Lecture II-4: The large N limit of the $\mathbb{CP}^{N-1}$ model

Edward Witten

Notes by Dan Freed

Remark. *The lecture treated the σ-model into projective space, but these notes cover the generalization to a Grassmannian, as requested in Problem Set 3.*

In this lecture we discuss the large N behavior of the two dimensional σ-model into the Grassmannian $Gr(k,N)$ of k dimensional subspaces of $\mathbb{C}^N$. Here k is fixed as $N\to\infty$. We also consider the real Grassmannian. Since Grassmannians have positive Ricci curvature these field theories are asymptotically free, but in any case our task is to investigate the infrared behavior.

The Euclidean action of the σ-model is

$$S[\phi]=\frac{1}{g^2}\int_\Sigma d^2x\,|d\phi|^2-i\theta\int_\Sigma \phi^*(\alpha), \tag{1}$$

where Σ is a Riemann surface, $\phi\colon \Sigma\to Gr(k,N)$ a map into the Grassmannian, $\alpha\in H^2\big(Gr(k,N),\mathbb{Z}\big)$ a generator of the cohomology, and g,θ are parameters of the theory. We specify the metric on $Gr(k,N)$ shortly. The second term is a topological term;[1] for Σ closed the integral is integer-valued. Thus a shift $\theta\to\theta+2\pi$ does not affect the model. The parameter θ is also a parameter of the quantum theory, but renormalization exchanges the dimensionless coupling constant g with a mass parameter μ.

The rescaled coupling constant $\tilde g$, defined by

$$g^2=\tilde g^2/N, \tag{2}$$

is more natural in the large N limit, as we will see.

[1]The factor of i is present in the *Euclidean* action so that the action conjugates under orientation reversal. In this way its continuation to Minkowski space is real. See Freed's notes *Actions and Reality* for more details.

§4.1. The Questions

We ask specific questions about the behavior of the model.

1. Is there a mass gap?
2. What is the θ dependence of the partition function?

   Remark. For Σ small the partition function

$$Z(\theta) = \int D\phi \; e^{-\frac{1}{g_{\text{eff}}^2}\int_\Sigma d^2x\,|d\phi|^2} \, e^{i\theta\int_\Sigma \phi^*(\alpha)}$$

   can be studied using perturbation theory. Here g_{eff} is the effective coupling, which varies with the distance scale in the theory set by the size of the surface Σ. (Classically the model is conformally invariant, so does not depend on the size of Σ, but quantum mechanically this is no longer true.) By asymptotic freedom this coupling is small at small distances, hence the assertion that perturbation theory applies in this regime. Now the classical solutions to (1) are harmonic maps $\Sigma \to Gr(k,N)$. Note that the only θ dependence is through the degree of ϕ, and so we split the integral as a sum over maps ϕ of varying degrees. In degree 0 we obtain constant maps and in general for degree n some moduli space $\mathcal{M}_n$ of harmonic maps, which are the *instantons* of this model. Degree 1 instantons are (anti)holomorphic maps. The perturbation expansion around these solutions has the rough form

$$(3) \qquad Z(\theta) \sim \frac{\mathrm{Vol}\big(Gr(k,N)\big)}{\sqrt{\det(\cdot)}}(1+\cdots) + \sum_{\pm} e^{\pm i\theta} e^{-cN/g_{\text{eff}}^2} \frac{\mathrm{Vol}\,\mathcal{M}_1}{\sqrt{\det(\cdot)}}(1+\cdots) + \text{higher instantons}$$

   The constant c in the exponential is the action of a 1-instanton, which is independent of N. The factor of N comes from (2). If $g_{\text{eff}} << 1$ we see that the θ dependent term of $Z(\theta)$ vanishes exponentially as $N \to \infty$. As the area of Σ increases the effective coupling g_{eff} also increases. Equation (3) is the answer for Σ compact and of small area, but we will find a vastly different result for $\Sigma = \mathbb{R}^2$.
3. *Symmetry Breaking.* Symmetry breaking in two dimensions is possible for a discrete symmetry, and in this model we have the parity symmetry

$$(4) \qquad \begin{aligned} P\colon \Sigma &\longmapsto \overline{\Sigma} \qquad\qquad (\overline{\Sigma} \text{ is } \Sigma \text{ with the opposite orientation}) \\ \theta &\longmapsto -\theta \end{aligned}$$

   For $\Sigma = \mathbb{R}^2$ we implement the orientation reversal by an orientation-reversing isometry, i.e., a reflection. Then for $\theta = 0$ and $\theta = \pi$ the parity symmetry P acts on a fixed theory and we ask if it is broken in the quantum theory.

4. The group $PSU(N)$ of isometries of the Grassmannian acts in the classical theory. Since continuous symmetry groups are unbroken in two dimensions, this symmetry acts in the quantum theory as well. However, it is possible that a realization of the theory has a symmetry group which is a cover of $PSU(N)$, the latter being the group which acts on the operator algebra. Does that happen here?

