. Therefore, the values of (n_m, n_e) for the massless states at the singularities are $(1, 0)$, $(1, 1)$ and $(1, 2)$.

Similarly, we can follow the monopoles from $N_f = 1$ to $N_f = 0$. One of the monopole points moves to infinity and the other two remain, giving the structure proposed in Ref. [9].

It is curious to note that for small quark mass the massless fields at the singularity are magnetic monopoles whereas for large mass some of them are electric charges – the elementary quarks. This continuous transformation from an elementary particle to a magnetic monopole is possible because of the nonabelian monodromies. The values of (n_m, n_e) of a massless particle at a singularity are determined by the monodromy around the singularity. This monodromy depends on a choice of base point and a path around the singularity. When these choices are changed, the monodromy is conjugated. As the mass changes the singularities move on the moduli space and correspondingly, the natural choice for the path along which the monodromy is computed is changed. Therefore, although the conjugacy class of the monodromy cannot change, the natural labeling by quantum numbers (n_m, n_e) can change. We are used to this phenomenon in the case of a magnetic monopole acquiring electric charge [23] by changing $\theta \to \theta + 2\pi$. Here, because of the other singularities, also the magnetic charge can change.

To summarize this discussion, for the theories with zero bare masses one has the following:

For $N_f = 0$ the global symmetry acting on the u plane is $\mathbb{Z}_2$. There are two singularities related by this symmetry with massless states $(n_m, n_e) = (1, 0)$ and $(n_m, n_e) = (1, 1)$.

For $N_f = 1$ the global symmetry of the u plane is $\mathbb{Z}_3$. There are three singularities related by this symmetry with massless states $(n_m, n_e) = (1,0)$, $(n_m, n_e) = (1,1)$ and $(n_m, n_e) = (1,2)$.

For $N_f = 2$ the global symmetry of the u plane is $\mathbb{Z}_2$. There are two singularities related by this symmetry. The massless states at one singularity have $(n_m, n_e) = (1,0)$ and are in one spinor of SO(4) while the massless states at the other singularity have $(n_m, n_e) = (1,1)$ and are in the other spinor of SO(4).

For $N_f = 3$ the u plane has no global symmetry. There are two singularities. In one of them there are four massless states with $(n_m, n_e) = (1,0)$ in a spinor representation of SO(6) with definite chirality; in the other there is a single state with $(n_m, n_e) = (2,1)$.

Although the arguments we suggested for the singularity structure are quite plausible, they are certainly not a rigorous proof. In the next sections we will study the consequences of this picture and supply what we regard as convincing evidence.

8. Low energy theory near the singularities

In this section we study the low energy effective field theory near the singularities. For simplicity, we will focus on the massless theories. It is straightforward to extend our considerations to the massive ones.

The light states near all the singularities are a photon multiplet and some charged fields. Using the duality transformation, the low energy theory near any of the singularities is an abelian gauge theory with some light hypermultiplets. This theory was studied in Section 2, but two important points should be noted. First, the $N = 2$ photon multiplet we encounter in our low energy effective theories is not the semiclassical photon. It is related to the semiclassical photon by an appropriate duality transformation. Similarly, the scalar a in Section 2 should be identified with a linear combination of a and a_D of the non-abelian theory. The second important difference is that the low energy theory also contains higher dimension operators. These include terms of the form $\left(A/\Lambda\right)^n W_\alpha^2$ in the gauge coupling. These terms break the U(1)$_\mathcal{R}$ symmetry of the abelian theory.

The flavor symmetry of our original massless theory is SO($2N_f$) and the light states at the singularities are in a spinor of SO($2N_f$) (except for one of the singularities of the $N_f = 3$ theory at which the massless multiplet is a flavor singlet). At first look, this might seem in contradiction with the SU(k) symmetry of the effective low energy theory under which the k light fields transform as the fundamental representation of SU(k). Fortunately, special properties of the relevant SO groups make this consistent. For $N_f = 2$ the two light states transform as $(\mathbf{2}, \mathbf{1})$ or $(\mathbf{1}, \mathbf{2})$ under SO(4) $\cong$ SU(2) × SU(2) and, either way, only a single SU(2) acts on the light fields and the two hypermultiplets transform as a doublet. Similarly, for $N_f = 3$ the four states at the singularity are in a spinor of SO(6) $\cong$ SU(4) and transform as the $\mathbf{4}$ of SU(4).

Having established that the symmetries of the low energy theory act correctly, we will now study the flat directions of this theory. They are the quantum moduli space of the original theory and should be connected smoothly to the semiclassical picture.

The low energy theories in all the singularities have flat directions of the a field along which all the hypermultiplets acquire a mass. As such they are smoothly connected with the Coulomb branch of the original non-abelian theory.

For $N_f = 1$ there is a single light hypermultiplet at every singularity and therefore there are no other flat directions. This is consistent with the absence of Higgs branches in the moduli space of the original theory.

For $N_f = 2$ the low energy theory at the singularity is QED with $k = 2$ hypermultiplets. As discussed in Section 2, this theory has M flat directions along which $SU(2) \times SU(2)_R$ is broken to $SU(2)_{R'}$ and the light fields transform as $\mathbf{3} \oplus \mathbf{1}$ of the unbroken symmetry (see Eqs. (2.3) and (2.4)). Adding to this symmetry the other $SU(2)$ global symmetry which does not act on the light fields, we conclude that along these M flat directions the symmetry breaking pattern is $SO(4) \times SU(2)_R \rightarrow SU(2) \times SU(2)_{R'}$ and the light fields transform like $(\mathbf{1}, \mathbf{3}) \oplus (\mathbf{1}, \mathbf{1})$ of the unbroken symmetry. This is exactly the pattern of symmetry breaking and light spectrum observed on the Higgs branch of the original theory in (3.5) and (3.6).

Classically, the $N_f = 2$ theory had two Higgs branches touching the Coulomb branch at the origin. Quantum mechanically, the two branches touch the Coulomb branch at different points but the metric on them and the pattern of symmetry breaking are the same as they are classically.

For $N_f = 3$ there are two singularities. In one of them $k = 1$; there are no M flat directions emanating from that point. In the other one there are $k = 4$ hypermultiplets in a spinor of $SO(6)$. Along the M flat directions $SU(4) \times SU(2)_R$ is broken to $SU(2) \times U(1) \times SU(2)_{R'}$ and the light fields transform as $(\mathbf{2}, 1, \mathbf{2}) \oplus (\mathbf{2}, -1, \mathbf{2}) \oplus (\mathbf{1}, 0, \mathbf{3}) \oplus (\mathbf{1}, 0, \mathbf{1})$ (see Eqs. (2.3) and (2.4)). Again, this is precisely the pattern of symmetry breaking and light spectrum observed in (3.5) and (3.6) on the Higgs branch of the original theory.

Note how the Higgs branch of the quantum moduli space has two weakly coupled limits. In one of them the weakly coupled particles are magnetically charged in a spinor of $SO(2N_f)$ and in the other limit they are quarks which are doublets of the $SU(2)$ gauge group and are components of a vector of $SO(2N_f)$ (which is broken to a subgroup). The original gauge symmetry looks like it is confined at one end because magnetic monopoles condense there. At the other end it looks like it is completely broken by the Higgs mechanism. Since our theory includes matter fields in the fundamental representation, there is no strict gauge invariant distinction between confinement and complete gauge symmetry breaking [14] and therefore there is no contradiction here. The gauge invariant order parameters – $\tilde{M}M$ at one end and $V^{rs} = Q^r Q^s$ at the other end – transform the same way under the global symmetry ($(\mathbf{3}, \mathbf{1})$ under $SO(4) \cong SU(2) \times SU(2)$ for $N_f = 2$ and $\mathbf{15}$ of $SO(6) \cong SU(4)$ for $N_f = 3$) and hence lead to the same pattern of symmetry breaking and to the same massless spectrum.

9. Breaking $N = 2$ to $N = 1$

In this section we break $N = 2$ supersymmetry to $N = 1$ by adding a mass term $m \operatorname{Tr} \Phi^2$ to the tree level superpotential (3.1). When $m \gg \Lambda$, the $N = 1$ chiral multiplet

in the adjoint representation Φ is heavy and can be integrated out. The resulting theory is $N = 1$ SUSY with gauge group SU(2) and $2N_f$ chiral doublets Q^r with $r = 1, \ldots, 2N_f$. An interesting term at tree level is a quartic term $(1/m)(Q^r Q^s)^2$ in the superpotential which breaks the global SU($2N_f$) symmetry of the $N = 1$ theory to SO($2N_f$). At one loop, the scale $\widetilde{\Lambda}_{N_f}$ of the $N = 1$ theory is given by $\widetilde{\Lambda}_{N_f}^{6-N_f} = m^2 \Lambda_{N_f}^{4-N_f}$. As $m \to \infty$ with $\widetilde{\Lambda}_{N_f}$ held fixed the quartic term in the superpotential is negligible and we should recover the known results of the $N = 1$ theory [4].

For small m we can use the low energy effective theory. The mass term is represented as a term mU in the superpotential. Since it has no critical points as a function of U, the only reason that there are any supersymmetric ground states at all is that new degrees of freedom become light and have to be included near the singularities. Near the singularities, one can use an effective Lagrangian like that of subsection 2.5 and approximate $U \approx u_0 \Lambda_{N_f}^2 + u_1 \Lambda_{N_f} A + \mathcal{O}(A^2)$ where u_0 and u_1 are dimensionless constants and A is the chiral superpartner of the light photon. This is exactly the Lagrangian we studied in subsection 2.5 with $\mu = mu_1 \Lambda_{N_f}$. As we saw there, the value of a is fixed at zero and therefore $u = u_0 \Lambda_{N_f}^2$. The matter fields M and $\widetilde{M}$ acquire expectation values breaking the U(1) gauge symmetry. Since these are magnetic monopoles, this means confinement of the original charges.

We see that the continuum of vacua on the Coulomb branch has disappeared and the surviving ground states are at the singularities. Every singularity leads to a vacuum.

Next we should identify what happens to the Higgs branches. We continue to use the effective Lagrangian of subsection 2.5. For $N_f = 2$ we have two regions described by QED with $k = 2$ where generically SO(4) is broken to SU(2) (see Eq. (2.11)) and at a special point it is broken to SU(2) × U(1) (see Eq. (2.12)). For $N_f = 3$ there is an isolated ground state (related to the condensation of the $(n_m, n_e) = (2, 1)$ monopoles) as well as a continuum. At the generic point in the continuum SO(6) is broken to SU(2) × U(1) (see Eq. (2.11)) and at a special point it is broken to SU(3) × U(1) (see Eq. (2.12)).

This low energy effective Lagrangian is a good description of the physics for small m. However, it might not be appropriate in the limit $m \to \infty$. There are two reasons for that. First, in that limit new states which are massive for any finite m can become massless and should be included in the Lagrangian. Second, as $m \to \infty$ we have to take Λ_{N_f} to zero in order to keep the low energy scale $\widetilde{\Lambda}_{N_f}$ fixed. This means that the different ground states on the Coulomb branch approach each other. The appropriate effective Lagrangian should describe all of them.

The degrees of freedom that we expect are those of the $N = 1$ theory that is obtained by integrating out ϕ. This theory can be usefully described [4] by an effective theory for the gauge invariant composite field $V^{[rs]} = Q^r Q^s$. It would be equivalent to use the composite monopole fields $Y_a^b = \widetilde{M}_a M^b$ with $a, b = 1, \ldots, k$. For $N_f = 3$ the fields V are in the **15** of SO(6) and so are the fields Y if Tr Y is removed. For $N_f = 2$ the fields V are in $(\mathbf{3}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{3})$ of SO(4). In terms of the monopole fields, these representations are obtained by considering the monopole bilinears Y in the two branches and removing their traces.

The effective superpotential for V can be constrained along the lines of Ref. [3]. We

can require it to respect all the symmetries of the theory – including those explicitly broken by m or the anomaly – if we assign appropriate transformation laws under such symmetries to m and Λ. We also demand that it be locally holomorphic in V, m and Λ_{N_f}, and that it has a finite limit as $m \to \infty$ with $\widetilde{\Lambda}_{N_f}$ held fixed which coincides with that of the $N = 1$ theory [4].

For $N_f = 2$ these considerations determine the superpotential

$$W = X(\mathrm{Pf}\ V - m^2 \Lambda_2^2) + \frac{1}{m} V^2, \tag{9.1}$$

where X is a Lagrange multiplier. For finite m it leads to two branches: $X = \pm 1/m$ with $V^{rs} = \pm \frac{1}{2} \epsilon^{rstu} V^{tu}$ and Pf $V = m^2 \Lambda^2$. These are the two Higgs branches we found before. As $m \to \infty$ more fields become massless, and we recover the full moduli space of the $N = 1$ theory which is constrained by Pf $V = m^2 \Lambda_2^2 = \widetilde{\Lambda}_2^4$ [4].

For $N_f = 3$ the superpotential is

$$-\frac{1}{m^2 \Lambda_3} \mathrm{Pf}\ V + \frac{1}{m} V^2. \tag{9.2}$$

For finite m the equation of motion of V has two types of solution. There is a continuum of states which we associate with the Higgs branch. There is also an isolated state with unbroken SU(4) at $V = 0$; we interpret this as the vacuum with the condensation of the $(n_{\mathrm{m}}, n_{\mathrm{e}}) = (2, 1)$ monopole. As $m \to \infty$, the isolated state merges into the continuum to form the moduli space of the $N = 1$ theory.

Note the following crucial point. For $N_f = 2$ the ground states of the massive theory are on the quantum moduli space of the corresponding $N = 1$ theory Pf $V = m^2 \Lambda_2^2$. This is not the case for $N_f = 3$. The states can be described by the order parameter V of the $N = 1$ theory but they occur for values of V that do not obey the equations of motion of the $N = 1$ theory (namely $\epsilon_{r_1 \ldots r_6} V^{r_1 r_2} V^{r_3 r_4} = 0$). The same phenomenon happens when other perturbations of the massless $N = 1$, $N_f = 3$ theory (like adding mass terms for the quarks [4] or gauging a subgroup of the global symmetry [5]) are considered. It arises because all the components of V are massless in the $N = 1$ theory at $V = 0$. Therefore, all of them should be kept in the effective Lagrangian.

In sum, we have found two different low energy effective Lagrangians for the theory broken to $N = 1$. One of them includes a photon and some monopole fields. The other includes only the fields V. For finite non-zero m they lead to the same physics for the massless modes and differ in the way they describe the massive fields. A low energy effective Lagrangian with a finite number of terms cannot be expected to describe massive fields correctly. At best it can give an approximate description of the light fields. The monopole Lagrangian has a smooth $m \to 0$ limit because it includes the fields which become massless in this limit. On the other hand, the Lagrangian with V has a smooth $m \to \infty$ limit because it includes the fields which become massless in that limit.

This picture also explains the phenomenon observed in Ref. [4] in the $N = 1$, $N_f = 3$ theory where at the origin of field space confinement (to the extent that it is well defined in a theory with matter fields in the fundamental representation) occurred without chiral

symmetry breaking. This is due to the condensation of monopoles which do not carry global quantum numbers, notably the $(n_m, n_e) = (2, 1)$ monopole.

We see here a new phenomenon in quantum field theory. Magnetic monopoles acquire global charges because of the existence of fermion zero modes. When these monopoles condense, they lead to chiral symmetry breaking. This leads us to suggest that to the extent that condensation of monopoles can be used to describe confinement in QCD, it can also be used to describe chiral symmetry breaking.

10. The singularities for $N_f = 4$

When the number of flavors is four the one-loop beta function vanishes. Because of the properties of $N = 2$, the beta function is also zero to all orders in perturbation theory. Does the exact beta function vanish? The non-perturbative contributions to the beta function can be studied by examining the low energy effective coupling τ. As in Ref. [8] we can examine $\partial^2 \tau(a)/\partial a^2$ by computing a matrix element of four fermions. However, as we explained above, because of the parity symmetry (3.2) the one instanton contribution to this matrix element vanishes when the number of flavors is non-zero. For $N_f = 0, 1, 2, 3$ the two instanton contribution to τ is non-zero. If it is also non-zero for $N_f = 4$, it is logarithmic in a, so that

$$\frac{\partial^2 \tau(a)}{\partial a^2} \sim e^{-(16\pi^2/g^2) + 2i\theta}/a^2 = e^{2\pi i \tau_{cl}}/a^2.$$

Including also multiple instanton contributions, $\tau(a)$ can have the form $c(e^{2\pi i \tau_{cl}}) \ln(a/\Lambda)$ leading to a term $\int d^2\theta\, c(e^{2\pi i \tau_{cl}}) \ln(\Lambda/\Lambda) W_\alpha^2$ in the low energy effective Lagrangian. The appearance of the scale Λ in this theory signals a non-perturbative anomaly both in conformal invariance and in $U(1)_{\mathcal{R}}$ (the anomaly in $U(1)_{\mathcal{R}}$ can be seen by performing a $U(1)_{\mathcal{R}}$ transformation on the low energy effective action). When the theory is put on manifolds with non-zero second Betti number, instantons in the low energy abelian theory lead to explicit exponentially small breaking of $U(1)_{\mathcal{R}}$ as a result of which all amplitudes in certain topological sectors would vanish. This seems bizarre. Another consequence of the logarithm in τ is that the metric on the moduli space, namely $\text{Im}\, \tau\, da\, d\bar{a}$, is not positive definite. Although it could perhaps be modified by other non-perturbative effects, we find this unlikely. So we will assume that $c = 0$ and the exact quantum theory is scale invariant. In any event, the results that we will obtain add considerably to the plausibility of our assumption of exact scale invariance.

Such a scale invariant theory is characterized by the classical dimensionless coupling constant $\tau = (\theta/\pi) + (8\pi i/g^2)$ because there is no dimensional transmutation. Its quantum moduli space is by scale invariance the same as the classical space, and the absence of corrections to τ implies that there are no corrections to

$$a = \tfrac{1}{2}\sqrt{2u},$$
$$a_D = \tau a. \tag{10.1}$$

Clearly, the only monodromy is the one around the origin, which is $P = -1$.

