LECTURE II-8, PART II: ABELIAN DUALITY IN FOUR DIMENSIONS AND $\mathrm{Sl}(2,\mathbb{Z})$

EDWARD WITTEN*

1. Duality and $\mathrm{Sl}(2,\mathbb{Z})$

In this second part of lecture II-8, we discuss abelian duality in four dimensions, and give an application to an $\mathrm{Sl}(2,\mathbb{Z})$ symmetry of the free $U(1)$ theory in four dimensions. We postpone discussion of $\mathrm{Sl}(2,\mathbb{Z})$ symmetries of non-free theories to a later lecture, since all known examples of that involve supersymmetry.

We work with a $U(1)$ bundle $\mathscr{L}$ on a 4-manifold M, and a connection A on $\mathscr{L}$, whose curvature is $F{=}F_A$. The gauge theory Lagrangian (in Euclidean signature) including the topological term is

$$\begin{aligned} \mathcal{L}(A) &= \frac{1}{4\mathrm{e}^2}\int d^4x\,\sqrt{g}F_{mn}F^{mn} + \frac{i\theta}{16\pi^2}\int d^4x\,\sqrt{g}\epsilon_{mnpq}F^{mn}F^{pq} \\ &= \frac{1}{2\mathrm{e}^2}\int F_A\wedge *F_A + \frac{i\theta}{4\pi^2}\int F_A\wedge F_A. \end{aligned} \tag{1.1}$$

We have used the standard normalization on the kinetic term, and have normalized the topological term so that replacing θ by $\theta{+}2\pi$ does not change the physics. (This property of the topological term derives from the fact that $c_1(\mathscr{L})^2 = \int(F_A/2\pi)\wedge(F_A/2\pi)$ is always an integer. Notice that on a spin manifold, $c_1(\mathscr{L})^2$ is always an *even* integer, and we gain an additional equivalence under replacement of θ by $\theta+\pi$.)

Let $\tau = \frac{\theta}{\pi} + \frac{2\pi i}{e^2} \in \mathfrak{h}$. As we have just observed, $\tau \mapsto \tau+2$ is a symmetry of this theory, and $\tau\mapsto\tau+1$ is a symmetry when working on a spin manifold. To extend this to an $\mathrm{Sl}(2,\mathbb{Z})$ action (in the spin manifold case) we also need a symmetry which maps τ to $-1/\tau$; this will be given by a *duality* transformation $F_A \leftrightarrow *F_C$ (with C being a new "dual" connection).

The computations for this duality transformation are similar to those in lecture II-7. We begin by defining $F_\pm = \frac{1}{2}(F_A \pm *F_A)$, and rewriting our Lagrangian (1.1) as

$$\begin{aligned} \mathcal{L}(A) &= \frac{i\overline{\tau}}{4\pi}\int F_+\wedge F_+ + \frac{i\tau}{4\pi}\int F_-\wedge F_- \\ &= \frac{i\overline{\tau}}{4\pi}\int \|F_+\|^2 - \frac{i\tau}{4\pi}\int \|F_-\|^2. \end{aligned} \tag{1.2}$$

Date: 6 March 1997.

*Notes by David R. Morrison.

Letting $\mathcal{G}$ denote the gauge group associated to A, the partition function for this theory can be written as

$$Z(\tau) = \frac{1}{\mathrm{vol}(\mathcal{G})} \sum_{\mathscr{L}} \int \mathcal{D}A\, e^{-\frac{i\overline{\tau}}{4\pi}\int \|F_+\|^2 + \frac{i\tau}{4\pi}\int \|F_-\|^2}. \tag{1.3}$$

Our earlier examples of duality began with a theory of a scalar field ϕ which entered into the Lagrangian only through its derivative $d\phi$ so that the theory had a symmetry under $\phi \mapsto \phi + c$ (with c constant); the first step in the duality transformation was to gauge this symmetry, introducing also an appropriate Lagrange multiplier field.