§4.2. An Equivalent Formulation

To study the model we rewrite it, that is, we construct an action with the same classical and quantum physics. As a preliminary we recall some basic geometry of the Grassmannian. Over $Gr(k,N)$ lies a canonical sequence of vector bundles

$$0 \to S \xrightarrow{s} Gr(k,N) \times \mathbb{C}^N \to Q \to 0, \tag{5}$$

where the fiber of S at a k-plane π is simply π viewed as a subspace of $\mathbb{C}^N$. Fix the standard metric on $\mathbb{C}^N$. It induces a metric on S and identifies $Q \cong S^{\perp}$. There is a canonical connection ∇ on S, obtained by projecting the natural connection on the trivial bundle $Gr(k,N) \times \mathbb{C}^N$. We easily compute

$$\nabla = d - s^* ds, \tag{6}$$

where s is the inclusion $S \xrightarrow{s} Gr(k,N) \times \mathbb{C}^N$. Then

$$\nabla s \colon T\big(Gr(k,N)\big) \longrightarrow \operatorname{Hom}(S, S^{\perp}) \tag{7}$$

is an isomorphism. We use it to induce a metric on $Gr(k,N)$, the metric needed to write down the σ-model action (1).

Now if $\phi\colon \Sigma \to Gr(k,N)$ we pullback (5) to obtain a sequence of bundles over Σ, and by (7) the lagrangian density of ϕ is

$$|d\phi|^2 = |(\phi^*\nabla)\phi^* s|^2. \tag{8}$$

Note $\phi^* s\colon \phi^* S \to \Sigma \times \mathbb{C}^N$ determines ϕ. The idea is to replace ϕ by such a bundle map, and so first to replace $\phi^* S$ by a fixed bundle. Note $\deg(\phi^* S) = \deg(\phi)$ so that the topology of $\phi^* S$ determines the cohomology class $\phi^*(\alpha)$, which appears in the second term of the action (1). Hence fix a vector bundle $E \to \Sigma$ of rank k and degree d. Also fix a hermitian metric on E. We introduce a new field

$$\hat{\phi}\colon E \longrightarrow \Sigma \times \mathbb{C}^N$$

which we constrain to be an isometric immersion:

$$\hat{\phi}^*\hat{\phi} = \mathrm{id}_E\,. \tag{9}$$

The image of $\hat{\phi}$ determines a map $\phi\colon \Sigma \to Gr(k,N)$ which is unchanged if we shift $\hat{\phi}$ by a unitary gauge transformation of E. To rewrite (8) in terms of $\hat{\phi}$ we need a connection on E, and as there is no natural choice we introduce a variable unitary connection A. Using $\hat{\phi}$ we identify E with a subbundle of the trivial bundle $\Sigma \times \mathbb{C}^N$, so can differentiate $\hat{\phi}$ using the usual derivative d. Writing A as a 1-form plus this trivial connection we find

$$\begin{aligned} |d_A\hat{\phi}|^2 &= |d\hat{\phi} + \hat{\phi}A|^2 \\ &= |d\hat{\phi}|^2 + 2\,\mathrm{Re}(d\hat{\phi}, \hat{\phi}A) + |A|^2, \end{aligned} \tag{10}$$

since $\hat{\phi}^*\hat{\phi} = \mathrm{id}_E$. This expression is quadratic in A, so if (10) is a classical lagrangian for A we can use the equations of motion to obtain

$$A_0 = -\hat{\phi}^* d\hat{\phi}. \tag{11}$$

Comparing with (6) we see that A_0 is the pullback of the canonical connection on S, and so by (8)

$$|d\phi|^2 = |d_{A_0}\hat{\phi}|^2.$$

In other words, the lagrangian (10) is equivalent to $|d\phi|^2$ for fields which satisfy the constraint (9). We impose the constraint via a lagrange multiplier field

$$\sigma\colon \Sigma \longrightarrow \mathrm{HermitianEnd}(E).$$

The θ term in the original action (1) can be computed using the (skew-Hermitian) curvature F_A via Chern-Weil theory. Altogether we obtain for our new action[2]

$$S[\hat{\phi}, A, \sigma] = \frac{1}{g^2}\int_\Sigma |d_A\hat{\phi}^2| - i\int_\Sigma \mathrm{Tr}\,\sigma(\hat{\phi}^*\hat{\phi} - \mathrm{id}_E) + \frac{\theta}{2\pi}\int_\Sigma \mathrm{Tr}\,F_A. \tag{12}$$

[2]Some explanation about the lagrange multiplier term is in order. It is based on the general formula

$$\int e^{i(\sigma,x)}d\sigma = \delta(x),$$

where σ lies in the dual space to x, the measure $d\sigma$ is suitable normalized, and $\delta(x)$ is the δ-distribution supported at the origin. In the Minkowski space lagrangian the lagrange multiplier term is

$$\int_\Sigma \mathrm{Tr}\,\sigma(\hat{\phi}^*\hat{\phi} - \mathrm{id}_E),$$

and σ should be interpreted as a 2-form (or density); it lies in the dual space to the function $\hat{\phi}^*\hat{\phi} - \mathrm{id}_E$. Rotation to Euclidean space yields the second term of (12).

As we have explained, the classical equations of motion (and other classical constructs) computed from (12) are equivalent to those computed from the original action (1). The classical computation which led to (11) is valid quantum mechanically since the dependence of (12) on A is quadratic and the Hessian is the identity operator (see (10)). (We ignore the constant determinant factor which we obtain from the A integral.) Thus the quantum physics is the same as well.

§4.3. The Large N Effective Theory

The argument here is almost identical to that for the large N σ-model into a sphere (see lecture II-3). So we will be brief.