The situation is more interesting when some masses m_i are not zero. Then, both scale invariance and $U(1)_{\mathcal{R}}$ are explicitly broken by the masses. As some of the masses go to infinity (with a suitable limit of τ) we should recover the quantum moduli space of the asymptotically free theories with $N_f = 0, 1, 2$ or 3.

More explicitly, if some masses m_i for $i = n + 1, \ldots, 4$ are taken to infinity, the right scaling limit is obtained by taking $\tau \to i\infty$ holding fixed

$$\Lambda_n^{4-n} \sim e^{\pi i \tau} \prod_i m_i = q^{1/2} \prod_i m_i \tag{10.2}$$

and u. Then, the low energy theory has $N_f = n$ flavors and scale parameter Λ_n. Various definitions of this scale (e.g. with other subtraction schemes) differ by a multiplicative constant of order one.

For example, if only one mass m_4 is not zero, the scaling limit should be that of the $N_f = 3$ theory. For weak coupling, its moduli space has two singularities at $u \ll m_4$ (where the theory flows to strong coupling) with four massless particles in one of them and a single massless particle in the other. Since for large u we recover the fourth flavor, there is another singularity where the fourth quark is massless.

More generally, if we weight each singularity by the number of massless hypermultiplets at that point, then the total weighted number of singularities in the complex u plane is always six.[4] If the masses are generic, there are six singularities each of weight one. If some masses are degenerate, some of the singularities can be combined to a smaller number of singularities of higher weight. Denote by k_l the weight of the lth singularity, so $\sum k_l = 6$. The value of the k_l at the singularities are constrained by the symmetries of the massive theory. We mention a few examples:

(1) $m_i = (m, 0, 0, 0)$ with global symmetry $SU(4) \times U(1)$. This is the case discussed above with three singularities and $k_l = (4, 1, 1)$. The four massless particles in the first singularity transform according to the fundamental representation of $SU(4)$.

(2) $m_i = (m, m, m, m)$ with global symmetry $SU(4) \times U(1)$. There are three singularities with $k_l = (1, 1, 4)$ where again the massless particles in the last singularity transform according to the fundamental representation of $SU(4)$. Note the similarity between this example and the previous one.

(3) $m_i = (m, m, 0, 0)$ with global symmetry $SU(2) \times SU(2) \times SU(2) \times U(1)$. There are three singularities with $k_l = (2, 2, 2)$ where the massless particles in every singularity are in a doublet of one of the $SU(2)$ factors.

(4) $m_i = (m + \mu, m - \mu, 0, 0)$ with non-zero $\mu \neq m$. The global symmetry is $SU(2) \times SU(2) \times U(1) \times U(1)$. For non-zero μ one of the singularities in case (3) splits and there are four singularities with $k_l = (1, 1, 2, 2)$ where the massless particles in the singularities with $k = 2$ are in a doublet of one of the $SU(2)$ factors. As μ varies between 0 and m, this example interpolates between examples (3) and (1).

(5) $m_i = (m, m, \mu, \mu)$ with non-zero $\mu \neq m$. The global symmetry is $SU(2) \times SU(2) \times U(1) \times U(1)$. For non-zero μ one of the singularities in case (3) splits and there are

---

[4] This will ultimately follow from the fact that the low energy physics is described by a curve $y^2 = F(x, u)$ where the discriminant of $F(x, u)$ with respect to x is a sixth order polynomial in u; its zeroes are the singularities.

four singularities with $k_l = (1, 1, 2, 2)$ where the massless particles in the singularities with $k = 2$ are in a doublet of one of the SU(2) factors. As μ varies between 0 and m, this example interpolates between examples (3) and (2). Note the similarity between this example and the previous one.

(6) $m_i = (m, m, m, 0)$ with global symmetry SU(3) × U(1) × U(1). There are four singularities with $k_l = (1, 1, 1, 3)$ where the massless particles in the singularity with $k = 3$ transform according to the fundamental representation of SU(3).

As $m \to \infty$ with an appropriate shift of u to remove the last of these singularities, we should recover the quantum moduli space of theories with fewer flavors.

The similarities between examples (4) and (5) (and between their special cases (1) and (2)) can be used as evidence for the triality of the theory. The S_3 automorphism of Spin(8) is generated by two transformations. They act on the masses (which are in the adjoint representation of Spin(8)) as

$$
\begin{aligned}
m_1 &\to m_1 , \\
m_2 &\to m_2 , \\
m_3 &\to m_3 , \\
m_4 &\to -m_4
\end{aligned}
\tag{10.3}
$$

which exchanges the two spinors keeping the vector fixed and

$$
\begin{aligned}
m_1 &\to \tfrac{1}{2}(m_1 + m_2 + m_3 + m_4) , \\
m_2 &\to \tfrac{1}{2}(m_1 + m_2 - m_3 - m_4) , \\
m_3 &\to \tfrac{1}{2}(m_1 - m_2 + m_3 - m_4) , \\
m_4 &\to \tfrac{1}{2}(m_1 - m_2 - m_3 + m_4)
\end{aligned}
\tag{10.4}
$$

which exchanges the vector with one of the spinors keeping the second spinor fixed. The transformation in (10.4) is particularly interesting. Since it exchanges electrons and monopoles, to map the theory to itself it must also act on τ. However, the action on τ should not affect the qualitative structure of the moduli space: the number of singularities and the values of the k_l up to permutation are independent of τ at least for generic τ. Returning to our examples, it is easy to check that the transformation (10.4) exchanges $m_i = (m, m, \mu, \mu) \leftrightarrow m_i = (m + \mu, m - \mu, 0, 0)$ and therefore triality would relate their moduli spaces and explain the similarity between them.

11. Preliminaries for determining the metric

In the remainder of this paper we will find the exact solution for the low energy effective action, metric on moduli space, and particle masses for the various theories. The basic idea, as in Ref. [9], involves introducing a suitable family of elliptic curves, and interpreting (a_D, a) as periods of an appropriate family of meromorphic one-forms.

11.1. $N_f = 0$

It is appropriate to start first by reconsidering the $N_f = 0$ theory. In Section 6 of Ref. [9], we described this theory by the family of curves

$$y^2 = (x - \Lambda^2)(x + \Lambda^2)(x - u). \tag{11.1}$$

(The renormalization scale Λ was set to 1 in that reference.) This is the modular curve for the group $\Gamma(2)$ consisting of integer-valued matrices

$$M = \begin{pmatrix} m & n \\ p & q \end{pmatrix} \tag{11.2}$$

with $\det M = 1$ and with n and p even.

However, if we wish to compare the $N_f = 0$ theory to theories with $N_f > 0$, it is natural to make a change of conventions that was mentioned at the end of the introduction. The magnetic and electric charges (n_m, n_e) were normalized in Ref. [9] so that they were arbitrary integers in the pure $N = 2$ (or $N = 4$) theory. However, for $N_f > 0$, there are particles with half-integral electric charge in those units. Therefore, for $N_f > 0$, it is convenient to multiply n_e by 2 – to restore its integrality – and divide a by 2 – to preserve the structure $Z = a_D n_m + a n_e$. In fact, it is all but necessary to make this change of conventions if one wishes to exhibit all the $SL(2, \mathbb{Z})$ symmetry for $N_f = 4$, while the original conventions are natural for exhibiting the duality symmetry for $N = 4$. The reason is that the $N = 4$ theory has an $SL(2, \mathbb{Z})$ duality symmetry acting on (a_D, a), while the $N_f = 4$ theory turns out to have an $SL(2, \mathbb{Z})$ symmetry acting on (a_D, a'), where $a' = a/2$.

Dividing a by 2 has the effect of conjugating the monodromy matrix M of (11.2) by

$$W = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \tag{11.3}$$

replacing M by

$$W^{-1}MW = \begin{pmatrix} m & 2n \\ p/2 & q \end{pmatrix}. \tag{11.4}$$

The matrices of this form make up the group $\Gamma_0(4)$ consisting of unimodular integer-valued matrices with the upper right entry divisible by four.

With the new conventions, the monodromy at infinity is

$$\begin{pmatrix} -1 & 4 \\ 0 & -1 \end{pmatrix} \tag{11.5}$$

(that is, under monodromy, a magnetic monopole picks up electric charge 4 instead of 2 since we are using a smaller unit of charge), and the monodromy due to a massless monopole is

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}. \tag{11.6}$$

These matrices generate $\Gamma_0(4)$, and in the new conventions, the $N_f = 0$ theory is described by the modular curve of $\Gamma_0(4)$. This is the family of curves

$$y^2 = x^3 - ux^2 + \frac{1}{4}\Lambda^4 x, \tag{11.7}$$

as we will explain presently. Like (11.1), (11.7) has a $\mathbb{Z}_4$ symmetry generated by $y \to iy$, $x \to -x$, $u \to -u$; only a $\mathbb{Z}_2$ quotient acts on the u plane. The two families of curves are related by what is called an "isogeny"; for fixed u the curve described by (11.7) is a double cover of the curve described by (11.1), and vice-versa.

Is it more natural to describe the pure $N = 2$ theory using the conventions that lead to (11.1) or the conventions that lead to (11.7)? That depends on the context. If one is viewing this theory as the low energy limit of the $N = 4$ theory perturbed by an $N = 2$ invariant mass term, then (11.1) is more natural; if one is viewing the same theory as the low energy limit of a theory with $N_f = 4$ (and bare masses for the hypermultiplets) then (11.7) is more natural.

11.2. Some convenient facts

To prove that (11.7) is the modular curve of $\Gamma_0(4)$ is a fairly simple exercise using the addition law on a cubic curve. [5] For our purposes, it will be more helpful to simply work out the singularities and the monodromies of the family.

A general cubic curve of the form

$$y^2 = F(x) = x^3 + \alpha x^2 + \beta x + \gamma = (x - e_1)(x - e_2)(x - e_3) \tag{11.8}$$

describes a double cover of the x plane branched over e_1, e_2, e_3, and ∞. This curve becomes singular when two of the branch points coincide – thus for $e_i = e_j$, or $e_i \to \infty$. For instance, for the family (11.7) the branch points are 0, $\frac{1}{2}(u \pm \sqrt{u^2 - \Lambda^4})$, and ∞, and singularities occur precisely for $u = \pm\Lambda^2$ and $u \to \infty$.

The singularities we will meet on the finite u plane will always be singularities at which precisely two branch points coincide. For $u \to \infty$, we will often have a more complicated configuration. Mathematically, a singularity at which precisely two branch points coincide is said to be "stable"; any singularity can be put in this form by reparametrizing x and y by u-dependent factors, as we will frequently do to understand the behavior for $u \to \infty$. A family of curves in which precisely two branch points are coinciding at, say, $u = 0$, always looks locally near the singularity like the family

$$y^2 = (x - 1)(x^2 - u^n) \tag{11.9}$$

---

[5] $\Gamma_0(4)$ parametrizes elliptic curves with a cyclic subgroup of order four. Such a subgroup is generated by the points $P_\pm$ with coordinates $x = \Lambda^2/2$, $y = \pm\frac{1}{2}\Lambda^2\sqrt{\Lambda^2 - u}$; this follows from the fact that the tangent line to P_+ or P_- passes through the point of order two with coordinates $x = y = 0$. (The fact that there is no natural way to pick a sign of $\sqrt{\Lambda^2 - u}$ means that only a subgroup of order 4, not a point of order 4, is naturally determined; this is appropriate for $\Gamma_0(4)$.) Conversely, given an elliptic curve with a cyclic subgroup T of order 4, one can pick coordinates so that the generators of T have $x = \Lambda^2/2$ and the point of order two has $x = y = 0$; then the curve takes the form of (11.7) with a uniquely determined u.

for some integer n. The monodromy is then conjugate to T^n where as usual

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{11.10}$$

This can be proved, for instance, by computing the periods of the holomorphic differential $\omega = dx/y$, that is the functions

$$\omega_1 = \int_{-u^{n/2}}^{u^{n/2}} \frac{dx}{y}$$

$$\omega_2 = \int_{u^{n/2}}^{1} \frac{dx}{y}, \tag{11.11}$$

near $u = 0$.

An important role in our analysis will be played by the discriminant, Δ, of a polynomial F. It is defined as

$$\Delta = \prod_{i<j} (e_i - e_j)^2 \tag{11.12}$$

with e_i the roots of the polynomial. Since Δ is symmetric under permutations of the roots, it can be expressed as a polynomial in the coefficients of F. For instance, for the cubic polynomial

$$F = x^3 + Bx^2 + Cx + D \tag{11.13}$$

the discriminant is

$$\Delta = -27D^2 + 18BCD + B^2C^2 - 4B^3D - 4C^3. \tag{11.14}$$

Obviously, the branch points of $y^2 = F(x, u)$ coincide precisely when $\Delta = 0$, so singularities (apart from $u = \infty$) are at zeroes of Δ. We can be somewhat more precise. Notice that the curve in (11.9) has branch points at $1, \infty$, and $\pm u^{n/2}$. In particular, its discriminant behaves as $\Delta \sim u^n$ for u near 0. Therefore, in the stable case the exponent of the monodromy is the order of vanishing of the discriminant.

For example, for the family of curves (11.7), the branch points are at $0, \infty$, and $\frac{1}{2}(u \pm \sqrt{u^2 - \Lambda^4})$, as we noted above. If $u = \Lambda^2 + \epsilon$, the branch points for ϵ small are located at approximately $0, \frac{1}{2}(\Lambda^2 \pm \Lambda\sqrt{2\epsilon})$ and ∞. Thus, the discriminant is proportional to ϵ, so $n = 1$ and the monodromy near $u = \Lambda^2$ is conjugate to T. In view of the $\mathbb{Z}_2$ symmetry of the u plane, the same is true of the monodromy near $u = -\Lambda^2$.

For u near ∞, the structure is slightly more complicated. The branch points for large u are at approximately $0, \Lambda^4/4u, u$ and ∞. The singularity at infinity is not stable since two pairs of branch points (0 and $\Lambda^4/4u$, and also u and ∞) coincide for $u \to \infty$. By a change of variables

$$x = x'u,$$

$$y = y'u^{3/2} \tag{11.15}$$

we get a stable situation with branch points approximately 0, $\Lambda^4/4u^2$, 1, and ∞ and only one pair that coincides for $u \to \infty$. This gives $\Delta \sim u^{-4}$, so the monodromy at infinity for the family of curves in the x'–y' plane is conjugate to T^4 if one circles counter-clockwise around the origin in the u^{-1} variable; a counter-clockwise circuit in u gives T^{-4}. Going back to the x–y plane, we must note that the differential form dx/y picks up a factor of $u^{1/2}$ from the change of variables in (11.15); hence its periods da_D/du and da/du pick up a factor of $u^{1/2}$. Since $u^{1/2}$ is odd under the monodromy at infinity, this gives an extra factor of P (the operator that acts as -1 on (a_D, a)) in the monodromy at infinity, so that the monodromy for the original family of curves is PT^{-4}.

It can be shown straightforwardly that there is only one representation of the fundamental group of the u plane punctured at $\pm\Lambda^2$ and ∞ with the monodromies conjugate to T, T, and PT^{-4} and a $\mathbb{Z}_2$ symmetry exchanging the first two singular points; this gives another way to prove that (11.7) is the right family. In fact, one can work out the required monodromies for all $N_f \leqslant 3$ using the general properties unearthed in Sections 7, 8. The monodromies at infinity are determined by the one-loop beta functions in the microscopic theory to be

$$\mathcal{M}_\infty = PT^{N_f - 4}, \tag{11.16}$$

just as we discussed for $N_f = 0$ in Ref. [9]. According to our proposal in the previous sections, with one exception for $N_f = 3$, the singularities at finite points in the u plane correspond to massless magnetic monopoles with one unit of magnetic charge. The monodromy around such a point is determined by the one-loop beta function in QED with k hypermultiplets. From the infrared behavior of QED as in Ref. [9], Section 5.4, it is conjugate to T^k. We pick a base point at infinity and then the quantum numbers (n_m, n_e) of the massless monopoles are well defined. The monodromy around a point with k magnetic monopoles with $(n_m = 1, n_e)$ is $(T^{n_e}S)T^k(T^{n_e}S)^{-1}$ as $T^{n_e}S$ conjugates a hypermultiplet with $(0, 1)$ to $(1, n_e)$. The monodromy for the $(2, 1)$ state in $N_f = 3$ can be determined similarly. Altogether the monodromies are

$$
\begin{array}{llll}
& STS^{-1}, \quad (T^2S)T(T^2S)^{-1}; & & \text{for } N_f = 0 \\
STS^{-1}, & (TS)T(TS)^{-1}, \quad (T^2S)T(T^2S)^{-1}; & & \text{for } N_f = 1 \\
& ST^2S^{-1}, \quad (TS)T^2(TS)^{-1}; & & \text{for } N_f = 2 \\
(ST^2S)T(ST^2S)^{-1}, & ST^4S^{-1}; & & \text{for } N_f = 3
\end{array}
\tag{11.17}
$$

It is straightforward to check that the product of the monodromies is $\mathcal{M}_\infty$ of (11.16). This is another check of our assertions about the singularities.

11.3. General structure of the curve

In trying to generalize (11.7) to $N_f > 0$, it is useful to first extract some properties of the $N_f = 0$ curve that we will try to generalize:
(1) The family is of the form $y^2 = F(x, u, \Lambda)$, where F is a polynomial in x, u, and Λ that is at most cubic in x and u.

(2) The part of $F(x, u, \Lambda)$ that is cubic in x and u is $F_0 = x^2(x - u)$.

(3) If one assigns the $U(1)_{\mathcal{R}}$ charges 4 to u and x, 2 to Λ, and 6 to y, then the family is $U(1)_{\mathcal{R}}$ invariant; and in particular F has charge 12.

(4) F can be written $F = F_0 + \Lambda^4 F_1$ where $F_1 = \frac{1}{4}x$.

For the time being we will simply assume that property (1) remains valid in the presence of matter hypermultiplets. (This property can be at least partly deduced from our later consideration of the $N = 4$ and $N_f = 4$ theories.)