The present theory is already a gauge theory, being a theory of a connection A which enters into the Lagrangian only through its curvature F_A, so that there is a symmetry under $A \mapsto A + B$ (with B a flat connection). We want to do the analogue of gauging this symmetry, by allowing B to be an arbitary connection on an arbitrary bundle, introducing a kind of "exotic gauge field" G which is a 2-form field, and extending the symmetry to

$$\begin{aligned} A &\to A + B \\ G &\to G + F_B. \end{aligned} \tag{1.4}$$

Then $\mathcal{F} := F_A - G$ plays the role of the "gauge-invariant" quantity, analogous to the covariant derivative of a scalar field. It is to be stressed that two G fields will be considered gauge-equivalent if they differ by $G \to G + F_B$ for F_B the curvature of any connection on any line bundle. In our analysis, we will assume for simplicity that there is no torsion in $H^2(M)$.

We need a "gauge-invariant" extension of our Lagrangian. We might try

$$\mathcal{L}(A, G) = \frac{i\overline{\tau}}{4\pi}\int \|\mathcal{F}_+\|^2 - \frac{i\tau}{4\pi}\int \|\mathcal{F}_-\|^2, \tag{1.5}$$

but this is too simple (because, for example, we could gauge $\mathcal{F}$ to zero). To improve this, we introduce a new connection C on a line bundle $\mathscr{N}$, with curvature F_C, and consider the Lagrangian

$$\mathcal{L}(A, G, C) = \frac{i\overline{\tau}}{4\pi}\int \|\mathcal{F}_+\|^2 - \frac{i\tau}{4\pi}\int \|\mathcal{F}_-\|^2 - \frac{i}{2\pi}\int F_C \wedge G. \tag{1.6}$$

The partition function for this new theory can be represented as a path integral, which includes sectors associated to all choices of bundles $\mathscr{L}$ and $\mathscr{N}$:

$$\frac{1}{\mathrm{vol}(\widetilde{\mathcal{G}})}\,\frac{1}{\mathrm{vol}(\mathcal{G})}\,\frac{1}{\mathrm{vol}(\mathcal{G}_C)} \sum_{\mathscr{L},\mathscr{N}} \int \mathcal{D}A\,\mathcal{D}G\,\mathcal{D}C\, e^{-\frac{i\overline{\tau}}{4\pi}\int \|\mathcal{F}_+\|^2 + \frac{i\tau}{4\pi}\int \|\mathcal{F}_-\|^2 + \frac{i}{2\pi}\int F_C\wedge G}, \tag{1.7}$$

where $\mathcal{G}$ and $\mathcal{G}_C$ denote the gauge groups associated to A and C, respectively, and $\widetilde{\mathcal{G}}$ denotes the "exotic" gauge group.

To see that this new theory is equivalent to the original one, we first do the C-integral in (1.7): write $C = C_0 + C'$, for C_0 a fixed connection on the line bundle $\mathscr{N}$. Then the

C' integral is

$$\frac{1}{\mathrm{vol}\,\mathcal{G}_C}\int \mathcal{D}C'\, e^{\frac{i}{2\pi}\int C'\wedge dG} = \delta(dG). \tag{1.8}$$

Thus, when we sum over $\mathscr{N}$ we find

$$\frac{1}{\mathrm{vol}\,\mathcal{G}_C}\sum_{\mathscr{N}}\int \mathcal{D}C e^{-\frac{i}{2\pi}\int F_C\wedge G} = \sum_{x\in H^2(M)} e^{i(x,G)}\delta(dG) = \delta(\left[\frac{G}{2\pi}\right]\in\mathbb{Z})\delta(dG). \tag{1.9}$$

The conditions that $dG=0$ and that $[G/2\pi]$ is an integral class precisely mean that G is of the form F_B for some connection on some line bundle and hence that G can be gauged to zero. After doing this, it follows that the partition function (1.7) coincides with $Z(\tau)$, and we recover the original theory.