First, the $\hat{\phi}$ integral is Gaussian, so the partition function is

$$\begin{aligned} Z &= \sum_E \int \frac{DA}{\mathrm{vol}} D\sigma D\hat{\phi}\, e^{-S[\hat{\phi},A,\sigma]} \\ &= \sum_E \int \frac{DA}{\mathrm{vol}} D\sigma \exp - \left[\mathrm{Tr} \ln \left(\frac{d_A^* d_A}{g^2} - i\sigma \right) + i \int_\Sigma \mathrm{Tr}\, \sigma + \frac{\theta}{2\pi} \int_\Sigma \mathrm{Tr}\, F_A \right]. \end{aligned} \tag{13}$$

Here $\sum_E$ is the sum over bundles E. We take $\Sigma = \mathbb{R}^2$. Then the sum over E is irrelevant, as is the topological term. In a first approach to this problem, the topology of the bundle will not be important, since it is spread over an infinite volume. Once we get a basic understanding of what the quantum vacuum looks like, it will not be hard to go back and see how the topology enters. We evaluate the leading behavior in large N by evaluating at a stationary point of the exponential in the integrand, i.e., a classical solution of the effective action.

The only Poincaré invariant (that is, Euclidean invariant) possibility for the gauge field is $A = 0$. In that case we rewrite minus the integrand in the last line of (13) as

$$S_{\mathrm{eff}}[A, \sigma] = N\, \mathrm{Tr} \ln \left(\frac{d^* d}{g^2} - i\sigma \right) + i \int \mathrm{Tr}\, \sigma,$$

where the operator in the first term acts on sections of E^*. Now the only Poincaré invariant possibility is σ constant, and so we diagonalize σ and pass to k one dimensional problems for an eigenvalue. At this point we rescale

$$\begin{aligned} \sigma &= N\tilde{\sigma} \\ g^2 &= \tilde{g}^2/N, \end{aligned} \tag{14}$$

and so for each eigenvalue we have exactly the problem we had for the large N σ-model in to a sphere. Thus the solution $\tilde{\sigma}_0$ has all eigenvalues equal and is specified by

$$-i\tilde{g}^2 \tilde{\sigma}_0 = \Lambda^2 e^{-4\pi/\tilde{g}^2} = M^2$$

for Λ an ultraviolet cutoff and now $\tilde{g}^2$ the running coupling constant. We define M^2 to be this dynamically generated mass squared.

So in the large N effective action $\tilde{\sigma}$ acquires a mass. Note that the A field has no transverse degrees of freedom—it is a gauge field in two dimensions—so does not enlarge the spectrum of the model (though, of course, it affects the Hamiltonian as we will see, and in fact diminishes the spectrum). Thus the large N limit has a mass gap. This is the answer to the first in our list of questions.

To answer the other questions we need to compute something more precise, namely the leading approximation to the large N effective action. This means that we do perturbation theory for the action (12) about the point $\hat{\phi}_0 = 0$, $A_0 = 0$, $\tilde{\sigma}_0 = iM^2/\tilde{g}^2$. So shifting $\tilde{\sigma}$ by $\tilde{\sigma}_0$ and rescaling $\hat{\phi}$ we have from (12)

$$S'[\hat{\phi}, A, \tilde{\sigma}] = \int_\Sigma |d_A\hat{\phi}|^2 \;+\; M^2\int_\Sigma |\hat{\phi}|^2 \;-\; i\tilde{g}^2\int_\Sigma \operatorname{Tr}\tilde{\sigma}\hat{\phi}^*\hat{\phi} \;+\; \frac{\theta}{2\pi}\int_\Sigma \operatorname{Tr} F_A. \tag{15}$$

Here we omit a constant term (which only shifts the partition function by a constant) and a linear term in $\tilde{\sigma}$ (since we are expanding around a solution of the effective action).

Figure 1: The inverse $\tilde{\sigma}$ propagator

The effective action is computed in perturbation theory using one particle irreducible Feynman diagrams with external lines for A and σ and with internal $\hat{\phi}$ lines. So the inverse propagator for $\tilde{\sigma}$ is computed from the diagram shown in Figure 1, where the solid line represents $\hat{\phi}$ and the dotted line represents σ. (Note that (15) has no quadratic term in $\tilde{\sigma}$, else Figure 1 would be a correction to such a term.) We evaluate the diagram in momentum space as

$$\begin{aligned}
-\tilde{g}^4 Nk\int\frac{d^2q}{(2\pi)^2}\,\frac{1}{q^2+M^2}\,\frac{1}{(p-q)^2+M^2} &= -\frac{\tilde{g}^4Nk}{4\pi^2}\int_0^1 d\alpha\int d^2q\,\frac{1}{[q^2+(M^2+\alpha(1-\alpha)p^2)]^2}\\
&= -\frac{\tilde{g}^4Nk}{\pi}\int_0^1\frac{d\alpha}{M^2+\alpha(1-\alpha)p^2}\\
&= -\frac{\tilde{g}^4Nk}{\pi M^2} + O(p^2) \qquad \text{as } p\to 0.
\end{aligned}\tag{16}$$

This corresponds to a term

$$C|\tilde{\sigma}|^2$$

in the effective action, with $C > 0$. (The minus sign comes since in the Euclidean theory diagrams compute negative contributions to the effective action.) This is the dominant term in the infrared, which means that $\tilde{\sigma}$ is massive and does not affect the long range behavior of the theory.