As for property (2), it is easy to see that it must be valid for $N_f \leqslant 3$. The cubic part of F (after absorbing a possible multiplicative constant in y) would in general take the form $(x - e_1 u)(x - e_2 u)(x - e_3 u)$ for some e_i. For large u the branch points are at $e_1 u, e_2 u, e_3 u$, and ∞. After dividing x by u, the branch points for $u \to \infty$ are at the e_i and ∞, so they are distinct if the e_i are distinct. Thus, if the e_i are distinct, the family of curves has no singularity for $u = \infty$. This is actually the correct behavior for the conformally invariant theories, but for $N_f \leqslant 3$ it would prevent one from getting for $u \to \infty$ the logarithm associated with asymptotic freedom as in (4.3). On the other hand, one does not want all three e_i equal, for then u could be eliminated from the cubic term by changing variables from x to $x' = x - e_1 u$; it would then again be impossible to get a logarithm. So we can assume for asymptotically free theories that $e_1 = e_2 \neq e_3$. By a redefinition of x by $x \to ax + bu$ and a rescaling of y, one can assume that $e_1 = 0$, $e_3 = 1$, and hence that the cubic part of F is $x^2(x - u)$.

As far as the third property is concerned, one expects that $U(1)_{\mathcal{R}}$ can be treated as a symmetry if Λ is assigned the correct charge. 4 and 2 are indeed the correct $U(1)_{\mathcal{R}}$ charges of u and Λ. For u, this is just a statement about the classical theory. To understand the charge of Λ, one notes that the one loop beta function and $U(1)_{\mathcal{R}}$ anomaly are such that a one-instanton amplitude is proportional to Λ^4 and violates $U(1)_{\mathcal{R}}$ by eight units. The physical meaning of x and y is somewhat mysterious, so we will have to accept their $U(1)_{\mathcal{R}}$ charges of 4 and 6 as an empirical fact.

It is now easy to interpret property (4). Given the anomalous $U(1)_{\mathcal{R}}$ conservation and the fact that F is polynomial, it is clear that in the weak coupling limit of the theory, obtained by taking $\Lambda \to 0$, F reduces to the cubic term F_0. Moreover, since it is only instantons that violate the $U(1)_{\mathcal{R}}$ charge if we do *not* assign a $U(1)_{\mathcal{R}}$ charge to Λ, and a one-instanton amplitude has the quantum numbers of Λ^4, the expansion of F in powers of Λ^4 can be interpreted as an expansion in the instanton number. The fact that $F = F_0 + \Lambda^4 F_1$ means that for $N_f = 0$, F has only the classical contribution F_0 and the one instanton contribution $\Lambda^4 F_1$.

12. The curve for $N_f = 1$

12.1. Massless $N_f = 1$

Our next goal is to adapt the principles just described to determine the families of curves that control the low energy behavior of the $N = 2$ theories with matter. In doing so, we let Λ_{N_f} be the renormalization scale parameter of the theory with N_f flavors. We consider first the case of $N_f = 1$ with zero bare mass.

For $N_f = 1$, the one-instanton amplitude is proportional to Λ_1^3, and an amplitude with r instantons is proportional to Λ_1^{3r}. However, for $N_f > 0$ and one or more massless hypermultiplets, the equation defining the curve can only receive contributions from terms with an even number of instantons. This is because an amplitude with an odd number of instantons violates the internal "parity" symmetry of Eq. (3.2), while the curve is invariant under this symmetry. For $N_f = 1$, given that Λ_1 has $U(1)_{\mathcal{R}}$ charge 2 and F has charge 12, the only possible term in F that depends on Λ_1 is a two-instanton term, which must be a constant times Λ_1^6. Therefore, the curve of the massless $N_f = 1$ theory must be

$$y^2 = x^2(x - u) + t\Lambda_1^6. \tag{12.1}$$

The constant t can be absorbed in a redefinition of Λ_1

$$\widetilde{\Lambda}_1^6 = t\Lambda_1^6. \tag{12.2}$$

Below we will determine t for a particular definition of Λ_1.

Let us now see that this curve has the right properties. First of all, we have the expected $\mathbb{Z}_3$ symmetry of the u plane, as (12.1) is invariant under multiplying x and u by a cube root of unity, with y invariant. (From the formulas below for a and a_D, it will be clear that a and a_D transform correctly under the symmetry.) To find the singularities, we compute the discriminant of the polynomial on the right-hand side of (12.1), with the result

$$\Delta = \widetilde{\Lambda}_1^6(4u^3 - 27\widetilde{\Lambda}_1^6). \tag{12.3}$$

Consequently, on the finite u plane there are three singular points at $u = e^{2\pi i n/3}3\widetilde{\Lambda}_1^2/4^{1/3}$, $(n = 1, 2, 3)$, permuted by the $\mathbb{Z}_3$ symmetry, and with monodromy conjugate to T. To find the behavior for large u, we note that for large u the branch points are approximately at $x = u, \infty$, and $\pm\widetilde{\Lambda}_1^3/\sqrt{u}$. Upon absorbing a factor of u in x and a factor of $u^{\frac{3}{2}}$ in y, we get a stable situation with branch points at $1, \infty$ and $\pm\widetilde{\Lambda}_1^3/u^{3/2}$. The discriminant is hence of order u^{-3} for $u \to \infty$, so the monodromy is conjugate to PT^{-3} (the factor of P arises from the factor of $u^{3/2}$ in y).

12.2. Massive $N_f = 1$

Now we consider the $N_f = 1$ theory with a non-zero bare mass m for the hypermultiplet. m-dependent terms must vanish for $\Lambda_1 \to 0$, since classically the mass m of the hypermultiplet does not affect the low energy couplings of the vector multiplet. m is odd under the internal "parity," so with $m \neq 0$, one can have contributions of odd instanton number to the equation defining the curve; they simply have to be odd in m. Since m has $U(1)_{\mathcal{R}}$ charge 2, the most general possibility is

$$y^2 = x^2(x - u) + t\Lambda_1^6 + m\Lambda_1^3(ax + bu) + cm^3\Lambda_1^3 \tag{12.4}$$

with constants a, b, and c that must be determined.

To do so we note that when m is large the quark can be integrated out. The low energy theory is the pure gauge $N = 2$ theory; its scale Λ_0 is determined by the one-loop beta function to be

$$\Lambda_0^4 = m\Lambda_1^3. \tag{12.5}$$

Λ_0 for the pure gauge theory was defined such that the singularities are at $u = \pm\Lambda_0^2$. Then we can use Eq. (12.5) as a definition of Λ_1. Other definitions corresponding to different subtraction schemes differ by a multiplicative constant in (12.5).

An obvious constraint on (12.4) is that in the limit of large m we recover the curve of the $N_f = 0$ theory. More precisely, we should consider the scaling limit $m \to \infty$, $\Lambda_1 \to 0$ with $\Lambda_0^4 = m\Lambda_1^3$ held fixed. Comparison of (12.4) and (11.7) shows that $a = \frac{1}{4}$ and $b = c = 0$. Therefore, the curve is

$$y^2 = x^2(x - u) + \frac{1}{4}m\Lambda_1{}^3 x + t\Lambda_1{}^6. \tag{12.6}$$

By computing the discriminant, one can verify that in the double scaling limit, one of the singularities moves to infinity; in fact the singularity in question is at

$$u \approx -\frac{m^2}{64t}, \qquad x \approx -\frac{8t\Lambda_1{}^3}{m}. \tag{12.7}$$

(It is not necessary here to actually consider the discriminant of the cubic polynomial on the right of (12.6); the term cubic in x is unimportant near the singularity, which is controlled by the discriminant and zeroes of the quadratic polynomial $-ux^2 + \frac{1}{4}m\Lambda_1{}^3 x + t\Lambda_1{}^6$.)

The constant t cannot be determined by the singularities and monodromies of the curve, since they are invariant under scaling of m or Λ_1. To fix t, we note that for large m, there should be a singularity at $u \approx m^2$ where one of the elementary quarks becomes massless. Comparing to (12.7), we see that $t = -1/64$, so the curve is

$$y^2 = x^2(x - u) + \frac{1}{4}m\Lambda_1{}^3 x - \frac{1}{64}\Lambda_1{}^6. \tag{12.8}$$

Of course, we see here the merits of the change of normalization that led us from (11.1) to (11.7). If one wishes to study the $N_f = 0$ theory as a low energy limit of a massive theory with $N_f > 0$, which has fields of one-half the W boson charge, it will appear naturally in the form of (11.7). Similarly (as we will see later), renormalization group flow from the mass-deformed $N = 4$ theory, which does not have isodoublets, will lead naturally to (11.1).

Note that given the $N_f = 1$ curve, the $N_f = 0$ curve can be immediately determined, but not the other way around. Of course, that is a manifestation of the usual irreversibility of the renormalization group.

13. The curve for $N_f = 2$

13.1. Massless $N_f = 2$

First we consider the $N_f = 2$ theory without masses. The instanton amplitude is now Λ_2^2, and the internal parity implies that without bare masses, only amplitudes with an even number of instantons contribute. Given that Λ_2 has $U(1)_{\mathcal{R}}$ charge 2 and the equation for the curve has charge 12, the curve can receive contributions only from instanton number 2 and is of the form

$$y^2 = x^2(x - u) + ax\Lambda_2^4 + bu\Lambda_2^4. \tag{13.1}$$

It remains to determine a and b.

Let us denote the polynomial on the right-hand side of (13.1) as F. The discriminant of F (with respect to x) is a polynomial in u that is quartic (or lower order, for some special values of a and b). Generically, this polynomial has four simple zeroes, so that the family of curves (13.1) has four singularities in the u plane, with monodromies conjugate to T. We want to find the values of a and b for which there are instead two singularities with monodromies conjugate to T^2. This can be done by studying the discriminant or more directly as follows.

As we have noted before, near a point $u = u_0$ at which F has a double root at $x = x_0$, F looks like $F = \text{const} \cdot ((x - x_0)^2 - (u - u_0)^n)$, for some n; the monodromy is then conjugate to T^n. The condition that $n > 1$ is that

$$F = \frac{\partial F}{\partial x} = \frac{\partial F}{\partial u} = 0 \tag{13.2}$$

at $x = x_0$, $u = u_0$. (This is equivalent to saying that the two dimensional variety defined by the equation $y^2 = F(x, u)$ in three variables x, u, and y has a singularity at $x = x_0$, $u = u_0$, $y = 0$.) One immediately sees that, with F as above, Eqs. (13.2) require $a = -b$. As in our discussion of $N_f = 1$, these considerations cannot determine the normalization of Λ_2 and we conclude that the family of curves is

$$y^2 = (x^2 - \widetilde{\Lambda}_2^4)(x - u) \tag{13.3}$$

for some $\widetilde{\Lambda}_2$. Below we will express it in terms of a convenient definition of Λ_2. This, indeed, is a family that we have seen before – it is essentially our old friend (11.1) of the $N_f = 0$ theory in the "old" normalization. Note in particular that the expected $\mathbb{Z}_2$ symmetry of the u plane has appeared.

13.2. Massive $N_f = 2$

Now we allow arbitrary bare masses m_1, m_2 for the quarks. The masses transform like $(3, 1) \oplus (1, 3)$ of the global $SO(4) \cong SU(2) \times SU(2)$. The polynomial defining the curve must be $SO(4)$ invariant. Also, terms that are odd (or even) under the internal parity which changes the sign of one of the m's must come from an odd (or even) number of instantons. The most general possible structure of the curve is then

$$y^2 = (x^2 - t\Lambda_2{}^4)(x - u) + m_1 m_2 \Lambda_2{}^2 (ax + bu) + c(m_1{}^2 + m_2{}^2)\Lambda_2{}^4 \qquad (13.4)$$

with constants a, b, and c that must be determined. We define Λ_2 such that when m_2 is large, the scale of the low energy $N_f = 1$ theory is $\Lambda_1^3 = m_2 \Lambda_2^2$. Then, we should also determine the constant t relating Λ_2 to $\widetilde{\Lambda}_2$.

We proceed as in our discussion of the massive $N_f = 1$ theory. The scaling limit $m_2 \to \infty$, $\Lambda_2 \to 0$ with $\Lambda_1^3 = m_2 \Lambda_2^2$ held fixed determines $a = \frac{1}{4}$, $b = 0$ and $c = -1/64$. The curve is then

$$y^2 = (x^2 - t\Lambda_2{}^4)(x - u) + \frac{1}{4} m_1 m_2 \Lambda_2{}^2 x - \frac{1}{64}(m_1{}^2 + m_2{}^2)\Lambda_2{}^4. \qquad (13.5)$$

As for $N_f = 1$, the constant t is determined by imposing that when any mass, say m_1, is large there is a singularity at $u = m_1^2$. Substituting $m_2 = 0$ and $u = m_1^2$ in the discriminant of (13.5), the leading term at large m_1 is $4m_1{}^8 \Lambda_2{}^4(t - 1/64)$; hence $t = 1/64$ and the curve is

$$y^2 = (x^2 - \frac{1}{64}\Lambda_2{}^4)(x - u) + \frac{1}{4} m_1 m_2 \Lambda_2{}^2 x - \frac{1}{64}(m_1{}^2 + m_2{}^2)\Lambda_2{}^4. \qquad (13.6)$$

As a check, we can compute the discriminant for $m_1 = m_2 = m$ and find

$$\Delta = \frac{\Lambda_2{}^4}{16}(u + m\Lambda_2 + \frac{1}{8}\Lambda_2^2)(u - m\Lambda_2 + \frac{1}{8}\Lambda_2^2)(u - m^2 - \frac{1}{8}\Lambda_2^2)^2. \qquad (13.7)$$

This shows that when the two quarks are degenerate there is a singularity at $u = m^2 + \Lambda_2^2/8$ with two massless particles (double zero of the discriminant) which for large m can be identified with that of the massless quarks. The reason for the two-fold degeneracy there is that the massless particles at the singularity are in a doublet of the global SU(2) symmetry of the massive theory. The other two singularities – which coincide for $m = 0$ – come from simple zeroes of the discriminant corresponding to vacua with one massless multiplet each.

14. The curve for $N_f = 3$

14.1. Massless $N_f = 3$

Now we will determine the curve for the $N_f = 3$ theory with zero bare masses. We do this using the expected singularity structure: there should be a singularity in the u plane with monodromy conjugate to T^4 and one with monodromy conjugate to T.

Since the $N_f = 3$ theory has no symmetry in the u plane, we may as well add a constant to u and assume that the singularity with monodromy T^4 is at $u = 0$. The curve is then given by $y^2 = F(x, u)$, where F is cubic in x and u, and at $u = 0$, $F(x, 0) = 0$ has a double root; by adding a constant to x, we may as well assume that the double root is at $x = 0$. In fact, F must have the form

$$F = a\Lambda_3{}^2 x^2 + bu^2 x + cux^2 + x^3. \qquad (14.1)$$

This polynomial has two roots within $O(u^2)$ of the origin, so the discriminant is proportional to u^4 for small u and the monodromy around $u = 0$ is conjugate to T^4. The structure in (14.1) was determined as follows. Terms u^n with $n \leqslant 3$ or xu^m with $m < 2$ would cause the zero of the discriminant at $u = 0$ to be of order lower than 4, so those terms have been suppressed. The coefficient of x^3 was set to one to agree with the classical limit.

To determine the parameters, we require that $b \neq 0$, since otherwise the curve is singular for all u, not the situation we want. We also ask that (14.1) has the right classical limit; this is equivalent to saying that the cubic part of F, which is $x^3 + cux^2 + bu^2x$, is equal to $(x - e_1u)^2(x - e_2u)$ with $e_1 \neq e_2$. Since $b \neq 0$, the only way to achieve this is to have $x^2 + cux + bu^2$ be a perfect square, say $(x + \alpha u)^2$. After rescaling x and y so that $\alpha = 1$, and renaming $x + u$ as x, we finally determine our curve:

$$y^2 = x^2(x - u) + t\Lambda_3{}^2(x - u)^2. \tag{14.2}$$

t is a constant which can be absorbed in a redefinition of Λ_3.

Notice that for $N_f = 3$, a one-instanton amplitude is proportional to Λ_3, so the term in (14.2) proportional to $\Lambda_3{}^2$ can be interpreted as a two-instanton effect. Thus, the contributions from an odd number of instantons vanish, as expected from the internal "parity."

The other singularity of this family, with monodromy conjugate to T, is at $u = -t\Lambda_3{}^2/4$. This singularity is presumably associated with vanishing mass of a state of $(n_m, n_e) = (2, 1)$.

14.2. Massive $N_f = 3$

We now determine the curve for the massive $N_f = 3$ theory. We repeat the steps we used for $N_f = 1, 2$. The masses m_1, m_2, m_3 are in the **15** of the global SO(6) symmetry. The equation for the curve must be compatible with SO(6) invariance and the anomalous U(1)$_\mathcal{R}$, and must reduce to the previous result when the masses vanish. Also, it should be invariant under the parity transformation which changes the sign of one of the masses and the sign of the instanton factor Λ_3. And it should flow to the massive $N_f = 2$ curve in the scaling limit of $m_3 \to \infty$ with $\Lambda_2^2 = m_3\Lambda_3$ fixed. The most general polynomial equation with these properties is

$$\begin{aligned}
y^2 &= x^2(x - u) + t\Lambda_3{}^2(x - u)^2 - \frac{1}{64}(m_1{}^2 + m_2{}^2 + m_3{}^2)\Lambda_3{}^2(x - u) \\
&\quad + \frac{1}{4}m_1m_2m_3\Lambda_3 x - \frac{1}{64}(m_1{}^2m_2{}^2 + m_2{}^2m_3{}^2 + m_1{}^2m_3{}^2)\Lambda_3{}^2 \\
&\quad + a(m_1{}^2 + m_2{}^2 + m_3{}^2)\Lambda_3{}^4 + bm_1m_2m_3\Lambda_3{}^3.
\end{aligned} \tag{14.3}$$

Here we encounter a new element that did not arise for $N_f = 1, 2$. The two constants a and b are not determined by the previous considerations nor are they a choice of scale (like t). To determine them we examine the small m theory.