Alternatively, we can evaluate the partition function (1.7) by gauging A to 0, using the "exotic" gauge invariance (which has an ordinary gauge invariance as an ambiguity). This leaves the path integral

$$\frac{1}{\mathrm{vol}\,\mathcal{G}_C}\sum_{\mathscr{N}}\int \mathcal{D}G\int \mathcal{D}C\, e^{-\frac{i\overline{\tau}}{4\pi}\int \|G_+\|^2+\frac{i\tau}{4\pi}\int\|G_-\|^2+\frac{i}{2\pi}\int F_C\wedge G}. \tag{1.10}$$

To evaluate the G integral, we complete the square, bearing in mind that

$$\int F_C\wedge G = \int (F_{C+}\wedge *G_+ - F_{C-}\wedge *G_-). \tag{1.11}$$

In fact, if we define $G' = G - \frac{1}{\overline{\tau}}F_{C+} + \frac{1}{\tau}F_{C-}$, then we can write the exponent from eq. (1.10) as

$$-\frac{i\overline{\tau}}{4\pi}\int\|G'_+\|^2 + \frac{i\tau}{4\pi}\int\|G'_-\|^2 + \frac{i}{4\pi\overline{\tau}}\int\|F_{C+}\|^2 - \frac{i}{4\pi\tau}\int\|F_{C-}\|^2. \tag{1.12}$$

When we carry out the G' integral, the first two terms give a Gaussian integral which contributes to the overall normalization; integrating out G' leaves the path integral

$$\frac{1}{\mathrm{vol}\,\mathcal{G}_C}\sum_{\mathscr{N}}\int \mathcal{D}C\, e^{\frac{i}{4\pi\overline{\tau}}\int\|F_{C+}\|^2-\frac{i}{4\pi\tau}\int\|F_{C-}\|^2}. \tag{1.13}$$

This is the same as the original path integral, but with τ replaced by $-1/\tau$, precisely what we wanted to show.

As we did in the case of two dimensions, it is possible to analyze the τ-dependence of the normalization of the path-integral, and obtain further interesting results. Some hint of the flavor of the results to be obtained this way is seen if we evaluate the Gaussian integral indicated above, which yields

$$\left(\frac{2\pi}{\sqrt{i\overline{\tau}}}\right)^{n_{2+}}\left(\frac{2\pi}{\sqrt{-i\tau}}\right)^{n_{2-}}, \tag{1.14}$$

where $n_{2\pm}$ denote the numbers of self-dual and anti-self-dual 2-forms. Of course, these numbers are infinite, so there must be some cancellation against other normalization factors. When this is worked out in detail,[1] the result is found to be

$$Z(\tau) = \tau^{-\frac{\chi+\sigma}{4}} \overline{\tau}^{-\frac{\chi-\sigma}{4}} Z(-1/\tau), \tag{1.15}$$

where χ and σ are the Euler number and signature of M, respectively. Thus, the partition function $Z(\tau)$ is actually a modular form for $\mathrm{Sl}(2, \mathbb{Z})$ (or for a subgroup, when the manifold is not spin) of weight $(\frac{\chi+\sigma}{4}, \frac{\chi-\sigma}{4})$.

We can also follow certain operator insertions through the duality transformation, as we did in lower dimensions. An insertion of $F_\pm$ in the original theory can be realized by inserting $\mathcal{F}_\pm = F_\pm - G$ in the extended theory, which can be written

$$\mathcal{F}_+ = F_+ - G'_+ - \frac{1}{\overline{\tau}} F_{C+}, \text{ or } \mathcal{F}_- = F_- - G'_- + \frac{1}{\tau} F_{C-}, \tag{1.16}$$

respectively, after making the change of variables to G'. Thus, when we gauge A to zero, and integrate out G', we are left with operator insertions proportional to $F_{C\pm}$, namely:

$$F_+ \mapsto (-1/\overline{\tau}) F_{C+}, \text{ and } F_- \mapsto (1/\tau) F_{C-}. \tag{1.17}$$

Notice that as a consequence of the τ-dependence of these mappings, a correlation function involving insertions of F_+ and F_- will have a different modular weight than that of the partition function.