FIGURE 2: THE INVERSE A PROPAGATOR

FIGURE 3: THE INTERACTION VERTICES FOR THE A FIELD

The inverse propagator for A is computed by the diagrams in Figure 2, which come from the first term in (15). Here the wavy line represents A. To compute these diagrams we need the Feynman rules for the vertices indicated in Figure 3, which correspond to the terms

$$2\,\mathrm{Re}(d\hat{\phi}, \hat{\phi}A)$$
$$|\hat{\phi}A|^2$$

in the action (15). The indices refer to a standard orthonormal basis for $\mathbb{R}^2$. The Feynman rule (in momentum space) for the second vertex is easy:

$$II = -\delta_{\mu\nu}.$$

(There is a minus sign since the Euclidean functional integral involves e^{-S}.) For the vertex I we must remember that $\hat{\phi}$ is complex and that A is skew-Hermitian:

$$2\,\mathrm{Re}(\partial_\mu\hat{\phi}, \hat{\phi}A_\mu) = -\partial_\mu\hat{\phi}A_\mu\hat{\phi}^* + \hat{\phi}A_\mu\partial_\mu\hat{\phi}^*.$$

Thus the vertex is

$$I = -i(k_1 - k_2)_\mu.$$

Note that one of the solid lines in the vertex represents $\hat{\phi}$ and the other solid line represents $\hat{\phi}^*$. So the sum of the diagrams in Figure 2 is

$$\int \frac{d^2q}{(2\pi)^2}\frac{(p-2q)_\mu(p-2q)_\nu}{(q^2+M^2)((p-q)^2+M^2)} \;-\; 2\delta_{\mu\nu}\int\frac{d^2q}{(2\pi)^2}\frac{1}{(q^2+M^2)}. \tag{17}$$

The factor of 2 in the second term is from the two ways of attaching the A lines to the external vertices; the corresponding factor of 2 in the first term is canceled by the symmetry which exchanges the two internal vertices in the first diagram. (That is, there is a factor of 1/2 from the expansion of the exponential, since we have two triple vertices.) Each term in (17) is divergent, but the divergences cancel in the difference, and after some computation similar to (16) the answer to leading order in p is

$$\frac{N}{12\pi M^2}(p_\mu p_\nu - p^2\delta_{\mu\nu}). \tag{18}$$

This corresponds to a term

$$\frac{N}{24\pi M^2}|F_A|^2 \tag{19}$$

in the effective action. (Again we must recall that diagrams contribute negatively to the effective action.) In fact, (18) and (19) correspond precisely if $k = 1$ (the σ-model into projective space), since then A is an abelian connection. In the nonabelian case there are cubic and quartic terms, but by gauge invariance their leading contribution must be as in (19). We remark that (19) is the lowest order gauge invariant term we can write, and so its appearance can be predicted from gauge invariance alone, but of course we must do a computation to determine the coefficient.

So, finally, the long distance behavior of the two dimensional σ-model into the Grassmannian $Gr(k,N)$, in the large N limit, is equivalent to the long distance behavior of a two dimensional gauge theory with gauge group $U(k)$ and charged massive scalars. The action is

$$\frac{N}{2e^2}\int_\Sigma |F_A|^2 + \frac{\theta}{2\pi}\int_\Sigma \mathrm{Tr}\,F_A + \int_\Sigma |d_A\phi|^2 + M^2\int_\Sigma |\phi|^2. \tag{20}$$

The first term in (20) was just computed, where the coupling e^2 summarizes the numerical factor in (19). The second term is the θ term in (15). The last two terms are the first two terms in (15), except we now drop the '$\hat{\ }$' for convenience. So ϕ is a section of $(E^*)^{\oplus N}$.

§4.4. Real Grassmannians

From the beginning we can replace the complex Grassmannian with the real Grassmannian. In that case the bundle E is real, $\hat{\phi}\colon E \to \Sigma \times \mathbb{R}^N$, and there is no other change except in the representation of the θ term. For simplicity we consider the σ-model into the Grassmannian $Gr^0_{\mathbb{R}}(k,N)$ of *oriented* k-planes in $\mathbb{R}^N$. (It is a double cover of the Grassmannian of unoriented k-planes.) Now $H^2\big(Gr^0_{\mathbb{R}}(k,N);\mathbb{Z}\big) \cong \mathbb{Z}/2\mathbb{Z}$ and the θ term is meant to detect this class. Thus the second term of (1) is replaced by

$$i\theta \deg_2 \alpha,$$

where $\deg_2$ is the mod 2 degree (0 or 1) and $\theta = 0$ or $\theta = \pi$. In the reformulation of the problem E is an oriented real k-plane bundle over Σ, and the topological term in (12) and subsequent formulas is

$$i\theta w_2(E)[\Sigma], \tag{21}$$

where w_2 is the second Stiefel-Whitney class. For $\Sigma = \mathbb{R}^2$ we can rewrite (21) in terms of a *Wilson line* operator. Namely, a field configuration with finite action (12) has A essentially flat at infinity. Since $\mathbb{R}^2$ is contractible we can lift the $SO(k)$ connection A to a Spin(k) connection $\tilde{A}$, and the holonomy on a large loop $C \subset \mathbb{R}^2$ is approximately ± 1 depending on the Stiefel-Whitney class of the induced bundle on S^2. (We let '-1' denote the nontrivial element of Spin(k) covering $1 \in SO(k)$.) Choose a representation R of Spin(k) and consider

$$W_R(C) = \frac{\operatorname{Tr}_R \operatorname{hol}_C(A)}{\dim R}. \tag{22}$$

In the limit where the loop C becomes large, this computes the exponential of (21), where $\theta = 0$ if R is a representation of $SO(k)$ and $\theta = \pi$ if R is a representation of Spin(k) but not of $SO(k)$.