For $m_1 = m_2 = m_3 = 0$ there are two singularities. One of them is at $u = 0$ with four massless particles and the other at $u = -t\Lambda_3{}^2/4$ where there is a single massless

particle. Correspondingly, the discriminant of the massless curve has a single zero at $u = -t\Lambda_3^2/4$ and a fourth order zero at $u = 0$. When some of the masses are turned on the zero of the discriminant at $u = -t\Lambda_3^2/4$ moves and the multiple zero at $u = 0$ splits. We will now examine this splitting and determine the constants a and b. It is enough to consider the case with $m_1 = m_2 = m_3 = m$. From our analysis of the monopole structure, we know that the zero at $u = 0$ should split to a single zero and a triple zero. The discriminant is of course holomorphic in all variables, and since the single and triple zero near $u = 0$ are unique, their positions also vary holomorphically. (By contrast, if a fourth order zero splits to two double zeroes, the formula for the locations of the two double roots can contain a square root branch cut.) Therefore, for small m the discriminant of the curve with three equal masses should have a zero at $u \sim m$. Expanding the discriminant of (14.3) in m with $u = \lambda m$ we find that actually

$$\Delta = -4m^2 t^2 \Lambda_3^8 \left[3\Lambda_3^2 at + m(t\Lambda b + 18\lambda a) + \mathcal{O}(m^2) \right]. \tag{14.4}$$

To get a zero of Δ for $m \to 0$ with λ of order 1, the coefficient of m^2 has to vanish. Since $t \neq 0$, this forces $a = 0$. Then, vanishing of the $\mathcal{O}(m^3)$ term forces $b = 0$.

To determine t we impose that for large m there is a singularity at $u = m^2$. The leading contribution to the discriminant at large m with $u = m^2$ is $\frac{1}{16}m^{10}\Lambda_3^2(1 + 64t)$ and therefore $t = -1/64$.

We conclude that the curve is

$$y^2 = x^2(x - u) - \frac{1}{64}\Lambda_3^2(x - u)^2 - \frac{1}{64}(m_1^2 + m_2^2 + m_3^2)\Lambda_3^2(x - u)$$
$$+ \frac{1}{4}m_1 m_2 m_3 \Lambda_3 x - \frac{1}{64}(m_1^2 m_2^2 + m_2^2 m_3^2 + m_1^2 m_3^2)\Lambda_3^2. \tag{14.5}$$

As a test we evaluate the discriminant for three equal masses m and find that it has a triple zero at $u = m^2 + m\Lambda_3/8$. The three particles at the singularity are in the fundamental representation of the global SU(3) symmetry. As m varies from zero to infinity they move from the origin where they are magnetic monopoles to infinity where they are interpreted as the elementary quarks.

15. Masses, periods and residues

As in Ref. [9], the particle masses and the low energy metric and couplings are now determined by equating a and a_D with periods of a certain meromorphic one-form λ on the curve E. λ has two characteristics: (i) λ may have poles but (as long as the monodromies are in SL(2, $\mathbb{Z}$)) its residues vanish; (ii) to achieve positivity of the metric on the quantum moduli space, its derivative with respect to u is proportional to dx/y.

Condition (i) means that the definition of a and a_D by contour integrals

$$a = \int_{\gamma_1} \lambda,$$

$$a_D = \int_{\gamma_2} \lambda, \tag{15.1}$$

– with γ_1 and γ_2 some contours on E – is invariant under deformation of the γ_i, even across poles of λ. This ensures that only the homology classes of the γ_i matter and reduces the monodromies to a group $SL(2, \mathbb{Z})$ that acts on $H_1(E, \mathbb{Z})$. In the presence of bare masses, this is too strong a condition since (as we saw originally in the discussion of $N = 2$ QED in Section 2.4) when the bare masses are non-zero the monodromies are not quite in $SL(2, \mathbb{Z})$.

As for condition (ii), the differential form dx/y has no poles and represents a cohomology class on E of type $(1, 0)$. Having $d\lambda/du = f(u)\, dx/y$ leads to positivity of the metric as we explained in Ref. [9]. The function $f(u)$ is determined by requiring the right behavior at the singularities, for instance $a \approx \frac{1}{2}\sqrt{2u}$ for large u. f is a constant for the same reasons as in Ref. [9]. The proper relation is in fact

$$\frac{d\lambda}{du} = \frac{\sqrt{2}}{8\pi} \frac{dx}{y}. \tag{15.2}$$

Up to an inessential sign, this is $1/2$ the value in the "old" conventions. By integration with respect to u, (15.2) determines λ (once the curve is known) for all values of N_f. (15.2) is only supposed to hold up to a total differential in x; λ is supposed to be meromorphic in x.

But when one obtains λ by integration of (15.2), does one get a result that obeys condition (i)? As an example,[6] consider the massive $N_f = 2$ theory with

$$y^2 = (x^2 - \frac{1}{64}\Lambda_2^4)(x - u) + \frac{1}{4}m_1 m_2 \Lambda_2^2 x - \frac{1}{64}(m_1^2 + m_2^2)\Lambda_2^4.$$

In this case, a meromorphic λ obeying (15.2) can be found by inspection:

$$\lambda = \frac{-\sqrt{2}}{4\pi} \frac{dx\, y}{x^2 - \frac{1}{64}\Lambda_2^4}. \tag{15.3}$$

Note that λ has poles at $x = \pm\frac{1}{8}\Lambda_2^2$ whose residues are[7]

$$\frac{\pm m_1 \pm m_2}{(2\pi i)2\sqrt{2}}. \tag{15.4}$$

Thus, condition (i), as we have stated it so far, holds only when the bare masses are zero.

---

[6] We consider this example because the theories with $N_f \geqslant 3$ require considerably more powerful methods – developed in the last section of this paper – to analyze condition (i). On the other hand, once the conditions are imposed for $N_f = 2$, there is no need to consider $N_f < 2$ separately since the correct behavior for those cases follows from $N_f = 2$ by the renormalization group flow.

[7] Note that the residues are independent of u. The reason for this is that $\omega = d\lambda/du$ has zero residues. The operation of taking the residue in x commutes with differentiation with respect to u, so the fact that ω has zero residues means that the residues of λ are annihilated by d/du.

More generally, when the bare masses are not zero, the residues of λ mean that the definition of a and a_D in (15.1) is *not* invariant under deforming the contours γ_i across the poles of λ. Under such a deformation, a and a_D would change by a constant – the residue of the pole.

It follows that the monodromies are no longer simply in $SL(2, \mathbb{Z})$. In defining a and a_D, one tries to take a smoothly varying family of contours that keeps away from the poles of λ. However, in looping around a closed path in the u plane, the contour might come back on the wrong side of a pole. When this happens, in addition to the $SL(2, \mathbb{Z})$ action, a and/or a_D will jump by a constant.

But from our analysis in Sections 2 and 6, we know that precisely such constants are needed when the bare masses m are not zero. In fact, according to Eq. (6.2), the constant jump should be of the form $\sum_i m_i S_i / \sqrt{2}$. The S_i are abelian conserved charges that appear in the central extension of the $N = 2$ algebra along with the electric and magnetic charges.

In the $N_f = 2$ theory, the S_i are integers for the fundamental particles and are non-zero half integers for monopoles. Therefore, the allowed jumps are linear combinations of $(\pm m_1 \pm m_2)/2\sqrt{2}$ – exactly as we see in (15.4). (Note that since a is a contour integral of λ, the jump in a is $2\pi i$ times the residue of λ.)

In sum, in the massive theory, we do not want the residues of λ to be zero. We want them to be linear in the masses and in particular independent of Λ_k. It is a non-trivial test of our answers for the curves that the residues of λ have the right form. In fact, the detailed structure of the equations for the curves could all have been determined from this one condition alone; that will be our strategy in Section 17.

Singularities from semi-classical states

In Ref. [9] and the present paper, we have assumed that singularities of the curve come not from massless non-abelian gauge fields but from massless hypermultiplets of spins $\leqslant 1/2$. $N = 2$ multiplets of such low spin are necessarily BPS-saturated, and – as BPS-saturated states have strong stability properties – it is natural to expect that such states would be the continuation to the strong coupling regime of states that can be seen semiclassically; for brevity we will refer to such states as semiclassical states.

Moreover, *ex post facto*, from the curves that we obtain, it is possible to show that (if the curves are correct) the singularities must be due to semiclassical states. The approach to showing this was explained in the last paragraph of Ref. [9]: one shows that one can interpolate from the semiclassical region of large u to the singularities along a path on which a/a_D is never real. The importance of the requirement that a/a_D is not real is that (see the end of Section 4 of Ref. [9]) as long as a/a_D never becomes real, the spectrum of BPS-saturated states cannot jump and therefore the BPS-saturated states are precisely the semiclassical ones.

The argument given at the end of Ref. [9] can be carried over to the models of the present paper, but for brevity we will do this only in the case which is perhaps most interesting: the $(n_m, n_e) = (2, 1)$ state for $N_f = 3$. What makes this the most interesting example is that, in fact, it is not known whether this state exists semiclassically; in view of what we are about to say, its semiclassical existence is a prediction of our analysis. This prediction should be testable using the methods of Refs. [18,25]. (Of course, our

proposal of $SL(2, \mathbb{Z})$ symmetry for $N_f = 4$ makes many similar predictions.)

In units with $\Lambda_3^2/64 = 1$, the massless $N_f = 3$ curve is

$$y^2 = x^2(x - u) - (x - u)^2. \tag{15.5}$$

The polynomial on the right-hand side has zeroes at $x_0 = u$ and at

$$x_\pm = \frac{1}{2}\left(1 \pm \sqrt{1 - 4u}\right). \tag{15.6}$$

In particular, at $u = 1/4$, $x_+ = x_-$, giving the singularity that we have attributed to a massless state of $(n_m, n_e) = (2, 1)$. To show that this state is semiclassical, we will interpolate on the positive u axis from the semiclassical regime of $u \rightarrow \infty$ to the singularity at $u = 1/4$. For $u > 1/4$, x_+ and x_- are complex conjugates.

We have

$$\frac{da}{du} = \int_{\gamma_1} \omega$$

$$\frac{da_D}{du} = \int_{\gamma_2} \omega, \tag{15.7}$$

where $\omega = (\sqrt{2}/8\pi)dx/y$, γ_1 is a circle in the x plane that loops around x_+ and x_- but not x_0, and γ_2 is a contour that loops round x_0 and x_+ but not x_-. Complex conjugation leaves x_0 alone and exchanges x_+ with x_-; hence γ_1 is invariant under complex conjugation, but complex conjugation turns γ_2 into a contour γ_3 that loops around x_0 and x_- while avoiding x_+. So a is real but the complex conjugate of a_D is given by

$$\frac{d\bar{a}_D}{du} = \int_{\gamma_3} \omega. \tag{15.8}$$

γ_3, however, is homotopic to the sum of $-\gamma_1$ and $-\gamma_2$. (The minus sign comes from keeping track of the orientations of the contours.) Hence, (15.8) gives

$$\bar{a}_D = -a - a_D. \tag{15.9}$$

In other words,[8]

$$a_D = -\frac{a}{2} + \text{imaginary}. \tag{15.10}$$

Now, jumping of BPS saturated states can only occur when a_D/a is real; in other words, for u positive and greater than $1/4$, it can only occur when the imaginary part

---

[8] This equation has the following interpretation. The curve (15.5) is real for real u, that is, the coefficients in the equation are real. There are two types of real elliptic curve: τ can have real part zero or $1/2$. (Thus τ is either invariant or transformed by the $SL(2, \mathbb{Z})$ transformation $\tau \rightarrow \tau - 1$ under the complex conjugation operation $\tau \rightarrow -\bar{\tau}$.) The two possibilities correspond in a suitable basis to a real, a_D imaginary, or a real, $a_D = -a/2 +$ imaginary. For $u > 1/4$ we have the second possibility.

of a_D vanishes and so $2a_D + a = 0$. When that happens, a BPS-saturated state of $(n_m, n_e) = (2, 1)$ becomes massless, if there is such a state.

If our curve is correct, there must be such a state at $u = 1/4$ since we need it to generate the singularity. Let u' be the smallest value of $u > 1/4$ at which the imaginary part of a_D vanishes. Then, one can interpolate from $u = 1/4$ to $u = u'$ without any jumping of BPS-saturated states; hence, the $(2, 1)$ state would again be massless at u'. This should produce an extra singularity that our curve does not have. So if the curve is correct, u' does not exist, Im a_D never vanishes on the positive u axis for $u > 1/4$, one can interpolate from $u = 1/4$ to $u = \infty$ without jumping, and the $(2, 1)$ state that gives the singularity at $u = 1/4$ must be semiclassical.

Of course, it would be desirable to sharpen this argument and prove directly from (15.7) that a_D is never real for $u > 1/4$.

16. Structure of the scale invariant theories

In this section, we will analyze the $N_f = 4$ and $N = 4$ theories. In the absence of bare masses, those theories are conformally invariant. In the presence of $N = 2$-invariant bare masses, their properties (or at least the properties that we can analyze) are much richer.

The challenge that these theories present is that there is a dimensionless coupling constant,

$$
\tau = \begin{cases} \dfrac{\theta}{\pi} + \dfrac{8\pi i}{g^2} & \text{for } N_f = 4 \\[2mm] \dfrac{\theta}{2\pi} + \dfrac{4\pi i}{g^2} & \text{for } N = 4. \end{cases} \tag{16.1}
$$

Therefore, in the curve $y^2 = F(x, u, m_i, \tau)$ that controls the low energy behavior, the coefficients are functions of τ that must be determined. This contrasts with $N_f < 4$ where, instead of τ, one has the renormalization scale Λ; dimensional analysis ensures that (if F is holomorphic and free of singularities [9]) the dependence on Λ is polynomial, so that there are only finitely many parameters to determine.

Of course, if one is willing to assume $SL(2, \mathbb{Z})$ duality, and if one can guess the modular weights of x, y, u, and the m_i, then the coefficients are modular forms, which are determined by $SL(2, \mathbb{Z})$ in terms of finitely many coefficients. The assumption of $SL(2, \mathbb{Z})$ would thus put us back in a situation similar to that which we have already encountered for $N_f < 4$. We will not follow that road because instead of assuming $SL(2, \mathbb{Z})$ invariance, we want to deduce it. Therefore, we face a more difficult task. We will in fact provide two different routes to the goal; in this section we analyze the $N_f = 4$ and $N = 4$ models by more careful application of the methods that we have used for $N_f \leqslant 3$, while in the next section we use a method suggested at the end of Section 15.

---

[9] Holomorphy of F is needed in order for the coefficient $\tau(u, m_i, \Lambda)$ in front of W_α^2 in the low energy theory to be holomorphic in the field u and the parameters m_i and Λ, as follows from Ref. [3]. Absence of singularities in F is less obvious.

16.1. The massless case

The first step is to find the right family of curves for the conformally invariant case, that is when the bare masses are zero. In this case, the classical formulas

$$
a = \begin{cases} \frac{1}{2}\sqrt{2u}, & \text{for } N_f = 4 \\ \sqrt{2u} & \text{for } N = 4, \end{cases}
$$
$$
a_D = \tau a \tag{16.2}
$$

are exact. So we wish to find a curve $y^2 = F(x, u, \tau)$ such that the differential form

$$
\omega = \begin{cases} \dfrac{\sqrt{2}}{8\pi} \dfrac{dx}{y} & \text{for } N_f = 4 \\[2mm] \dfrac{\sqrt{2}}{4\pi} \dfrac{dx}{y} & \text{for } N = 4, \end{cases} \tag{16.3}
$$

has the periods $(\partial a_D/\partial u, \partial a/\partial u)$, with (a_D, a) given in (16.2).

Now, a genus one curve E and a differential form with periods a multiple of $(\tau, 1)$ can be found as follows. Let E be the quotient of the complex z plane by the lattice generated by π and $\pi\tau$. Let $\omega_0 = dz$. Obviously, the periods of ω_0 (integrated on contours that run from 0 to π and from 0 to $\pi\tau$, respectively) are π and $\pi\tau$.

To find an algebraic description of E (see Section I.6 of Ref. [26] or Section III.3 of Ref. [27]), one introduces the Weierstrass $\mathcal{P}$ function, which obeys

$$
\mathcal{P}(z) = \mathcal{P}(z+1) = \mathcal{P}(z+\tau) = \mathcal{P}(-z) \tag{16.4}
$$

and has for its only singularity on E a double pole at the origin. If one sets $x_0 = \mathcal{P}(z)$, $y_0 = \mathcal{P}'(z)$, then one finds

$$
y_0^2 = 4x_0^3 - g_2(\tau)x_0 - g_3(\tau). \tag{16.5}
$$

Here $g_2 = 60\pi^{-4}G_4(\tau)$, $g_3 = 140\pi^{-6}G_6(\tau)$, and G_4, G_6 are the usual Eisenstein series

$$
G_4(\tau) = \sum_{m,n \in \mathbb{Z}_{\neq 0}} \frac{1}{(m\tau + n)^4},
$$
$$
G_6(\tau) = \sum_{m,n \in \mathbb{Z}_{\neq 0}} \frac{1}{(m\tau + n)^6} \tag{16.6}
$$

which define modular forms for $SL(2, \mathbb{Z})$ of weight 4 and 6, respectively. Since the definition of x_0 and y_0 was such that $y_0 = dx_0/dz$, one also can rewrite $\omega_0 = dz$ as

$$
\omega_0 = \frac{dx_0}{y_0}. \tag{16.7}
$$

Now, set $x = x_0 u$, $y = \frac{1}{2}y_0 u^{3/2}$, and

$$
\omega = \begin{cases} \dfrac{\sqrt{2/u}}{4\pi}\omega_0 = \dfrac{\sqrt{2}}{8\pi}\dfrac{dx}{y} & \text{for } N_f = 4 \\[2mm] \dfrac{\sqrt{2/u}}{2\pi}\omega_0 = \dfrac{\sqrt{2}}{4\pi}\dfrac{dx}{y} & \text{for } N = 4. \end{cases} \tag{16.8}
$$

The equation for the curve becomes

$$y^2 = x^3 - \frac{1}{4} g_2(\tau) x u^2 - \frac{1}{4} g_3(\tau) u^3. \tag{16.9}$$

This change of variables and in particular the normalization of u is motivated by the following requirement. For weak coupling ($\tau \to i\infty$) we should recover our curve $y^2 = F_0(x, u) = x^2(x - u)$. It is easy to check from the asymptotic behavior $g_2 = \frac{4}{3} + \mathcal{O}(q)$, $g_3 = \frac{8}{27} + \mathcal{O}(q)$ that after replacing x in (16.9) by $x - u/3$ we find F_0.