2. The Hamiltonian formalism

Returning to the case that the gauge group is $U(1)$, let us briefly discuss abelian four-dimensional duality in a Hamiltonian framework. Take a 4-manifold of the form $M_3 \times \mathbb{R}$, where $\mathbb{R}$ is a timelike direction. Note that this is a spin manifold, so we expect full $Sl(2, \mathbb{Z})$ symmetry. For simplicity we suppose that there is no torsion in $H_1(M_3)$. Each class $x \in H^2(M_3)$ determines a complex line bundle $\mathscr{L}_x$ on the 3-manifold M_3 (satisfying $c_1(\mathscr{L}_x) = x$). The Hilbert space for our theory on the 3-manifold M_3 can be written in the form

$$\mathcal{H}_\tau(M_3) = \bigoplus_{x \in H^2(M_3, \mathbb{Z})} \mathcal{H}_x, \tag{2.1}$$

where $\mathcal{H}_x$ is the Hilbert space which comes from quantizing connections on $\mathscr{L}_x$. (On the left, we have explicitly indicated the dependence on the coupling constant τ.) To construct $\mathcal{H}_x$, write an arbitrary connection in the form $A = A_0 + \beta$, where A_0 is a harmonic connection (a connection whose curvature is a harmonic two-form) and β is a 1-form which is co-closed. Let $\mathcal{T}_x$ be the space of harmonic connections on the line bundle $\mathcal{L}_x$. Then the quantization yields

$$\mathcal{H}_x = \mathcal{H}_\beta \otimes L^2(\mathcal{T}_x). \tag{2.2}$$

Here $\mathcal{H}_\beta$ is a Hilbert space obtained by quantizing the space of β's, and $L^2(\mathcal{T}_x)$ is just the space of L^2 functions on $\mathcal{T}_x$.

[1] E. Witten, *On S-duality in abelian gauge theory*, Selecta Math (N.S.) **1** (1995), 383–410.

Note that the factor $\mathcal{H}_\beta$ is independent of x, since the space of co-closed one-forms is defined with no reference to x. Duality maps $\mathcal{H}_\beta$ to itself while acting separately on $\mathcal{H}' = \oplus_x L^2(\mathcal{T}_x)$. The duality action on $\mathcal{H}_\beta$ follows from the operator mapping in (1.17).

The action of duality on $\mathcal{H}'$ can be described as follows. Note that the $\mathcal{T}_x$'s are all principal homogeneous spaces acted on by the torus $H^1(M_3, \mathbb{R}/\mathbb{Z})$, which parametrizes flat line bundles on M_3; the action is defined by tensoring any given line bundle with connection by a flat line bundle determined by a point in $H^1(M_3, \mathbb{R}/\mathbb{Z})$. Let y denote a character of the abelian group $H^1(M_3, \mathbb{R}/\mathbb{Z})$. There is a decomposition $L^2(\mathcal{T}_x) = \oplus_y \mathcal{T}_{x,y}$, where $\mathcal{T}_{x,y}$ is the subspace of $L^2(\mathcal{T}_x)$ transforming in the character y. Each $\mathcal{T}_{x,y}$ is one-dimensional. Hence

$$\mathcal{H}' = \oplus_{x,y} \mathcal{T}_{x,y} \tag{2.3}$$

Note that, by Poincaré and Pontryagin duality, the character group of $H^1(M_3, \mathbb{R}/\mathbb{Z})$ is $H^2(M_3, \mathbb{Z})$. Thus, x and y take values in the same space. It is hence possible to exchange them, and this is what the $\tau \to -1/\tau$ transformation does (more precisely, it acts by $(x, y) \to (-y, x)$). Thus duality exchanges a classical notion – the decomposition with respect to x – with a quantum notion – the decomposition with respect to y. The claim about how the duality acts will be justified below where we introduce the operators Q_E and Q_M.