The generalization to an arbitrary connected compact Lie group G is clear. A G bundle over Σ has a characteristic class in $H^2(\Sigma, \pi_1 G)$. It pairs with a homomorphism $e^{i\theta}\colon \pi_1 G \to \mathbb{T}$ to give a term in the exponentiated action. Here $\mathbb{T} = U(1)$ is the circle group. On the other hand a representation of the simply connected covering group induces a homomorphism $e^{i\theta}\colon \pi_1 G \to \mathbb{T}$, so we can use the Wilson operator (22) to write the θ term on $\mathbb{R}^2$.

§4.5. Pure Gauge Theory

We still must determine the quantum behavior of the theory with effective action (20). For this, we will first practice by analyzing the pure gauge theory in two dimensions, whose action is the sum of the first two terms of (20). We may as well consider an arbitrary connected compact Lie group G. The norm in the first term of (20) is defined via a bi-invariant metric on G. We quantize the theory on the circle S^1_V of circumference (=volume) V. In general, to quantize a gauge theory in n dimensions on an n-manifold Y, we consider connections on $\mathbb{R} \times Y$ in *temporal gauge* and take solutions to the equations of motion up to time-independent gauge transformations. (See Kazhdan's lectures on the quantization of gauge theories.) For pure gauge theory without the θ term the resulting space is the (co)tangent bundle of the space of connections on Y modulo gauge transformations. For $Y = S^1_V$ we first fix a basepoint; then a connection up to gauge equivalence is specified by the holonomy, an element of G. A change of basepoint conjugates the holonomy, and so the quantum Hilbert space is

$$\mathcal{H}_{\text{gauge}}(\theta{=}0) \;=\; L^2(G)^G,$$

the space of conjugacy invariant functions on G. A basis for this space is the set of characters of irreducible representations.

Next we compute the Hamiltonian. Let $A_t = A_t(x)dx$ be a connection on $\mathbb{R} \times S^1_V$ (with coordinates t, x) in temporal gauge, relative to some trivialization, and let $g_t \in G$ be the corresponding path of holonomies. In a gauge where A_t is constant, we have $g_t = e^{VA_t}$. Then the first term in the action (20) is

$$\frac{N}{2e^2} \iint \left| \frac{dA}{dt} \right| dtdx = \frac{N}{2e^2 V} \int |\dot{g}|^2 \, dt. \tag{23}$$

This is the lagrangian for a classical particle of mass N/e^2V moving on the Riemannian manifold G; the corresponding quantum Hamiltonian is

$$H_{\text{gauge}} = \frac{e^2 V}{2N} \Delta_G, \tag{24}$$

where Δ_G is the laplacian on G. The eigenfunctions are the characters of the irreducible representations with eigenvalues proportional to the Casimir.

Now consider the θ term in (20). For $G = U(k)$ we have a natural closed imaginary 1-form $\alpha \in i\Omega^1_G$ which is the trace of the Maurer-Cartan form. Then in terms of the path g_t of holonomies the θ term in the action is

$$\frac{\theta}{2\pi} \int g^* \alpha. \tag{25}$$

This is a topological term—it is invariant under reparametrizations of the path g_t. More geometrically, $\theta\alpha$ is a flat connection on a topologically trivial hermitian line bundle L_θ over G, and up to equivalence it is given by an element in $H^1(G;\mathbb{T})$. Then (25) is parallel transport in this flat bundle.[3] More generally, we can "twist" our mechanical system by any hermitian line bundle L with connection. Physically this describes a particle moving in an electromagnetic field. The quantum Hilbert space is the the space of sections of L with Hamiltonian the laplacian for such sections. In our case we obtain the space

$$\mathcal{H}_{\mathrm{gauge}}(\theta) = L^2(G, L_\theta)^G$$

of invariant sections with Hamiltonian (24). For arbitrary G an element $e^{i\theta} \in H^1(G;\mathbb{T})$ corresponds to a homomorphism $e^{i\theta}\colon \pi_1 G \to \mathbb{T}$. Recall that $\pi_1 G$ is a subgroup of the center of G. Then the eigenfunctions of the laplacian acting on $\mathcal{H}_{\mathrm{gauge}}(\theta)$ are the characters of representations of the simply connected cover of G whose restriction to $\pi_1 G$ is $e^{i\theta}$; the eigenvalue is again proportional to the Casimir.

For the unitary group $G = U(k)$ we identify $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ as before. The smallest Casimir occurs for a representation $\det^{\tilde\theta/2\pi}$, where $\tilde\theta \in \mathbb{R}$ is a representative of θ of smallest absolute value. If $\theta \neq \pi$ there is a unique such $\tilde\theta$, but for $\theta = \pi$ there are two possibilities: $\tilde\theta = \pi$ and $\tilde\theta = -\pi$. For $G = SO(k)$ the simply connected cover is $\tilde G = \mathrm{Spin}(k)$. For $\theta = 0$ there is a unique lowest representation, the trivial representation of $SO(k)$. For $\theta = \pi$ the lowest representation is the spin representation if k is odd, and the two half spin representations if k is even. In the large volume limit only the lowest eigenvalue survives, so we have two vacuua for $\theta = \pi$ in the complex $\bigl(U(k)\bigr)$ case and for k even in the real $\bigl(SO(k)\bigr)$ case. Observe that the two vacuua correspond under the involution $g \mapsto -1 \cdot g$ in these cases. (Notice also that the bundle $L_{\theta=0}$ is real.) We will see that this vacuum structure persists when we add matter. Thus parity symmetry (4) is spontaneously broken at $\theta = \pi$ for $G = U(k)$ and $G = SO(2\ell)$. This answers the third of our questions.