The periods of ω are now $\frac{1}{4} \sqrt{2/u}(1, \tau)$ for $N_f = 4$ and $\frac{1}{2} \sqrt{2/u}(1, \tau)$ for $N = 4$ and (since in general the periods of ω are da/du and da_D/du), one has

$$a = \begin{cases} \frac{1}{2} \sqrt{2u} & \text{for } N_f = 4 \\ \sqrt{2u} & \text{for } N = 4 \end{cases}$$
$$a_D = \tau a, \tag{16.10}$$

as desired.

Thus, we have determined the appropriate curve for the massless theory. The coefficients are modular forms. This is not really a new test of S-duality; it is equivalent to the fact that the metric of the classical theory is S-dual, which is one of the traditional pieces of evidence for S-duality.

Spin structures

Eq. (16.9) for the curve can be factored

$$y^2 = (x - e_1(\tau) u)(x - e_2(\tau) u)(x - e_3(\tau) u), \tag{16.11}$$

with the e_i being the roots of the cubic polynomial $4x^3 - g_2 x - g_3$; they obey $e_1 + e_2 + e_3 = 0$. The classical formulas for the e_i (Ref. [27], p. 69; note we set $\omega_1 = \pi$) are

$$e_1 - e_2 = \theta_3^4(0, \tau),$$
$$e_3 - e_2 = \theta_1^4(0, \tau),$$
$$e_1 - e_3 = \theta_2^4(0, \tau), \tag{16.12}$$

where θ_i are the θ functions

$$\theta_1(0, \tau) = \sum_{n \in \mathbb{Z}} q^{(n+1/2)^2/2},$$
$$\theta_2(0, \tau) = \sum_{n \in \mathbb{Z}} (-1)^n q^{n^2/2},$$
$$\theta_3(0, \tau) = \sum_{n \in \mathbb{Z}} q^{n^2/2} \tag{16.13}$$

and hence

$$e_1 = \frac{2}{3} + 16q + 16q^2 + \mathcal{O}(q^3),$$
$$e_2 = -\frac{1}{3} - 8q^{1/2} - 8q - 32q^{3/2} - 8q^2 + \mathcal{O}(q^{5/2}),$$

$$e_3 = -\frac{1}{3} + 8q^{1/2} - 8q + 32q^{3/2} - 8q^2 + \mathcal{O}(q^{5/2}) \tag{16.14}$$

(note that unlike Ref. [27], we use $q = e^{2\pi i \tau}$). As g_2 and g_3 are modular forms of weight 4 and 6, the e_i are of weight 2 (indeed, the theta functions are of weight $1/2$). However, the e_i are not modular forms for $SL(2, \mathbb{Z})$ because there is no modular-invariant way to select a particular root of the cubic. Rather, the e_i are modular forms of three different (conjugate) subgroups of $SL(2, \mathbb{Z})$, each of index three.

Actually, there is a natural one-to-one association of the e_i with the even spin structures on E. This can be seen as follows, beginning with the description of E as the quotient of the z-plane by a lattice. The even spin structures of E are in natural correspondence with the non-zero half-lattice points $z = 1/2, \tau/2$, and $(\tau + 1)/2$. Since $\mathcal{P}(z) = \mathcal{P}(-z)$, the derivative $y = \mathcal{P}'(z)$ vanishes at these points (indeed, up to a lattice translation those points are invariant under $z \leftrightarrow -z$). So the half-lattice points are zeroes of y. Looking at Eq. (16.11), we see therefore that the half-lattice points have $x = e_i u$, for $i = 1, 2$, or 3. So the even spin structures (or half-lattice points) correspond to the e_i. (And each e_i is a modular form for the subgroup of $SL(2, \mathbb{Z})$ that fixes that spin structure – these subgroups are conjugate to $\Gamma_0(2)$, the subgroup of $SL(2, \mathbb{Z})$ obtained by requiring that the upper right entry be even.)

For the $N = 4$ theory, the relation of the e_i to the spin structures will have no particular importance. For $N_f = 4$, it is very important since, as we know already, $SL(2, \mathbb{Z})$ permutes the three eight-dimensional representations of $\mathrm{Spin}(8)$ in the same way that it acts on the spin structures. So the three e_i are permuted under $\mathrm{Spin}(8)$ triality.

16.2. The curve for N = 4

We recall that $N = 4$ supersymmetric Yang–Mills theory can be regarded as $N = 2$ super Yang–Mills theory with an additional matter field that is a hypermultiplet in the adjoint representation of the gauge group. One can give a bare mass m to that hypermultiplet, explicitly breaking $N = 4$ to $N = 2$. In this subsection we analyze the resulting theory, for gauge group $SU(2)$.

We first consider the theory for weak coupling, that is for $|q| \ll 1$, with $m \neq 0$. There is one singularity at $u \approx \frac{1}{4} m^2$ where a component H of the elementary hypermultiplet is massless. This gives a monodromy conjugate to T^2.[10] In addition, at an energy of order $\Lambda_0 \sim q^{1/4} m$, the theory evolves to a strongly coupled pure $N = 2$ gauge theory. As was explained in Ref. [9], this theory has two singularities, associated with massless monopoles, with monodromies conjugate to T^2. So altogether, we get three singularities, each conjugate to T^2.[11]

---

[10] The elementary hypermultiplet for $N = 4$ has twice the electric charge of the hyperdoublets that we have considered earlier. As a massless hyperdoublet gives monodromy T, and the one-loop beta function which determines the monodromy is proportional to the square of the charge, the massless H particle would give monodromy T^4 in the conventions of the $N_f = 4$ theory. With the $N = 4$ conventions, the monodromy is T^2.

[11] These three singularities are permuted under monodromies in q and m. This is the reason the $N = 4$ conventions in which they are all conjugate to T^2 are preferable to the $N_f \neq 0$ conventions in which one is

Of course, the above analysis was valid for very weak coupling. Could it be, for instance, that what we described above as one singularity conjugate to T^2 is really a pair of singularities conjugate to T, separated by an amount that vanishes for weak coupling? $SL(2, \mathbb{Z})$ group theory alone would permit this, but it is impossible because each of the singularities arises when a *single* hypermultiplet becomes massless.

So we are looking for a family of curves

$$y^2 = F(x, u) \tag{16.15}$$

(with cubic F) that – as u varies – has precisely three singularities each conjugate to T^2. The restriction that this places on F was explained in our discussion of the $N_f = 2$ theory in Section 13.1. There is a singularity at u_0 with monodromy T^n for $n > 1$ (generically n will be 2) if and only if for some x_0,

$$F = \frac{\partial F}{\partial x} = \frac{\partial F}{\partial u} = 0 \tag{16.16}$$

at $x = x_0$, $u = u_0$. Conditions (16.16) mean that the curve $F(x, u) = 0$ has a singularity at $(x, u) = (x_0, u_0)$.

Therefore, we are looking for a plane cubic curve $F(x, u) = 0$ with three distinct singularities. The possible singularities of a plane cubic curve can be completely classified. If F is an irreducible polynomial, there is at most one singularity (a node or cusp). If $F = F_1 F_2$, with F_1 linear in x and u and F_2 quadratic and irreducible, there are precisely two singularities (perhaps at infinity), namely the points where $F_1 = F_2 = 0$. The only way to get three singularities is to have $F = F_1 F_2 F_3$, where the three factors are linear; the three singularities are the points $F_i = F_j = 0$ for any two distinct i and j.

To reproduce the known $m = 0$ limit of F, the F_i must be (up to a scalar multiple and a permutation of i) $F_i = x - e_i u - f_i$ where f_i are functions of m and τ only and vanish at $m = 0$. Moreover, by adding constants (that is, functions of m and τ only) to x and u, one can eliminate two of the three f_i. Since we did not assign any physical meaning to x we can take the freedom to shift it. However, we want to preserve $u = \langle \mathrm{Tr}\, \phi^2 \rangle$. Therefore, we will denote the shifted u by $\tilde{u}$ and will later determine the relation between them. To keep the symmetry under permuting the e_i, we shift x and u such that $f_i = e_i^2 f$. Then the equation of the mass-deformed $N = 4$ theory becomes

$$y^2 = (x - e_1 \tilde{u} - e_1^2 f)(x - e_2 \tilde{u} - e_2^2 f)(x - e_3 \tilde{u} - e_3^2 f). \tag{16.17}$$

We still need to find f and to determine $\tilde{u}$ in terms of u.

The relation between u and $\tilde{u}$ is determined by examining the theory at weak coupling; i.e. in the limit $\tau \to i\infty$. In this limit we should reproduce our weak coupling curve $y^2 = F_0 = x^2(x - u)$. This motivates us to change variables to

$$\tilde{u} = u - \tfrac{1}{2} e_1 f,$$
$$x \to x - \tfrac{1}{2} e_1 u + \tfrac{1}{2} e_1^2 f \tag{16.18}$$

---

conjugate to T^4 and the others to T.

in (16.17). The family of curves becomes

$$y^2 = (x - c_1 u)(x + c_2 u - c_2(c_1 + c_2)f)(x - c_2 u + c_2(c_1 - c_2)f) \qquad (16.19)$$

with $c_1 = \frac{3}{2} e_1$ and $c_2 = \frac{1}{2}(e_3 - e_2)$. In this form it is easy to study the weak coupling limit. For a smooth limit, $f_0 = f(\tau = i\infty)$ should be finite. Using $c_1(\tau = i\infty) = 1$ and $c_2(\tau = i\infty) = 0$, the exact curve (16.19) becomes $y^2 = F_0 = x^2(x - u)$ as required. Therefore, in the form (16.19) the family of curves is expressed in terms of $u = \langle \text{Tr} \, \phi^2 \rangle$.

We can now relate $f_0 = f(\tau = i\infty)$ to the mass m. We do that by examining the singularities of (16.19). The roots of the equation are at $x_1 = c_1 u$ and $x_{2,3} = \pm c_2(-u + (c_1 \pm c_2)f)$. A singularity occurs when $x_i = x_j$ (for any two distinct i and j). This occurs for

$$u_1 = \frac{3}{2} e_1 f = c_1 f,$$
$$u_{2,3} = \pm \frac{1}{2}(e_3 - e_2)f = \pm c_2 f. \qquad (16.20)$$

In the weak coupling limit $c_1 \approx 1$, $c_2 \approx 8q^{1/2}$ and hence $u_1 \approx f_0$ and $u_{2,3} \approx \pm 8q^{1/2} f_0$. We interpret the singularity at u_1 as associated with a massless elementary field. It should be at $a = m/\sqrt{2}$. For weak coupling and in the $N = 4$ normalization this is at $u \approx \frac{1}{2} a^2 = m^2/4$. Therefore,

$$f_0 = f(\tau = i\infty) = m^2/4. \qquad (16.21)$$

The other two singularities at $u_{2,3} \approx \pm 8q^{1/2} f_0$ are interpreted as the two singularities of the low energy pure gauge $N = 2$ theory. More precisely, we can now take the scaling limit $q \to 0$, $f_0 = m^2/4 \to \infty$ holding $\Lambda_0^2 = 2q^{1/2} m^2$ fixed. In this limit (16.19) becomes

$$y^2 = (x - u)(x^2 - \Lambda_0^4) \qquad (16.22)$$

which is exactly the curve of the expected low energy pure gauge $N = 2$ theory in the $N = 4$ conventions with scale Λ_0.

We still need to determine $f(\tau)$. To do that we will consider the residues of the differential form λ. As explained in Section 15 (see the footnote preceding Eq. (15.4)), they are independent of u and hence, on dimensional grounds, are proportional to m. The proportionality factor must be independent of τ, since the residues are related to the central extension in the $N = 2$ algebra, which is independent of τ. We will explain the general theory of these residues in Section 17, but for the moment, we give instead the following indirect argument which shows that f is independent of τ.

Let us assume first that f is independent of τ. Then, Eq. (16.17) has simple modular properties. y, x, $\tilde{u}$, e_i and f have modular weights 6, 4, 2, 2 and 0 respectively. Therefore the differential form λ determined by $d\lambda/du \sim dx/y$ transforms like a differential form of weight zero under $SL(2, \mathbb{Z})$ and the same is true of its residues. We also know that the residues do not diverge for $\tau \to i\infty$ (since in (16.21) we showed that at any rate the assumption that f is constant is valid for $\tau \to i\infty$). A modular function of weight zero that is bounded at infinity is a constant, so if f is constant, the residues are constants.

It is now trivial to determine the residues for arbitrary f. Indeed, on dimensional grounds, the residues are proportional to m, but m only enters through the function

$f(\tau, m) = m^2 f_1(\tau)$. Hence, the residues are a constant times $m\sqrt{f_1(\tau)}$. The residues, therefore, are independent of τ if and only if f_1 is a constant. The constant is known since we have determined the behavior for $\tau \to i\infty$:

$$f = \frac{1}{4}m^2. \tag{16.23}$$

We conclude that the curve governing the low energy behavior of the mass-deformed $N = 4$ theory is

$$y^2 = (x - e_1\widetilde{u} - \frac{1}{4}e_1{}^2 m^2)(x - e_2\widetilde{u} - \frac{1}{4}e_2{}^2 m^2)(x - e_3\widetilde{u} - \frac{1}{4}e_3{}^2 m^2) \tag{16.24}$$

with

$$u = \langle \mathrm{Tr}\, \phi^2 \rangle = \widetilde{u} + \frac{1}{8}e_1 m^2. \tag{16.25}$$

This formula is completely $\mathrm{SL}(2, \mathbb{Z})$ invariant; the coefficients are modular forms. Since the formula is not limited to weak coupling, this $\mathrm{SL}(2, \mathbb{Z})$ invariance is a genuine, new, strong coupling test of $\mathrm{SL}(2, \mathbb{Z})$ invariance of the $N = 4$ theory. (Also, we learn that the $N = 2$-invariant bare mass preserves $\mathrm{SL}(2, \mathbb{Z})$.)

$\mathrm{SL}(2, \mathbb{Z})$ invariance may, however, be lost in various weak coupling limits. For instance, for the weak coupling limit $\tau \to i\infty$ the natural variable is $u = \langle \mathrm{Tr}\, \phi^2 \rangle$ which differs from $\widetilde{u}$ by an additive renormalization (16.25). Unlike $\widetilde{u}$, u does not transform like a modular form. Furthermore, in the $\tau \to i\infty$ limit we defined a scaling theory by taking also $m \to \infty$ holding $\Lambda_0^4 = 4m^4 e^{2\pi i\tau}$ fixed. By $\mathrm{SL}(2, \mathbb{Z})$ we can find other weakly coupled theories and scaling limits in which $\tau \to p/q$ with p/q an arbitrary rational number. The theory in these limits is strongly coupled in the original variables – the elementary gauge fields – but weakly coupled in dual variables, ϕ_d, which were interpreted as monopoles in the original theory. The natural parameter is then $u_d = \langle \mathrm{Tr}\, \phi_d{}^2 \rangle$. It is related to u or $\widetilde{u}$ by a shift and a modular transformation. In a suitable limit with $\tau \to p/q$, $m \to \infty$ one gets a pure gauge $N = 2$ theory. Thus, the pure $N = 2$ theory that arises in the scaling limits does not have $\mathrm{SL}(2, \mathbb{Z})$ symmetry; $\mathrm{SL}(2, \mathbb{Z})$ merely permutes the possible equivalent scaling limits.

16.3. The curve for $N_f = 4$

In this subsection we determine the curve for the $N_f = 4$ theory with arbitrary masses (m_1, m_2, m_3, m_4). Among other things we will establish triality and $\mathrm{SL}(2, \mathbb{Z})$ invariance.

We start by considering the situation for $m_i = (m, m, 0, 0)$. As we discussed in Section 10, in this case we expect to find three singularities with two massless particles in each. According to our general discussion in Section 11, this means that the monodromy around any of them is conjugate to T^2. This is exactly the situation we encountered in our discussion of the curve for $N = 4$, so with these masses the curve is

$$y^2 = \prod_i (x - e_i\widetilde{u} - e_i^2 f). \tag{16.26}$$

Here $\tilde{u}$ is related to $u = \langle \mathrm{Tr}\, \phi^2 \rangle$ by a constant shift as in (16.18) and f is proportional to m^2 and *a priori* may depend on τ.

As for $N = 4$, since the residues of λ should be τ independent, f must be a constant. In order to determine the constant we again consider this curve in the weak coupling limit $\tau \to i\infty$. Again, we shift x and $\tilde{u}$ as in (16.18) and take the limit $q \to 0$, $f \to \infty$ with qf^2 held fixed. The low energy theory is that of the massless $N_f = 2$ theory. Indeed, we find the curve (13.3) with $\tilde{\Lambda}_2{}^2 = \Lambda_2{}^2/8 = 8q^{1/2}f$. We also find a singularity at $u \approx f$ which we interpret as associated with a massless elementary quark and hence it should be at $a = m/\sqrt{2}$. In the $N_f = 4$ normalization this means that it is at $u \approx 2a^2 = m^2$. Hence, $f = m^2$.