Upon quantization—and suppressing θ for a moment—one writes the four-dimensional curvature as $F'_A + e^2 \pi_A dt$, where F'_A is a two-form on M_3 and π_A—a one-form on M_3—is the momentum conjugate to the connection A. The Hamiltonian becomes

$$H = \frac{1}{2\mathrm{e}^2} \int F_{A_0}^2 + \frac{\mathrm{e}^2}{2} \nabla_{A_0} + H(\beta). \tag{2.4}$$

Here $H(\beta)$ is the part of the Hamiltonian that acts on $\mathcal{H}_\beta$. The other terms act on $\mathcal{H}'$. The first term is the magnetic energy of the harmonic connection A_0; it comes from the part of the Lagrangian quadratic in F'_A and is a multiple of $\int_{M_3} x \wedge * x$. The second term, which comes from the part of the Lagrangian quadratic in π_A, is the electric energy, the Laplacian on $\mathcal{T}_x$; it is a multiple of $\int_{M_3} y \wedge * y$.

Including the θ term shifts the quantization. In fact, the canonical momentum $F_A^\vee = * \pi_A$ as determined from the original Lagrangian (1.1) is

$$F_A^\vee = 2\pi i \frac{\delta S}{\delta F_A} = \frac{2\pi i}{\mathrm{e}^2} * F_A - \frac{\theta}{\pi} F_A \tag{2.5}$$

At non-zero θ, one has not $\mathcal{H}' = \oplus_x L^2(\mathcal{T}_x)$ but $\mathcal{H}' = \oplus_x \Gamma_{L^2}(\mathcal{T}_x, \mathcal{S}_\theta)$, where $\mathcal{S}_\theta$ is a certain flat line bundle over $\mathcal{T}_x$ and Γ_{L^2} is the space of L^2 sections. I leave it as an exercise to the reader to identify $\mathcal{S}_\theta$. Of course, $\mathcal{S}_\theta$ is trivial at $\theta = 0$ and only depends on θ modulo 2π.

Rewriting the formula for $F_A^\vee$ in terms of τ, we can determine how this operator transforms under $\tau \to -1/\tau$. Indeed, under the operator mapping (1.17) one gets

$$F_A^\vee = -\overline{\tau} F_+ + \tau F_- \mapsto F_{C+} + F_{C-} = F_C. \tag{2.6}$$

We will also need the "dual" version of this computation:

$$F_A = F_+ + F_- \mapsto \overline{(-1/\tau)} F_{C+} - (-1/\tau) F_{C-} = -F_C^\vee. \tag{2.7}$$

In the Hamiltonian formalism, for any 2-cycle $\Sigma \subset M_3$ we can define associated "electric" and "magnetic" operators on the Hilbert space $\mathcal{H}_\tau(M_3)$ for any 2-cycle $\Sigma \subset M_3$, by

$$\begin{aligned} Q_E(\Sigma) &= \int_\Sigma \frac{F_A^\vee}{2\pi} = \int_\Sigma \frac{F_C}{2\pi} \\ Q_M(\Sigma) &= \int_\Sigma \frac{F_A}{2\pi} = -\int_\Sigma \frac{F_C^\vee}{2\pi}. \end{aligned} \tag{2.8}$$

(These operators only depend on the class of Σ in $H_2(M_3, \mathbb{Z})$.) Clearly, under $\tau \to -1/\tau$, that is under $A \to C$, one has $Q_E \to Q_M$, $Q_M \to -Q_E$. Since x and y are the eigenvalues of Q_M and Q_E, this means that $\tau \to -1/\tau$ acts by $(x, y) \to (-y, x)$. Moreover, from the explicit formula (2.5) for $F_A^\vee$, one sees that when θ is increased by 2π (an operation which leaves the Hilbert space unchanged), the operator Q_M is unaltered, but the operator Q_E maps to $Q_E + Q_M$.

The statements made in the last paragraph can be combined to the following: $Sl(2, \mathbb{Z})$ acts on $\mathcal{H}'$ via the natural action of $Sl(2, \mathbb{Z})$ on the pair (x, y).