As the volume $V \to \infty$ the eigenvalues of (24) become widely separated. In particular, in the infinite volume limit there is no state of finite energy above the vacuum (or vacuua). So the physical Hilbert space in infinite volume consists only of the vacuum (or vacuua)—pure gauge theory on $\mathbb{R}^2$ is trivial.

Specialize to $G = U(1)$. Then the *vacuum energy density*, which is the minimum eigenvalue of the Hamiltonian (24) divided by the volume V, is

$$E_{\mathrm{vac}}(\theta) = \frac{e^2}{2N} \min_{n\in\mathbb{Z}} (n - \frac{\theta}{2\pi})^2. \tag{26}$$

Notice that the derivative of E_{vac} has a discontinuity at $\theta = \pi$.

For $G = U(k)$ formula (26) is simply multiplied by k.

The partition function $Z_\Sigma(\theta)$ of the pure gauge theory for $\Sigma = [0,T] \times S^1_V$ has the Hilbert space interpretation

$$Z_\Sigma(\theta) \sim \langle \Omega | e^{-TH(\theta)} | \Omega \rangle \quad \text{as } T \to \infty,$$

[3] In general, the classical field theory action on a manifold with boundary—here the interval—is not a number, but we do not stress that point here.

where $H(\theta)$ is the Hamiltonian and Ω the vacuum. For the pure gauge theory, we obtain from (26) for $G = U(1)$

$$Z_\Sigma(\theta) \sim \exp\bigl(cTV - \frac{e^2TV}{2N}\min_{n\in\mathbb{Z}}(n - \frac{\theta}{2\pi})^2\bigr) \quad \text{as } T \to \infty. \tag{27}$$

Here c is a constant which represents the indeterminacy of the path integral, or equivalently the fact that we are free to add a constant (independent of θ) to the Hamiltonian. This is quite different than the prediction (3) from the instanton sum. Note that due to the classical conformal invariance of the σ-model, the instantons used in deriving (3) do not have a definite size; they can be rescaled. For that reason the instanton sum is not reliable at large distances. In any case (27) answers question 2 for pure gauge theory.

§4.6. Classical Electromagnetism in Two Dimensions

Before analyzing the quantum gauge theory with bosonic matter—the ϕ field in the action (20)—we discuss the classical physics. For the classical analysis we work in the $G = U(1)$ theory. The classical equations—Maxwell's equations—for pure gauge theory are

$$\frac{N}{e^2} d_A^* F_A = 0. \tag{28}$$

Since we are in two spacetime dimensions, this implies that the electric field $f = f_A = *F_A$ is a constant. Let 'x' denote the coordinate on space, which is simply a copy of $\mathbb{R}$. Add a point charge of charge 1 at $x = x_0$. Then the electric field f as a function of x satisfies (28) with a right hand side due to the charge:

$$\frac{N}{e^2}\frac{df}{dx} = -\delta(x - x_0).$$

Thus the value of the electric field jumps by $-e^2/N$ across a charge. (See Figure 4.) Allowing for many charges, and assuming all of them are multiples of the basic charge (which is a conclusion of the quantum theory), we see that

$$\theta = \frac{2\pi N f}{e^2} \in \mathbb{R}/2\pi\mathbb{Z} \tag{29}$$

is a constant. So there is an angle in the classical theory (assuming charge quantization).

$$f \quad \uparrow \quad f - \frac{e^2}{N} \qquad\qquad\qquad f \quad \downarrow \quad f + \frac{e^2}{N}$$

FIGURE 4: CHANGE IN ELECTRIC FIELD ACROSS A POSITIVE OR NEGATIVE CHARGE

The classical energy density of an electric field f is computed from the action (23) (where $f = dA/dt$) to be $\frac{N}{2e^2} f^2$. As in the quantum theory, for a fixed value of θ in (29) there is a unique minimum obtained at some $f = f_0$ as long as $\theta \neq \pi$. Any valid configuration must have finite energy compared to the vacuum energy. Thus if $\theta \neq \pi$ we must have $f(x) \to f_0$ as $x \to \pm\infty$. Using the formula above for the jump of the electric field across a charge, we see that the total charge of a finite energy configuration is zero. This means that there is *confinement*—it is impossible to have a single charged particle or any other isolated collection of charges with nonzero total charge. On the other hand, for $\theta = \pi$ we have minimum energy density at $f_0 = \pm e^2/2N$, and so for any finite energy configuration $f(x) \to \pm e^2/2N$ as $x \to \pm\infty$. Thus there is a finite energy configuration with a single particle: the electric field satisfies $f(-\infty) = e^2/2N$, $f(\infty) = -e^2/2N$. (See Figure 5.) In this case there is no confinement. Also, in this case there are four components of finite energy configurations, depending on the value of f at $\pm\infty$.

$$\frac{e^2}{2N} \quad \uparrow \quad \frac{-e^2}{2N}$$

FIGURE 5: A ONE PARTICLE STATE FOR $\theta = \pi$

Notice that whereas in three space dimensions the Coulomb potential between point charges separated by distance r is proportional to $1/r$, the Coulomb potential in one space dimension is proportional to r. This means the potential energy grows as the charges separate, which is another way of understanding that confinement occurs.