We now turn to the theory with arbitrary m_i. We first impose the global SO(8) symmetry and construct its low dimension invariants. There is a unique quadratic invariant

$$R = \frac{1}{2} \sum_i m_i^2. \tag{16.27}$$

There are three linearly independent quartic invariants. We take them to be R^2 and

$$
\begin{aligned}
T_1 &= \frac{1}{12} \sum_{i>j} m_i^2 m_j^2 - \frac{1}{24} \sum_i m_i^4, \\
T_2 &= -\frac{1}{2} \prod_i m_i - \frac{1}{24} \sum_{i>j} m_i^2 m_j^2 + \frac{1}{48} \sum_i m_i^4, \\
T_3 &= \frac{1}{2} \prod_i m_i - \frac{1}{24} \sum_{i>j} m_i^2 m_j^2 + \frac{1}{48} \sum_i m_i^4
\end{aligned}
\tag{16.28}
$$

with $\sum T_i = 0$. The reason for writing them like that is that the T_i are permuted under the triality automorphism of SO(8) (which acts on the masses as in Eqs. (10.3) and (10.4)); we anticipate (but do not assume) that triality is a symmetry of the theory. There are four six order SO(8) invariants. We take them to be R^3, RT_i and

$$N = \frac{3}{16} \sum_{i>j>k} m_i^2 m_j^2 m_k^2 - \frac{1}{96} \sum_{i \neq j} m_i^2 m_j^4 + \frac{1}{96} \sum_i m_i^6. \tag{16.29}$$

R and N are invariant under triality.

In trying to generalize the curve for $m_i = (m, m, 0, 0)$ to arbitrary m we impose that:
(1) The limit of the curve as any masses go to zero is smooth and hence it is polynomial in m_i.
(2) $U(1)_{\mathcal{R}}$ (or equivalently dimensional analysis) constrains the powers of m_i. This is achieved by assigning charges $4, 4, 6, 2$ to $\tilde{u}, x, y, m_i$.
(3) For $m_i = (m, m, 0, 0)$ we recover the curve (16.26).

Since for $m_i = (m, m, 0, 0)$ we have $R \neq 0$, $T_i = N = 0$, the most general form of the curve consistent with these conditions is

$$y^2 = W_1 W_2 W_3 + x \sum_i T_i f_i + \tilde{u} \sum_i T_i g_i + R \sum_i T_i h_i + pN \tag{16.30}$$

with

$$W_i = x - e_i \widetilde{u} - e_i^2 R \tag{16.31}$$

and f_i, g_i, h_i and p are functions of τ to be determined.

For $m_i = (m, m, 0, 0)$ the curve has three singularities at $\widetilde{u}_i = e_i m^2$ with two massless particles in each. Therefore, the discriminant of the curve Δ has three double zeros at these values of $\widetilde{u}$. We now consider the situation with $m_i = (m + \mu, m - \mu, 0, 0)$. According to the discussion in Section 10, for non-zero μ one of the three singularities should split and the other two can move but remain double zeroes. Our weak coupling limit above identified the singularity at $\widetilde{u}_1$ with the one at which elementary quarks are massless. Therefore, this one should split. We will now determine some of the coefficients by demanding that the zero at $\widetilde{u}_2$ moves but remains a double zero.

Since the curve is holomorphic in μ^2, so is the discriminant Δ. For $|\mu| \ll |m|$ its double zero at $\widetilde{u}_2$ starts moving at order μ^2 to $\widetilde{u}_2 = e_2 m^2 + \mu^2 \lambda$. For it to remain a double zero, the order μ^2 term in $\Delta(\widetilde{u} = e_2 m^2 + \mu^2 \lambda)$ should vanish. It is straightforward to calculate this term. It is proportional to

$$-e_1 e_3 (f_2 + f_3 - 2f_1) + e_2 (g_2 + g_3 - 2g_1) + (h_2 + h_3 - 2h_1) + p. \tag{16.32}$$

We get one constraint by setting (16.32) to zero. Repeating this at $\widetilde{u}_3$ we get a similar equation with the subscripts 2 and 3 interchanged. Four more equations (related by other permutations of the subscripts) are obtained by studying the cases $m_i = (m, m, \mu, \mu)$, and $m_i = (m, m, \mu, -\mu)$. To organize the equations, we break the symmetry between $\widetilde{u}_2$ and $\widetilde{u}_3$ by deciding that in the first of these cases the zero at $\widetilde{u}_2$ splits and in the other the zero at $\widetilde{u}_3$ splits. The opposite choice leads to similar results with e_2 and e_3 interchanged. Using these six equations we determine the ten unknowns f_i, g_i, h_i, p in terms of four unknowns F, G, H, A:

$$
\begin{aligned}
f_1 &= A(e_2 - e_3) + F, \\
f_2 &= A(e_3 - e_1) + F, \\
f_3 &= A(e_1 - e_2) + F, \\
g_i &= -f_i e_i + G, \\
h_i &= -f_i e_i^2 + H, \\
p &= -(e_1 - e_2)(e_2 - e_3)(e_3 - e_1)A.
\end{aligned}
\tag{16.33}
$$

Since $\sum T_i = 0$, the values of F, G, H do not affect the curve and hence can be set to zero.

In order to determine the unknown function $A(\tau)$ we consider the μ dependence more fully. One approach would be to study higher order terms in the expansion around $\mu = 0$, and require that the double zeroes of Δ remain double zeroes. Instead, we examine the curve for $\mu = m$. For this value there are three massless quarks, and we expect the two unsplit double zeros to merge, giving Δ a fourth order zero (which we have previously encountered in the $N_f = 3$ theory). Δ has this fourth order zero if and only if

$$A = (e_1 - e_2)(e_2 - e_3)(e_3 - e_1) \tag{16.34}$$

leading to the curve

$$y^2 = W_1 W_2 W_3 + A(W_1 T_1(e_2 - e_3) + W_2 T_2(e_3 - e_1)$$
$$+ W_3 T_3(e_1 - e_2)) - A^2 N \tag{16.35}$$

again with

$$W_i = x - e_i \tilde{u} - e_i^2 R,$$
$$A = (e_1 - e_2)(e_2 - e_3)(e_3 - e_1),$$
$$R = \frac{1}{2} \sum_i m_i^2,$$
$$T_1 = \frac{1}{12} \sum_{i>j} m_i^2 m_j^2 - \frac{1}{24} \sum_i m_i^4,$$
$$T_2 = -\frac{1}{2} \prod_i m_i - \frac{1}{24} \sum_{i>j} m_i^2 m_j^2 + \frac{1}{48} \sum_i m_i^4,$$
$$T_3 = \frac{1}{2} \prod_i m_i - \frac{1}{24} \sum_{i>j} m_i^2 m_j^2 + \frac{1}{48} \sum_i m_i^4,$$
$$N = \frac{3}{16} \sum_{i>j>k} m_i^2 m_j^2 m_k^2 - \frac{1}{96} \sum_{i \neq j} m_i^2 m_j^4 + \frac{1}{96} \sum_i m_i^6. \tag{16.36}$$

We see that our final answer is modular invariant. To be precise, full $SL(2, \mathbb{Z})$ invariance, which permutes the e_i, is a symmetry if combined with Spin(8) triality, which permutes the T_i. This is a strong indication that the full theory is dual in the way described in Sections 6 and 10.

Weak coupling and scaling limits

As for $N = 4$, the theory has infinitely many weakly coupled limits related by $SL(2, \mathbb{Z})$. The obvious one is $\tau \to i\infty$; others are at $\tau = p/q$ with p/q rational. The weakly coupled variables are different in the various limits (the monopoles in one limit are the quarks in another limit). Correspondingly, $u = \langle \text{Tr} \, \phi^2 \rangle$ is different in the different scaling theories. Let us focus on $\tau \to i\infty$. We define u and shift x in a way similar to (16.18)

$$u = \tilde{u} + \frac{1}{2} e_1 R,$$
$$x \to x - \frac{1}{2} e_1 u + \frac{1}{2} e_1^2 R. \tag{16.37}$$

Substituting (16.37) in (16.35) we find

$$y^2 = (x^2 - c_2^2 u^2)(x - c_1 u) - c_2^2 (x - c_1 u)^2 \sum_i m_i^2 - c_2^2 (c_1^2 - c_2^2)(x - c_1 u) \sum_{i>j} m_i^2 m_j^2$$
$$+ 2c_2(c_1^2 - c_2^2)(c_1 x - c_2^2 u) m_1 m_2 m_3 m_4 - c_2^2 (c_1^2 - c_2^2)^2 \sum_{i>j>k} m_i^2 m_j^2 m_k^2 \tag{16.38}$$

where, as before, $c_1 = \frac{3}{2} e_1$ and $c_2 = \frac{1}{2}(e_3 - e_2)$.

We can now analyze the renormalization group flow from $N_f = 4$ to $N_f < 4$. In Sections 12–14 we have already verified the flows from $N_f = 3$ to $N_f = 0, 1, 2$, so it is sufficient here to consider the flow from $N_f = 4$ to the massive $N_f = 3$ theory. To do this, we take the limit $\tau \to i\infty$, $m_4 \to \infty$, keeping fixed m_1, m_2, m_3 and

$$\Lambda_3 = 64 q^{1/2} m_4. \tag{16.39}$$

(Recall that $q^{1/2}$ is the one instanton factor. The reason for the factor of 64 is that in Section 14 we took $\Lambda_2{}^2 = \Lambda_3 m_3$ but the above discussion of (16.26) implies that the flow from $N_f = 4$ to massless $N_f = 2$ gives $\Lambda_2{}^2 = 64 q^{1/2} m_3 m_4$. Thus the 64 in (16.39) is needed to agree with our previous definition of Λ_3.) Taking this limit using $c_1 \approx 1$ and $c_2 \approx 8 q^{1/2}$ leads to

$$\begin{aligned}
y^2 = x^2(x - u) &- \frac{1}{64}\Lambda_3{}^2(x - u)^2 - \frac{1}{64}(m_1{}^2 + m_2{}^2 + m_3{}^2)\Lambda_3{}^2(x - u) \\
&+ \frac{1}{4}m_1 m_2 m_3 \Lambda_3 x - \frac{1}{64}(m_1{}^2 m_2{}^2 + m_2{}^2 m_3{}^2 + m_1{}^2 m_3{}^2)\Lambda_3{}^2.
\end{aligned} \tag{16.40}$$

which is the same as the massive $N_f = 3$ curve (14.5).

As for the $N = 4$ theory, scaling limits around other weakly coupled points lead to equivalent theories in terms of dual degrees of freedom.

17. The theory of the residues

One important difference between the mathematical structure of the $N = 2$ theories with matter considered in this paper and the "pure gauge theory" studied in Ref. [9] is that in the presence of matter the monodromies do not simply transform (a_D, a) linearly, by $SL(2, \mathbb{Z})$ transformations; a and a_D also pick up additive constants under monodromy. As we explained in Section 15, these constants can be detected as the residues of the differential form λ. Since the jumps in a or a_D are integral linear combinations of $m_i/\sqrt{2}$ (with m_i the bare masses) and are $2\pi i$ times the residues of λ, the residues of λ should be of the form

$$\operatorname{Res}\lambda = \sum_i \frac{n_i m_i}{2\pi i \sqrt{2}}, \quad \text{with} \quad n_i \in \mathbb{Z}. \tag{17.1}$$

We have not so far verified or exploited this condition in full. It was verified in Section 15 for the massive $N_f = 2$ theory and exploited only in a very limited way in Section 16 for $N = 4$ and $N_f = 4$. The reason that we have not yet used the full force of the residue condition is that in fact, except in special cases in which λ can be found by inspection, implementing this condition requires a fairly elaborate machinery. This machinery will be developed in the present section and used to give a new derivation of the curves for $N_f = 4$ and $N = 4$ (the others can be obtained, of course, by renormalization group flow). Since the new derivation does in fact give results that agree with what we obtained previously, this will also show that the previously obtained curves do have residues of the right form.

In general, in this paper up to the present point we have followed a scenic route, starting with simple cases (the $N_f = 0$ theory), gradually adding and understanding new ingredients, climbing upstream to larger N_f (against the renormalization group current that flows to smaller N_f), and finally understanding what from this point of view are the most challenging cases of $N_f = 4$ and $N = 4$. The analysis of the residues presented in the present section has the opposite flavor; after building up the necessary apparatus, the machinery is easily applied directly to $N_f = 4$ and $N = 4$, and gives the answer after a very short calculation. This approach certainly provides additional insight into some questions like why triality holds for $N_f = 4$ and may also be useful in generalizing to gauge groups other than SU(2).

17.1. The meaning of the residues

The $N_f = 4$ theory is controlled by a curve $y^2 = F(x, u, m_i, \tau)$ and a differential form λ obeying

$$\frac{d\lambda}{du} = \omega + \text{exact form in } x \tag{17.2}$$

with

$$\omega = \frac{\sqrt{2}}{8\pi} \frac{dx}{y}. \tag{17.3}$$

For $N = 4$ the structure is the same, except that 8π is replaced by 4π. F should be such that the residues of λ are linear in the quark bare masses. This is a severe restriction on F; we will see that it determines F uniquely (up to the usual changes of variables) independently of most of the arguments that we have used up to this point.

As a preliminary, let us write (17.2) in a more symmetrical form. If $\lambda = dx\, a(x, u)$, then (17.2) means

$$\frac{\sqrt{2}}{8\pi} \frac{dx}{y} = dx\, \frac{\partial a}{\partial u} + dx\, \frac{\partial}{\partial x} f(x, u); \tag{17.4}$$

the arbitrary total x-derivative $dx\, \partial f/\partial x$ is allowed because it does not contribute to the periods. Eq. (17.4) can be understood much better if written symmetrically in x and u. Henceforth, instead of using a one-form $\omega = (\sqrt{2}/8\pi) \cdot dx/y$, we will use a two-form

$$\omega = \frac{\sqrt{2}}{8\pi} \frac{dx\, du}{y}. \tag{17.5}$$

Similarly, we combine the functions a, f appearing in (17.4) into a one-form $\lambda = -a(x, u)dx + f(x, u)du$. The change in notation for ω and λ should cause no confusion. Then Eq. (17.4) can be more elegantly written

$$\omega = d\lambda. \tag{17.6}$$

The meaning of the problem of finding λ can now be stated. Let X be the (noncompact) complex surface defined by the equation $y^2 = F(x, u)$ (we suppress the parameters

m_i and τ). Being closed, ω defines an element $[\omega] \in H^2(X, \mathbb{C})$. A smooth differential λ obeying (17.6) exists if and only if $[\omega] = 0$. Moreover, by standard theorems, in the absence of restrictions on the growth of λ at infinity, if λ exists it can be chosen to be holomorphic and of type $(1, 0)$.

If on the other hand $[\omega] \neq 0$, then (17.6) has no smooth, much less holomorphic, solution. However, X has the property that if one throws away a sufficient number of complex curves C_a, then $X' = X - \cup_a C_a$ has $H^2(X', \mathbb{C}) = 0$. (The necessary C_a are explicitly described later.) So if we restrict to X', the cohomology class of ω vanishes and λ exists. λ may however have poles on the C_a, perhaps with residues, which we call $\text{Res}_{C_a}(\lambda)$.[12] If λ does have residues, then $d\lambda$ contains delta functions, and if one works on X instead of X', one really has not (17.6) but

$$\omega = d\lambda - 2\pi i \sum_a \text{Res}_{C_a}(\lambda) \cdot [C_a] \tag{17.7}$$

where $[C_a]$ (which represents the cohomology class known as the Poincaré dual of C_a) is a delta function supported on C_a.

In cohomology, (17.7) simply means

$$[\omega] = -2\pi i \sum_a \text{Res}_{C_a}(\lambda) \cdot [C_a]. \tag{17.8}$$

Thus, if we pick the C_a so that the $[C_a]$ are a basis of $H^2(X, \mathbb{C})$, then the residues $\text{Res}_{C_a}(\lambda)$ are simply the coefficients of the expansion of $[\omega]$ in terms of the $[C_a]$. To find the residues we need not actually find λ; it suffices to understand the cohomology class of ω by any method that may be available.

For instance, if X were compact, we could proceed as follows. First compute the intersection matrix

$$M_{ab} = \#(C_a \cdot C_b) \tag{17.9}$$

(that is, the number of intersection points of C_a and C_b, after perhaps perturbing the C_a so that they intersect generically). This is an invertible matrix. Second, calculate the periods

$$c_a = \int_{C_a} \omega. \tag{17.10}$$

Then

$$[\omega] = \sum_{a,b} c_a M^{-1}{}_{ab} [C_b]. \tag{17.11}$$

Comparing to (17.8), we get

$$\text{Res}_{C_a}(\lambda) = -\frac{1}{2\pi i} \sum_b M^{-1}_{ab} c_b. \tag{17.12}$$

---

[12] The residues of λ along C_a are constants, since $d\,\text{Res}_{C_a}(\lambda) = \text{Res}_{C_a} d\lambda = \text{Res}_{C_a}\omega = 0$.

We will eventually follow that strategy after compactifying X and modifying ω so as to have no pole at infinity. With this in view, we will somewhat loosely and prematurely call the c_b the "periods" of ω.

17.2. The cohomology of X

It will be useful to know something about the cohomology of the complex manifold X described by the equation

$$y^2 = F(x,u) = (x - e_1 u)(x - e_2 u)(x - e_3 u) + \text{lower order terms}. \qquad (17.13)$$

It is helpful to compactify X, which will be needed anyway to do the calculation just mentioned. We do this by introducing another variable v and making the equation homogeneous. First we extend F to a polynomial $F(x,u,v)$ homogeneous of degree 3 (such that $F(x,u,1)$ is the original F). We could now consider the homogeneous equation $vy^2 = F(x,u,v)$, with x,u,v,y all of degree 1. However, things work out more easily if we instead take x,u,v,y to be of degree $1,1,1,2$; so we study the homogenous equation

$$y^2 = v F(x,u,v). \qquad (17.14)$$

It is helpful to first look at a more general equation

$$y^2 = G(x,u,v) \qquad (17.15)$$

with a generic, irreducible G homogeneous of degree 4. The variety Z defined by this equation has the following properties. There is a $\mathbb{Z}_2$ symmetry $\alpha : y \rightarrow -y$. The differential form of interest, $\omega = \mathrm{d}x\,\mathrm{d}u/y$, is odd under the symmetry, so we are mainly interested in the part of the cohomology of Z that is odd. Using methods familiar to physicists from the study of Calabi–Yau manifolds,[13] one can show that the odd part of $H^2(Z)$ is seven dimensional.

The situation is somewhat different for the special quartic polynomial $G(x,u,v) = v F(x,u,v)$, since the manifold $\overline{X}$ defined by (17.13) has (for generic F) three singularities, at $y = v = x - e_i u = 0$, for $i = 1, 2$, or 3. To understand the structure of the singularities, set $u = 1$ by scaling, and set $w = (x - e_i u)(e_1 - e_2)(e_1 - e_3)$. The behavior near the singularity is then

$$y^2 = vw + \text{higher order terms}. \qquad (17.16)$$

This is known as an A_1 singularity. It is actually a $\mathbb{Z}_2$ orbifold singularity,[14] which is a harmless kind of singularity for our purposes; for instance, there is no special subtlety in describing $H^2(\overline{X}, \mathbb{C})$ by differential forms.

---

[13] See Ref. [28] for an introduction to the requisite methods. Z has $b_1 = b_3 = 0$ by the Lefschetz hyperplane theorem. Its Euler characteristic is 10 so $b_2 = 8$. The α-invariant part of the cohomology of Z can be computed from α-invariant differential forms, so coincides with the cohomology of $\mathbf{CP}^2$. Hence, the part of $H^2(Z)$ that is even under α is one dimensional, and the odd part is seven dimensional.

[14] This was noted in Section 3 where we encountered the same type of singularity in a different way.

However, when a complex surface develops an A_1 singularity, the second Betti number drops by 1. Since the odd part of $H^2(Z)$ is seven dimensional, and $\overline{X}$ is a specialization of Z to a case with three A_1 singularities, the odd part of $H^2(\overline{X})$ is four dimensional. [15] As we will see, this occurrence of the number four is no coincidence: it is related to the fact that conformal invariance with matter hyperdoublets requires $N_f = 4$.

Extending ω

Let us discuss the behavior of ω under compactification. The homogeneous version of ω is

$$\omega = \frac{v\,\mathrm{d}x\,\mathrm{d}u + x\,\mathrm{d}u\,\mathrm{d}v + u\,\mathrm{d}v\,\mathrm{d}x}{vy}. \tag{17.17}$$

The point of this formula is that (i) it reduces to the old one if we set $v = 1$; (ii) it is invariant under scaling of the homogeneous coordinates

$$\begin{aligned}
\delta x &= \epsilon x\,, \\
\delta u &= \epsilon u\,, \\
\delta v &= \epsilon v\,, \\
\delta y &= 2\epsilon y;
\end{aligned} \tag{17.18}$$

(iii) it vanishes if contracted with the vector field in (17.18) so it can be interpreted as a pull-back from the weighted projective space of (x, u, v, y).

"Infinity" in $\overline{X}$ is just the region with $v = 0$ (which one misses if one sets $v = 1$). It is evident in (17.17) that ω has a pole at $v = 0$. The equation $y^2 = vF(x, u, v)$ shows that near $v = 0$ on the double cover, $v \sim y^2$, so y is the good coordinate near $v = 0$. As $\mathrm{d}v/v \sim 2\,\mathrm{d}y/y$, ω looks near $y = 0$, in, say, a coordinate system with $u = 1$, like

$$\omega \sim \frac{\mathrm{d}y}{y^2}\,\mathrm{d}x. \tag{17.19}$$

Thus, there is a pole at $y = 0$, but the residue vanishes. Because the residue vanishes, ω can be interpreted as a cohomology class not just on X but on $\overline{X}$.

One could modify ω near infinity, preserving the fact that it is closed, but losing the fact that it is holomorphic and of type $(2, 0)$, so that ω extends over infinity as a closed two-form and so defines a cohomology class of $\overline{X}$. The ability to so interpret ω makes it possible to calculate using intersections and periods, as we will.

17.3. Finding the curves

Now we would like to find a suitable set of curves C_a on which λ will have poles. In this section we set $v = 1$ and work on the uncompactified manifold X given by $y^2 = F(x, u)$.

---

[15] It is the odd part whose dimension drops when the singularity develops, since the even part, which is generated by the Kähler class, certainly survives.

To guess what the C_a may be, let us look back to some of the models treated earlier, for instance the massive $N_f = 2$ theory. In Eq. (15.3), we found an explicit formula for λ in this theory; it had poles at

$$x = \pm \frac{1}{8} \Lambda_2{}^2,$$

$$y = \pm \frac{i}{8} \Lambda_2{}^2 (m_1 \pm m_2). \tag{17.20}$$

This can be interpreted as follows. The equation $y^2 = F(x, u)$ describes a double cover of the x–u plane. The equation

$$x = \pm \Lambda_2{}^2/8 \tag{17.21}$$

describes (for a given choice of the sign) a line in that plane. A generic line L in the x–u plane would be described by

$$x = \beta u + \theta, \tag{17.22}$$

for some β, θ. The double cover $y^2 = F(x, u)$, restricted to the line L, would be given by an equation

$$y^2 = g(u) \tag{17.23}$$

where generically $g(u)$ is not the square of a polynomial. That being so, the double cover of L is generically an irreducible complex curve, which is obviously invariant under $\alpha : y \rightarrow -y$. That is not what we want, because ω is *odd* under α; we need cycles on X that are odd. For the special case that the parameters in (17.22) are as in (17.21), the double cover (17.23) reduces to $y^2 = -(m_1 \pm m_2)^2 \Lambda_2{}^4/64$. Here the right-hand side is the square of a polynomial (in fact a constant), and so the double cover consists of two branches, with $y = i(m_1 \pm m_2) \Lambda_2{}^2/8$ or $y = -i(m_1 \pm m_2) \Lambda_2{}^2/8$. If we call the two branches $D_{+,\pm}$ and $D-, \pm$ (the second subscript denotes the sign for x in (17.21), and the first is the sign of y), then the differences $C_\pm = D_{+,\pm} - D_{-,\pm}$ are divisors that are odd under α. The explicit determination of the residue of λ in Section 15 amounted to expressing the cohomology class $[\omega]$ as a linear combination of C_+ and C_-.

We can imitate this in the case at hand

$$y^2 = (x - e_1 u)(x - e_2 u)(x - e_3 u) + F_1(x, u) \tag{17.24}$$

(where $F_1(x, u)$ is of degree $\leqslant 2$ in x and u). We introduce the line L of (17.22), and consider again the double cover of L deduced from (17.24). This is described by an equation $y^2 = g(u)$ where for generic u, $g(u)$ is cubic. If g is cubic, it cannot be a square, so $y^2 = g(u)$ describes an irreducible cover of L, necessarily invariant under $y \leftrightarrow -y$.

Therefore, we must adjust the coefficients in (17.22) to kill the cubic term in $g(u)$. This requires that we set $\beta = e_i$ for some i. Then g is quadratic in u and the equation for the double cover is of the general form

$$y^2 = Au^2 + Bu + C. \tag{17.25}$$

The polynomial on the right-hand side of (17.25) is a square when and only when the discriminant vanishes,

$$0 = \Delta = B^2 - 4AC, \tag{17.26}$$

When $\Delta = 0$, (17.25) can be written in the form $y^2 = A(u - t)^2$ for some t, and its solutions consist of two branches $y = \pm\sqrt{A}(u - t)$. If we call these branches D_+ and D_-, we get a divisor $C = D_+ - D_-$ that is odd under α.

Now actually, A is linear in θ, B is quadratic, and C is cubic. So the discriminant is quartic in θ and has four zeroes θ_a. For each we get a line L_a whose double cover has two components $D_{\pm,a}$. So we get four odd divisors $C_a = D_{+,a} - D_{-,a}$. As we discussed in the last subsection, four is the dimension of the odd part of the cohomology of X, so there are as many C_a as we would need for a basis of that odd part. We will see that they are a basis.

But actually, the above construction began by setting $\beta = e_i$ for some fixed i. We could have carried out the above steps for any $i = 1, 2$, or 3. So making the dependence on i explicit, we have divisors $D^{(i)}_{\pm,a}$, and α-odd divisors $C^{(i)}_a$. Of course, Spin(8) triality permutes the (i) superscript.

Intersection pairings

We would like to prove that the $C^{(i)}_a$ of fixed i do furnish a basis of the odd part of the cohomology. Since there are four of them, it suffices to prove that they are linearly independent; for this purpose, it is enough to work on $\overline{X}$ (where the intersection pairings are topological invariants) and prove that the matrix of intersection pairings is non-degenerate.

So we have to calculate $C^{(i)}_a \cap C^{(i)}_b$. For $a \neq b$, this vanishes, for the following reason. The divisors $D^{(i)}_{\pm,a}$ are given by

$$\begin{aligned} x &= e_i u + \theta_a v, \\ y &= \pm\sqrt{A}(u - t_a v)v. \end{aligned} \tag{17.27}$$

(Since we are working on $\overline{X}$, we have restored v. In doing this, we made the equations homogeneous, remembering that x, u, v, y have degree $1, 1, 1, 2$.) Any two of these curves (for distinct a and b, but regardless of the independent choices of $\pm$ signs) meet precisely at the $\mathbb{Z}_2$ orbifold point $x - e_1 u = y = v = 0$, so the pairings of these curves with distinct a and b are

$$D^{(i)}_{\pm,a} \cap D^{(i)}_{\pm,b} = \frac{1}{2} \tag{17.28}$$

where the two $\pm$ signs are chosen independently. (The $1/2$ comes because an intersection at the orbifold point is counted with weight $1/2$.) Recalling that $C^{(i)}_a = D^{(i)}_{+,a} - D^{(i)}_{-,a}$ we get $C^{(i)}_a \cap C^{(i)}_b = 0$ for $a \neq b$. However, $C^{(i)}_a \cap C^{(i)}_a = -4$.[16] Putting these results together,

$$C^{(i)}_a \cap C^{(i)}_b = -4\delta_{ab}. \tag{17.29}$$

---

[16] The value is obviously independent of i and a. That it is -4 requires a slightly more detailed analysis that we omit. The value -4 in fact follows from our determination below that $C^{(i)}_a \cap C^{(j)}_b = \pm 2$ for $i \neq j$.

This matrix is in particular nondegenerate, showing that for fixed i, the $C^{(i)}{}_a$ give a basis of the relevant piece of the cohomology.

Triality

This result may seem esoteric, but by extending it a bit, we will see Spin(8) triality at work in the classical geometry. To this end, we want to calculate the intersections $C^{(i)}{}_a \cap C^{(j)}{}_b$ for $i \neq j$. We claim that

$$C^{(i)}{}_a \cap C^{(j)}{}_b = \pm 2, \quad \text{for } i \neq j. \tag{17.30}$$

Granted this, let us see how the statement is connected with Spin(8) triality.

Since the $C^{(i)}{}_a$ of fixed i are a basis of the odd part of the cohomology, the $C^{(j)}{}_a$ can be expanded as linear combinations of the $C^{(i)}{}_a$. In fact, we will expand everything in terms of the $C^{(1)}{}_a$. Comparing (17.30) to (17.29), we see that the expansion coefficients are all $\pm 1/2$. In the above we adopted no particular convention as to what was $D^{(i)}{}_{+,a}$ and what was $D^{(i)}{}_{-,a}$, so we have not fixed the signs of the $C^{(i)}{}_a$. Picking an arbitrary sign for $C^{(2)}{}_1$, we can fix the signs of all the $C^{(1)}{}_a$ by requiring that the expansion of $C^{(2)}{}_1$ has all plus signs:

$$C^{(2)}{}_1 = \frac{1}{2} \left(C^{(1)}{}_1 + C^{(1)}{}_2 + C^{(1)}{}_3 + C^{(1)}{}_4 \right). \tag{17.31}$$

Now, we consider $C^{(2)}{}_a$ with $a > 1$. We can carry out a relabeling of the a index, since its meaning has not been fixed in any special way. The $C^{(2)}{}_a$, $a > 1$ are orthogonal to $C^{(2)}{}_1$, and have expansion coefficients $\pm 1/2$ in terms of the $C^{(1)}{}_a$. These conditions are enough to ensure that, up to permutations on the a index and changes in sign of $C^{(2)}{}_a$,

$$C^{(2)}{}_2 = \frac{1}{2} \left(C^{(1)}{}_1 + C^{(1)}{}_2 - C^{(1)}{}_3 - C^{(1)}{}_4 \right),$$
$$C^{(2)}{}_3 = \frac{1}{2} \left(C^{(1)}{}_1 - C^{(1)}{}_2 + C^{(1)}{}_3 - C^{(1)}{}_4 \right),$$
$$C^{(2)}{}_4 = \frac{1}{2} \left(C^{(1)}{}_1 - C^{(1)}{}_2 - C^{(1)}{}_3 + C^{(1)}{}_4 \right). \tag{17.32}$$

It remains to consider the $C^{(3)}{}_a$. They must have expansion coefficients of $\pm 1/2$ in terms of either the $C^{(1)}{}_a$ or the $C^{(2)}{}_a$. This ensures that, up to permutations and sign changes of the $C^{(3)}{}_a$, they are given by

$$C^{(3)}{}_1 = \frac{1}{2} \left(C^{(1)}{}_1 + C^{(1)}{}_2 + C^{(1)}{}_3 - C^{(1)}{}_4 \right),$$
$$C^{(3)}{}_2 = \frac{1}{2} \left(C^{(1)}{}_1 + C^{(1)}{}_2 - C^{(1)}{}_3 + C^{(1)}{}_4 \right),$$
$$C^{(3)}{}_3 = \frac{1}{2} \left(C^{(1)}{}_1 - C^{(1)}{}_2 + C^{(1)}{}_3 + C^{(1)}{}_4 \right),$$
$$C^{(3)}{}_4 = \frac{1}{2} \left(-C^{(1)}{}_1 + C^{(1)}{}_2 + C^{(1)}{}_3 + C^{(1)}{}_4 \right). \tag{17.33}$$

Let us compare this to Spin(8) triality. The $N_f = 4$ theory has four masses $m_1, m_2, m_3,$ m_4 which are the "eigenvalues" of a mass matrix that is in the adjoint representation of Spin(8). So they transform under triality like the weights of Spin(8). Under the exchange of the vector with the positive chirality spinor (leaving the negative chirality spinor fixed) the masses transform by a formula that was already presented in (10.4):

$$m'_1 = \frac{1}{2} (m_1 + m_2 + m_3 + m_4) \ ,$$
$$m'_2 = \frac{1}{2} (m_1 + m_2 - m_3 - m_4) \ ,$$
$$m'_3 = \frac{1}{2} (m_1 - m_2 + m_3 - m_4) \ ,$$
$$m'_4 = \frac{1}{2} (m_1 - m_2 - m_3 + m_4) \ . \tag{17.34}$$

Under the exchange of the vector with the other spinor, the masses transform to

$$m''_1 = \frac{1}{2} (m_1 + m_2 + m_3 - m_4) \ ,$$
$$m''_2 = \frac{1}{2} (m_1 + m_2 - m_3 + m_4) \ ,$$
$$m''_3 = \frac{1}{2} (m_1 - m_2 + m_3 + m_4) \ ,$$
$$m''_4 = \frac{1}{2} (-m_1 + m_2 + m_3 + m_4) \ . \tag{17.35}$$

These formulas have precisely the same structure as (17.31), (17.32), and (17.33)!

This means the following. Suppose that we introduce masses $m_i, i = 1, \ldots, 4$ in the $N_f = 4$ theory and determine the equation $y^2 = F(x, u, m_i, \tau)$ by requiring that the "periods" of $\omega = (\sqrt{2}/8\pi) \, dx \, du/y$ are proportional to the m_i in the sense that

$$[\omega] = -\frac{1}{\sqrt{2}} \sum_a m_a [C^{(1)}{}_a]. \tag{17.36}$$

The above equations will then ensure that it is also true that

$$[\omega] = -\frac{1}{\sqrt{2}} \sum_a m'_a [C^{(2)}{}_a] = -\frac{1}{\sqrt{2}} \sum_a m''_a [C^{(3)}{}_a]. \tag{17.37}$$

In brief, ω is triality invariant; this is achieved because triality permutes $C^{(1)}{}_a, C^{(2)}{}_a,$ and $C^{(3)}{}_a$ and also $m_a, m'_a,$ and m''_a in the same way. In this sense, the structure of the complex manifold X makes triality invariance of the physics possible.

Actually, the formulas (17.34) and (17.35) were not quite uniquely determined, since triality is only uniquely defined up to a Weyl transformation. The Weyl group acts on the m's by permutations and sign changes, so the arbitrariness that was fixed to write the triality transformation as in (17.34) and (17.35) has the same structure as the arbitrariness that was fixed in writing the transformations of the C's.

Eq. (17.37) could similarly be modified by a Weyl transformation (permutations and pairwise sign changes of the m's) without spoiling triality. This would not be an

essential change. It would also be possible to multiply the right-hand side by a constant. Since we have not proved that the $C^{(i)}{}_a$ are an *integral* basis for the cohomology, and also because we have not analyzed the monodromies of a and a_D precisely enough, we cannot assert now that (17.37) is correctly normalized. (This may be the reason for a factor of two that will appear later.)

It remains to justify (17.30). Two lines $L : x = e_i u + \theta_{ia}$ and $L' : x = e_j u + \theta_{jb}$ with $i \neq j$ are not parallel and intersect at a point P on the x–u plane. On the double cover $y^2 = F(x, u)$ there are two points $P_\pm$ lying above P. Each double cover $D^{(i)}{}_{\pm,a}$ of L and each double cover $D^{(j)}{}_{\pm,b}$ of L' contains either P_+ or P_-. If for instance, we fix conventions so that P_+ is contained in $D^{(i)}{}_{+,a}$ and $D^{(j)}{}_{+,b}$ and P_- in the others, then the $D^{(i)}{}_{+,a} \cap D^{(j)}{}_{+,b} = D^{(i)}{}_{-,a} \cap D^{(j)}{}_{-,b} = 1$, while the other intersections are zero. So $C^{(i)}{}_a \cap C^{(j)}{}_b = 2$. With other conventions, we could get $C^{(i)}{}_a \cap C^{(j)}{}_b = -2$. So the intersections are ± 2, as claimed in (17.30). (There is no way to pick conventions so that the intersection is $+2$ for all i, j, a, b.)