§4.7. Quantum theory with matter

Now we want to show that there is confinement in the quantum theory of the lagrangian (20) as long as $\theta \neq \pi$. This is the assertion that every finite energy configuration in the quantum theory has total charge zero.

As a preliminary, consider the theory of matter only (no gauge field). In the simplest case $k = N = 1$ there is a single free complex scalar field ϕ with mass M. The Hilbert space $\mathcal{H}_\phi$ of this theory is the completed symmetric algebra of $W \oplus \overline{W}$, where W is the scalar representation of Poincaré with mass M. There is a global $U(1)$ symmetry which rotates ϕ. The corresponding quantum operator Q_ϕ, the Noether charge, has value 1 on W and -1 on $\overline{W}$, so value $p - q$ on $\operatorname{Sym}^p W \otimes \operatorname{Sym}^q \overline{W}$. A state in this subspace represents p positively charged particles and q negatively charged particles. For arbitrary k, N we have kN copies of this picture. In particular, for $k = 1$ there is a global $SU(N)$ symmetry (beyond the $U(1)$ symmetry already discussed.)

Now add the gauge field. For $k = N = 1$ we have a $U(1)$ gauge theory with a single charged scalar field. The global $U(1)$ symmetry of the preceding paragraph is now a local symmetry. Consider first the case $\theta = 0$. For small coupling e^2 we construct the quantum Hilbert space of the joint system by quantizing the symplectic manifold of classical solutions to (20). Thus take Σ to be the cylinder $\mathbb{R} \times S^1_V$. Then the space of classical solutions is a vector bundle over the cotangent bundle $T^*\mathcal{A}$ to the space $\mathcal{A}$ of connections on S^1_V; the fiber of this vector bundle at the trivial connection $A = 0$ is the real symplectic vector space underlying $W \otimes \overline{W}$. (Note we quantize by a complex polarization, which leads to the Hilbert space $\mathcal{H}_\phi$ described above.) To implement gauge symmetry we must take the symplectic quotient by the group $\mathcal{G}$ of gauge transformations. Fix a basepoint (infinity) on S^1_V. Then the subgroup $\mathcal{G}_0$ of gauge transformations which equal the identity at the basepoint acts freely on $\mathcal{A}$, and the quotient is identified with $U(1)$ via holonomy. In the pure gauge theory the symplectic quotient of $T^*\mathcal{A}$ by $\mathcal{G}_0$ is diffeomorphic to $T^*U(1)$; its quantization $L^2\big(U(1)\big)$ was discussed previously. As the length $V \to \infty$ recall that the only state which remains of finite energy is the vacuum state. Classically, the vacuum corresponds to the zero section, a lagrangian submanifold of $T^*U(1)$. The subgroup $U(1) \subset \mathcal{G}$ of constant gauge transformations acts trivially on $\mathcal{A}$—hence trivially on $T^*\mathcal{A}$—so does not enter into pure gauge theory. In the theory with matter the symplectic quotient by $\mathcal{G}_0$ is a vector bundle over $T^*U(1)$ with fiber $W \oplus \overline{W}$ at the identity. Now the constant gauge transformations act nontrivially in the fibers by scalar multiplication. To implement the symmetry we have two choices: we can quantize the symplectic quotient or we can consider the subspace of Hilbert space annihilated by the corresponding quantum operator N_e. Pursuing first the latter, we see that in the infinite volume limit the quantum Hilbert space before implementing the $U(1)$ is simply $\mathcal{H}_\phi$, since the gauge field contributes only the vacuum state. The operator N_e is simply equal to Q_ϕ. Thus the Hilbert space of the theory is the subspace of states with total charge $Q_\phi = 0$. Therefore, just as in the classical theory we have confinement. There is a mass gap, and the smallest mass is $2M$. In the theory with small coupling, the qualitative picture is the same.

If instead we take the symplectic quotient of $W \oplus \overline{W}$ by $U(1)$, we are led to a singular space. The moment map is $\mu(w, \bar{w}') = |w|^2 - |\bar{w}'|^2$, and $\mu^{-1}(0)$ is singular at $(0,0)$. In the quantization this singular point corresponds to the vacuum, and it is not hard to see heuristically that we are led to the same description as before.

Now allow $\theta \neq 0$. In the language of geometric quantization the prequantum line bundle over $T^*U(1)$ is now twisted by the pullback of $L_\theta \to U(1)$. Quantum states correspond to "Bohr-Sommerfeld" leaves of the given polarization, which are equally spaced parallel circles in the cylinder $T^*U(1)$. More precisely, if p is the (momentum) coordinate in the cotangent space, the circles occur at $p = n - \theta/2\pi$ for integral n. The energy of the corresponding quantum state (in pure gauge theory) is given in (26), and as $V \to \infty$ we only keep the closest circle(s) to $p = 0$, which corresponds to the vacuum state. If $\theta \neq \pi$ there is a unique such circle, and we have the same picture of confinement as above. For $\theta = \pi$ there are two such circles, corresponding to the two vacuum states. Thus in the theory with matter, before dividing out by the constant gauge transformations, we must quantize two disjoint copies of $W \oplus \overline{W}$ to obtain $\mathcal{H}_\phi \oplus \mathcal{H}_\phi$. If, as before, we assign zero charge to each of the vacuum states, then the operator N_e which corresponds to the $U(1)$ symmetry is $Q_\phi \oplus Q_\phi$. The subspace of $\mathcal{H}_\phi \oplus \mathcal{H}_\phi$ annihilated by $Q_\phi \oplus Q_\phi$ is simply two copies of the system seen previously, each copy with a vacuum. These are two "realizations" of the quantum theory, each with a mass gap of size $2M$. It turns out—and this is hard to explain from this viewpoint—that we can also assign *different* charges to the two vacuua. In that case the operator N_e is $(Q_\phi - 1) \oplus Q_\phi$ or $Q_\phi \oplus (Q_\phi + 1)$. The kernel in each case has one particle states of mass M, and there is no confinement. There is no vacuum state in either realization. The four realizations correspond to the classical picture of the previous section.