17.4. Determination of the equation for $N_f = 4$

We have finally assembled the tools to determine the precise function F in our equations $y^2 = F(x, u)$. First we do this for the theory with $N_f = 4$. We will determine the expansion of ω in terms of the $C^{(i)}{}_a$ for some given i; we may as well pick $i = 1$. (The expansions in terms of the $C^{(j)}{}_a$ for $j \neq 1$ are then determined by the above triality formulas.) By requiring that $\omega = -(1/\sqrt{2}) \sum_a m_a [C^{(1)}{}_a]$, F will be determined.

In coordinates with $v = 1$, the cubic part of F is $(x - e_1 u)(x - e_2 u)(x - e_3 u)$. Since e_1 is in any case singled out by the decision to expand in the $C^{(1)}{}_a$, it is convenient to make the change of variables $x - e_1 u \rightarrow x$. The lines $x = e_1 u + \theta$ are now described simply by

$$x = \theta. \tag{17.38}$$

Also, if we set

$$\alpha = e_2 - e_1, \qquad \beta = e_3 - e_1, \tag{17.39}$$

the cubic part of F is $x(x - \alpha u)(x - \beta u)$. The quadratic part of F is a linear combination of x^2, xu, and u^2. We can eliminate any two of the three by shifting x and u by constants. We choose to set the coefficients of xu and u^2 to zero. The general structure is then

$$F = x(x - \alpha u)(x - \beta u) + ax^2 + bx + cu + d. \tag{17.40}$$

Restricted to the line (17.38), one has

$$F = \theta \alpha \beta u^2 + (c - \theta^2(\alpha + \beta))u + \theta^3 + a\theta^2 + b\theta + d. \tag{17.41}$$

If we write the right-hand side as $Au^2 + Bu + C$, then the discriminant $\Delta = B^2 - 4AC$ is

$$\Delta = \theta^4(\alpha - \beta)^2 - 4a\alpha\beta\theta^3 + (-2c(\alpha + \beta) - 4b\alpha\beta)\theta^2 - 4\alpha\beta d\theta + c^2. \tag{17.42}$$

Δ has four roots θ_a, $a = 1, \ldots, 4$. For $\theta = \theta_a$, the equation $y^2 = F$ takes the form

$$y^2 = \theta_a \alpha \beta (u - u_0)^2 \tag{17.43}$$

for some u_0. Restoring the v dependence, in homogenous coordinates, the equations for the divisors $D^{(1)}{}_{\pm,a}$ take the form

$$\begin{aligned} x &= \theta v, \\ y &= \pm(\theta_a \alpha \beta)^{1/2} v(u - u_0 v). \end{aligned} \tag{17.44}$$

Computation of periods

Now we want to expand $[\omega]$ in terms of these divisors. The main point is to study the behavior of ω near ∞ (that is, $v = 0$), where we can set $u = 1$. The equation $y^2 = vF(x, u, v)$ becomes in this coordinate system

$$y^2 = vx(x - \alpha)(x - \beta) + av^2 x^2 + bv^3 x + cv^3 + dv^4. \tag{17.45}$$

So near the singularity at $x = v = 0$, we get

$$2y \cdot dy = v\alpha\beta \, dx + \ldots \tag{17.46}$$

where the $\ldots$ are terms proportional to dv (which will drop out when we compute ω) or v^2, vx (which are negligible near $v = x = 0$). Inserting (17.46) in $\omega = (\sqrt{2}/8\pi) \, dv \, dx/vy$, we get

$$\omega \sim \frac{\sqrt{2}}{4\pi} \frac{dv \, dy}{v^2 \alpha \beta}. \tag{17.47}$$

This means that near $v = 0$ we can write $\omega = d\lambda$ with

$$\lambda = -\frac{\sqrt{2}}{4\pi} \frac{y \, dv}{v^2 \alpha \beta}. \tag{17.48}$$

Let X_ϵ be the region with $|v| < \epsilon$. We want to modify ω inside X_ϵ to eliminate the pole at $v = 0$, while preserving the fact that ω is closed. The ω so modified extends over a neighborhood of the singularity at $x = v = 0$ in the compactification of X. (ω could be extended over all of $\overline{X}$, but that will not be necessary.)

To make the modification, let $\lambda' = f(v)\lambda$, with f a smooth function such that $f(v) = 1$ for $|v| > \epsilon$ and $f \sim |v|^2$ for $v \to 0$. Now, leave ω unchanged outside of X_ϵ, but inside X_ϵ take

$$\omega = d\lambda'. \tag{17.49}$$

We now have all the information required to compute the desired residues, which, according to (17.10) and (17.12), are equivalent to the integrals of ω over $C^{(1)}{}_a = D^{(1)}{}_{+,a} - D^{(1)}{}_{-,a}$. The integral over $C^{(1)}{}_a$ is twice the integral over $D = D^{(1)}{}_{+,a}$, using the symmetry under $y \leftrightarrow -y$:

$$c_a = \int_{C^{(1)}{}_a} \omega = 2 \int_D \omega. \tag{17.50}$$

Let D_ϵ be the part of D with $|v| < \epsilon$. Then

$$\int\limits_{D} \omega = \int\limits_{D_\epsilon} \omega, \tag{17.51}$$

as outside D_ϵ, ω is of type $(2, 0)$. Inside D_ϵ, $\omega = d\lambda'$, so

$$\int\limits_{D_\epsilon} \omega = \int\limits_{D_\epsilon} d\lambda' = \oint\limits_{|v|=\epsilon} \lambda' = \oint\limits_{|v|=\epsilon} \lambda = -\frac{\sqrt{2}}{4\pi} \oint\limits_{|v|=\epsilon} \frac{y \, dv}{v^2 \alpha\beta}$$

$$= -\frac{\sqrt{2}}{4\pi} \oint\limits_{|v|=\epsilon} \frac{\sqrt{\theta_a \alpha\beta} \, dv}{v\alpha\beta} = -\frac{i}{\sqrt{2}} \sqrt{\frac{\theta_a}{\alpha\beta}}. \tag{17.52}$$

The next to last step uses the fact that near $v = 0$ on D,

$$y \sim (\theta_a \alpha\beta)^{1/2} v \tag{17.53}$$

according to (17.44). So

$$c_a = -\sqrt{2} i \sqrt{\frac{\theta_a}{\alpha\beta}}. \tag{17.54}$$

Now we can determine the desired residues from (17.12) – using the fact that the matrix M is $M = -4$ according to (17.29). We get

$$\mathrm{Res}_{C^{(1)}{}_a} \lambda = -\frac{1}{4\pi\sqrt{2}} \sqrt{\frac{\theta_a}{\alpha\beta}}. \tag{17.55}$$

Here of course θ_a is any of the roots of the discriminant (17.42).

Final steps

On the other hand, we want the residues to be $m_a/2\pi i \sqrt{2}$, with m_a the masses. So the four zeroes θ_a of the discriminant should be $\theta_a = -4\alpha\beta m_a^2$. For the discriminant to have these roots means that (17.42) can be rewritten as follows:

$$\theta^4 (\alpha - \beta)^2 - 4a\alpha\beta\theta^3 + (-2c(\alpha + \beta) - 4b\alpha\beta)\theta^2 - 4\alpha\beta d\theta + c^2$$

$$= (\alpha - \beta)^2 \prod_{a=1}^{4} \left(\theta + 4\alpha\beta m_a^2\right). \tag{17.56}$$

Simply by equating the coefficients of different powers of θ, we now determine all the unknown quantities: [17]

---

[17] In solving the equations, one has to take a square root; a sign change in the square root is equivalent to a change in sign of one of the four masses. A similar and related choice was needed in Section 16 when we broke the symmetry between e_2 and e_3.

$$a = -(\alpha - \beta)^2 \sum_a m_a{}^2$$

$$b = -4(\alpha - \beta)^2 \alpha\beta \sum_{a<b} m_a{}^2 m_b{}^2 + 8\alpha\beta(\alpha^2 - \beta^2) \prod_{a=1}^4 m_a$$

$$c = -16(\alpha - \beta)\alpha^2\beta^2 \prod_{a=1}^4 m_a$$

$$d = -16(\alpha - \beta)^2 \alpha^2 \beta^2 \sum_{a<b<c} m_a{}^2 m_b{}^2 m_c{}^2. \tag{17.57}$$

So we determine finally the equation governing the low energy behavior of the $N_f = 4$ theory:

$$y^2 = x(x - \alpha u)(x - \beta u) - (\alpha - \beta)^2 x^2 \sum_a m_a{}^2$$

$$+ x\left(-4(\alpha - \beta)^2 \alpha\beta \sum_{a<b} m_a{}^2 m_b{}^2 + 8\alpha\beta(\alpha^2 - \beta^2) \prod_{a=1}^4 m_a \right)$$

$$- 16u(\alpha - \beta)\alpha^2\beta^2 \prod_{a=1}^4 m_a - 16(\alpha - \beta)^2 \alpha^2\beta^2 \sum_{a<b<c} m_a{}^2 m_b{}^2 m_c{}^2. \tag{17.58}$$

In particular, the right-hand side is a polynomial in x, u, and the m_i, as we assumed in all of our analyses of the various models. In fact, (17.58) can be seen to be equivalent to the result obtained in Section 16 (the manifestly triality invariant expression (16.35) and its shifted form (16.38)), verifying that the result of Section 16 is compatible with the residue condition. To see that, substitute in (16.38) $x \rightarrow x + c_1 u$ and use $c_1 = \frac{3}{2} e_1 = -\frac{1}{2}(\alpha + \beta)$ and $c_2 = \frac{1}{2}(e_3 - e_2) = \frac{1}{2}(\beta - \alpha)$ to derive (17.58). Actually, to agree with Section 16, it is also necessary to divide m by 2; we do not understand the origin of this discrepancy, but it may have to do with subtleties in normalizing the topological computation that were mentioned in the next to last paragraph of Section 17.3.

Most of $SL(2, \mathbb{Z})$ invariance is obvious in (17.58) since $\alpha = e_2 - e_1$ and $\beta = e_3 - e_1$ are modular forms of weight two for $\Gamma(2)$. Actually, the full $SL(2, \mathbb{Z})$ and triality are guaranteed, for the following reason. The condition on the residues of ω is consistent with S-duality and triality, as we saw in equation (17.37). Furthermore, in using that condition to find F, the "boundary" condition, given by the cubic function $(x - e_1 u)(x - e_2 u)(x - e_3 u)$, is also S-dual and triality-invariant. Moreover, the condition on ω led to a unique determination of F, up to the possibility of redefining x and adding a constant to u. Therefore, up to such transformations, the solution must have all the symmetries of the boundary conditions and, in particular, triality and $SL(2, \mathbb{Z})$.

17.5. The N = 4 Theory

In Eq. (16.17), we determined the general structure of the curve for $N = 4$:

$$y^2 = (x - e_1\widetilde{u} - e_1{}^2 f)(x - e_2\widetilde{u} - e_2{}^2 f)(x - e_3\widetilde{u} - e_3{}^2 f). \qquad (17.59)$$

We then showed f to be τ-independent by an indirect argument involving the residues and showed it to be $m^2/4$ by comparing to the behavior at infinity. Our purpose here is to analyze the residues for this curve more directly and completely.

First we discuss the classical geometry of the situation. The complex manifold X given by the above equation has three singularities where

$$0 = y = x - e_i\widetilde{u} - e_i{}^2 f = x - e_j\widetilde{u} - e_j{}^2 f \qquad (17.60)$$

for any distinct i, j. These are $\mathbf{A}_1$ singularities. We already know that for generic cubic F, the part of the cohomology of the complex manifold $y^2 = F(x, \widetilde{u})$ that is odd under $y \leftrightarrow -y$ is four dimensional. Each time an $\mathbf{A}_1$ singularity appears, the dimension of that part of the cohomology decreases by one. For the special case of the equation in (16.17), there are three singularities, so the odd part of the cohomology is one dimensional. The number one is no coincidence; the mass-deformed $N = 4$ theory has one mass parameter m.

Since the relevant part of the cohomology is one dimensional, there is one period or residue, and we will use it to determine f.

As in the discussion of $N_f = 4$, we consider a line in the x–$\widetilde{u}$ plane of the form

$$x = e_1\widetilde{u} + e_1{}^2 f + \theta. \qquad (17.61)$$

Restricted to that line, Eq. (17.59) becomes

$$y^2 = \theta \left(\theta + (e_1 - e_2)\widetilde{u} + (e_1{}^2 - e_2{}^2)f\right)\left(\theta + (e_1 - e_3)\widetilde{u} + (e_1{}^2 - e_3{}^2)f\right). \qquad (17.62)$$

The condition that the right-hand side is a perfect square gives

$$\theta = -(e_1 - e_2)(e_1 - e_3)f. \qquad (17.63)$$

For that value of θ, we get two curves D_+ and D_- given by (17.61) together with

$$y = \pm \left(\theta(e_1 - e_2)(e_1 - e_3)\right)^{1/2}(\widetilde{u} - \widetilde{u}_0). \qquad (17.64)$$

Taking $C = D_+ - D_-$ gives one divisor odd under $y \leftrightarrow -y$. One is enough as the odd part of the cohomology is one dimensional. One can show that

$$C \cap C = -2. \qquad (17.65)$$

Now we can repeat the derivation of (17.55). The only real difference in determining $c = \int_C \omega$ is that now $\omega = (\sqrt{2}/4\pi)\,dv\,dx/vy$. So we get

$$c = -2i\sqrt{2}\sqrt{\frac{\theta}{(e_1 - e_2)(e_1 - e_3)}}. \qquad (17.66)$$

In determining the residue λ, we must also remember that according to (17.65), the intersection matrix is now $M = -2$ instead of -4. So we get

$$\text{Res}_C(\lambda) = -\frac{i\sqrt{2}}{2\pi i}\left(\frac{\theta}{(e_1 - e_2)(e_1 - e_3)}\right)^{1/2}. \tag{17.67}$$

Setting this residue equal to $m/2\pi i\sqrt{2}$, we get $\theta^{1/2} = im((e_1 - e_2)(e_1 - e_3))^{1/2}/2$. Comparing to (17.63) gives $f = m^2/4$, so finally the curve of the mass-deformed $N = 4$ theory is

$$y^2 = (x - e_1\widetilde{u} - \frac{1}{4}e_1{}^2 m^2)(x - e_2\widetilde{u} - \frac{1}{4}e_2{}^2 m^2)(x - e_3\widetilde{u} - \frac{1}{4}e_3{}^2 m^2), \tag{17.68}$$

in agreement with the result obtained in Section 16.2.

Acknowledgements

References

[1] M.A. Shifman and A.I. Vainshtein, Nucl. Phys. B 277 (1986) 456; B 359 (1991) 571.
[2] D. Amati, K. Konishi, Y. Meurice, G.C. Rossi and G. Veneziano, Phys. Rep. 162 (1988) 169 and references therein.
[3] N. Seiberg, Phys. Lett. B 318 (1993) 469.
[4] N. Seiberg, Phys. Rev. D 49 (1994) 6857, hep-th/9402044.
[5] K. Intriligator, R. Leigh and N. Seiberg, Phys. Rev. D 50 (1994) 1092, hep-th/9403198.
[6] V. Kaplunovsky and J. Louis, hep-th/9402005.
[7] S.J. Gates, Jr., Nucl. Phys. B 238 (1984) 349;
    B. De Wit and A. Van Proeyen, Nucl. Phys. B 245 (1984) 89.
[8] N. Seiberg, Phys. Lett. B 206 (1988) 75.
[9] N. Seiberg and E. Witten, RU-94-52, IAS-94-43, hep-th/9407087.
[10] I. Affleck, M. Dine and N. Seiberg, Nucl. Phys. B 241 (1984) 493; B 256 (1985) 557.
[11] M.K. Prasad and C.M. Sommerfield, Phys. Rev. Lett. 35 (1975) 760.
[12] E.B. Bogomol'nyi, Sov. J. Nucl. Phys. 24 (1976) 449.
[13] E. Witten and D. Olive, Phys. Lett. B 78 (1978) 97.
[14] T. Banks and E. Rabinovici, Nucl. Phys. B 160 (1979) 349;
     E. Fradkin and S. Shenker, Phys. Rev. D 19 (1979) 3682.
[15] C. Montonen and D. Olive, Phys. Lett. B 72 (1977) 117;
     P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. B 125 (1977) 1.
[16] J. Cardy and E. Rabinovici, Nucl. Phys. B 205 (1982) 1;
     J. Cardy, Nucl. Phys. B 205 (1982) 17.
[17] A. Shapere and F. Wilczek, Nucl. Phys. B 320 (1989) 669.
[18] A. Sen, TIFR-TH-94-08, hep-th/9402002; G. Segal, to appear
[19] J. Wess and J. Bagger, Supersymmetry and supergravity (Princeton Univ. Press, Princeton, 1982).
[20] R. Jackiw, C. Rebbi, Phys. Rev. D 13 (1976) 3398.

[21] C. R. Nohl, Phys. Rev. D 12 (1975) 1840.
[22] M.F. Atiyah and N.J. Hitchin, The geometry and dynamics of magnetic monopoles (Oxford Univ. Press, Oxford, 1988).
[23] E. Witten, Phys. Lett. B 86 (1979) 283.
[24] S. Cecotti, P. Fendley, K. Intriligator and C. Vafa, Nucl. Phys. B 386 (1992) 405;
     S. Cecotti and C. Vafa, Commun. Math. Phys. 158 (1993) 569.
[25] N. S. Manton and B. J. Schroers, Ann. Phys. 225 (1993) 290.
[26] N. Koblitz, Introduction to elliptic curves and modular forms (Springer, Berlin, 1984).
[27] K. Chandrasekharan, Elliptic functions (Springer, Berlin, 1985).
[28] T. Hubsch, Calabi-Yau manifolds: A bestiary for physicists (World-Scientific, Singapore, 1992).