To justify these heuristic pictures we compute. From (20) we see that the Hamiltonian is

$$H = \int_{-\infty}^{\infty} dx\, \big(\frac{1}{2}|\pi_\phi|^2 + |d_A\phi|^2 + M^2|\phi|^2 + \frac{N}{2e^2}|f_A|^2\big), \tag{30}$$

where ϕ is a field on $\mathbb{R}$. Here π_ϕ is the conjugate momentum to ϕ, and $f = f_A$ is the Hodge star of the curvature as before. We derive an effective Hamiltonian for ϕ by plugging in the equation of motion of A. The latter is obtained by varying the lagrangian (20) with respect to A:

$$\frac{N}{e^2}\frac{df}{dx} = j, \tag{31}$$

where the current is

$$j = \phi\pi_\phi - \overline{\phi\pi_\phi}.$$

We integrate (31) to obtain

$$f(x) = \frac{e^2}{N}\left\{\int_{-\infty}^{\infty} dy\, G(x-y)j(y) + c\right\} \tag{32}$$

for some constant c, where

$$G(x-y) = \begin{cases} 1, & x \geq y; \\ 0, & x < y. \end{cases}$$

Note that

$$\lim_{x\to-\infty} f(x) = \frac{e^2 c}{N}. \tag{33}$$

Plugging into (30) we see

$$H = H_0 + \Delta H,$$

where H_0 is the Hamiltonian for the free scalar particle (if we compute at $A = 0$), and

$$\Delta H = \frac{e^2}{2N}\int_{-\infty}^{\infty} dx\, \left\{\int_{-\infty}^{\infty} dy\, G(x-y)j(y) + c\right\}^2.$$

The perturbation term ΔH is nonlocal and singular.

Now we must determine the subspace of states Ψ on which ΔH is finite. Consider first $c = 0$. Then $\langle\Psi|\Delta H|\Psi\rangle$ is finite if and only if

$$\lim_{x\to\infty} \langle\Psi|f(x)|\Psi\rangle = 0,$$

where $f(x)$ is defined by (32). Now

$$\lim_{x\to\infty} \int_{-\infty}^{\infty} dy\, G(x-y)j(y) = \int_{-\infty}^{\infty} dy\, j(y) = Q_\phi$$

is the charge operator in the Hilbert space $\mathcal{H}_\phi$. So if $c = 0$ we have confinement:

$$\langle\Psi|Q_\phi|\Psi\rangle = 0.$$

In general, c is related to θ through (33) and (29):

$$|c| = \min_n \left(\frac{\theta}{2\pi} + n\right).$$

This is the value of the electric field at $-\infty$. Also, we subtract an (infinite) constant from ΔH to account for the nonzero energy at ∞:

$$(\Delta H)_{\text{normalized}} = \frac{e^2}{2N} \int_{-\infty}^{\infty} dx \left[\left\{\int_{-\infty}^{\infty} dy\, G(x-y)j(y) + c\right\}^2 - c^2\right].$$

This is finite on states Ψ which satisfy

$$Q_\phi(Q_\phi + 2c)|\Psi\rangle = 0, \tag{34}$$

i.e., $Q_\phi|\Psi\rangle = 0$ or $Q_\phi|\Psi\rangle = -2c$. For $\theta \neq \pi$ the only possibility is $Q_\phi|\Psi\rangle = 0$ and we have confinement. For $\theta = \pi$ we have either $c = \pm 1/2$, and (34) is satisfied by states with $Q_\phi|\Psi\rangle = 0$ or $Q_\phi|\Psi\rangle = \mp 1$. As in the classical theory, this gives four sectors. We denote them $\mathcal{H}_{++}, \mathcal{H}_{+-}, \mathcal{H}_{-+}, \mathcal{H}_{--}$ according to the value of the electric field at $-\infty$ and $+\infty$. There is a vacuum state and confinement in $\mathcal{H}_{++}, \mathcal{H}_{--}$; there is neither a vacuum nor confinement in $\mathcal{H}_{+-}, \mathcal{H}_{-+}$.

In the confining cases the symmetry group $PSU(N)$ acts. In the sectors of the $\theta = \pi$ theory with no confinement the covering group $SU(N)$ acts.

The story for $k > 1$ is similar.

Finally, we make a remark about the electric charge. In the quantum picture it is an operator Q_e, and from Noether's theorem applied to (20) we compute the relation to N_e:

$$Q_e = \frac{e^2}{N}(N_e + \frac{\theta}{2\pi}).$$

The eigenvalues of N_e are integral, but those of Q_e are shifted from $\frac{e^2}{N}\mathbb{Z}$ if $\theta \neq 0$. Note the flow (monodromy) in the eigenvalues of Q_e as θ runs around the circle from $\theta = 0$ to $\theta = 2\pi$. We will encounter this phenomenon again in four dimensional gauge theory.