Generalized Symmetries in Field Theory

Riccardo Argurio
Université Libre de Bruxelles

Saalburg Summer School, September 2026

Contents

1 Introduction

In these lectures we will discuss symmetries in Quantum Field Theories (QFTs). It is difficult to over-emphasize how much symmetries are important in the study of QFTs. They are at the same time crucial in the definition of (classes of) QFTs, and the most useful tool to study and classify the observable physical effects that such theories describe. In the following, we will try to be as generic as possible concerning the QFTs we want to discuss. One should bear in mind that QFTs are the basic framework for high energy physics, and also for many condensed matter/many-body systems. They are also an extremely useful tool in mathematical physics, or mathematics in general, especially when one considers families of QFTs, including QFTs on arbitrary spacetime manifolds. This is why we will most of the time think of QFTs which have little relation to the most phenomenologically sound ones..

We will actually start from classical Field Theory, and a good starting point for our journey is Noether's theorem. A symmetry of a classical field theory is usually understood in terms of the action S from which the equations of motion (EOM) of the theory follow. A symmetry is a transformation of the fields, that we can generically indicate here as φ

$$\varphi \to \varphi' = R(\varphi) \tag{1.1}$$

such that the action is invariant

$$S[\varphi'] = S[\varphi] \ . \tag{1.2}$$

To be more precise, the fields $\varphi(x)$ are functions from the spacetime manifold $\mathcal{M}$ to a vector space given by a representation of the symmetries of $\mathcal{M}$. The action $S[\varphi]$ is a functional over a sensible set of such functions φ, that yields a real number for Lorentzian QFTs, but generically a complex number for Euclidean QFTs. For local QFTs, S is usually written as an integral over

$\mathcal{M}$ of a spacetime density, the Lagrangian $\mathcal{L}[\varphi]$, which depends only φ and its derivatives, evaluated all at the same point x.

The above is all that we need to know before deriving Noether's theorem (which we will do in the next section). Things get more interesting when going to Quantum Field Theory. In QFT, the main object is still the action $S[\varphi]$, but in that it appears as a weight in the path integral, which is a formal integration over the set of functions φ

$$Z = \int \mathcal{D}\varphi \; e^{iS[\varphi]} \qquad \text{[Lorentzian]} \tag{1.3}$$

or

$$Z = \int \mathcal{D}\varphi \; e^{-S[\varphi]} \qquad \text{[Euclidean]} \tag{1.4}$$

Z is the partition function, it is just a number, but we will see that we can make it depend on some parameters so that it can potentially give all the observables of the QFT. A transformation of the fields φ is a symmetry of the QFT if Z is unchanged. This is at the same time stronger and weaker than the classical requirement. Indeed we see that $S[\varphi'] = S[\varphi]$ is not enough, we also need the measure to be invariant $\mathcal{D}\varphi = \mathcal{D}\varphi'$. Or, to be more precise, we need $\mathcal{D}\varphi \; e^{-S}$ to be invariant. This means that strictly speaking, S need not be exactly invariant, shifts in $2\pi i\mathbb{Z}$ are allowed. We will see examples of this.

Now that we have defined what are symmetries of a QFT, we should be more specific on what are the transformations acting on the fields φ. Here we are immediately brought to the branch of mathematics that originates in the study of symmetries, namely group theory, and algebra. Symmetries are naturally associated to group elements, and the fields transforming under them will belong to representations of such groups. The main goal of these lectures is to actually see that symmetries of QFTs are more general than groups, but groups still give a very good intuition of how symmetries act in a QFT. For instance, we will distinguish between finite, discrete, and continuous symmetries, and between abelian and non-abelian ones.

Symmetries not only act on the fundamental fields through which one defines a QFT, but also on all its observable objects: typically functionals of the fundamental fields. Equivalently, one can think of the spectrum of the theory (which as we will see, consists both of particle-like states and extended defects). All physical observables must transform under the symmetries, namely they must be in a representation of the group (when the symmetries indeed form a group).

Let us finally mention that symmetries are also important when they are not there, namely when they are broken, either explicitly (the QFT is not invariant under the symmetry), or spontaneously (the QFT is invariant but the vacuum is not). What breaks the symmetry is usually very much constrained by the symmetry itself.

These lecture notes summarize, with a large bias towards the personal interests of the author, recent research that originates from the seminal paper:

- D. Gaiotto, A. Kapustin, N. Seiberg and B. Willett, "Generalized Global Symmetries," JHEP **02** (2015), 172 [arXiv:1412.5148 [hep-th]].

By now, many lecture notes exist that are accessible on the arXiv. A selection of the ones that are perhaps closest in spirit to the present ones, namely with more emphasis on basic QFT examples from high energies, is the following:

- P. R. S. Gomes, "An introduction to higher-form symmetries," SciPost Phys. Lect. Notes **74** (2023), 1 [arXiv:2303.01817 [hep-th]].

- T. D. Brennan and S. Hong, "Introduction to Generalized Global Symmetries in QFT and Particle Physics," [arXiv:2306.00912 [hep-ph]].

- N. Iqbal, "Jena lectures on generalized global symmetries: principles and applications," [arXiv:2407.20815 [hep-th]].

2 From Noether's theorem to topological defects

Let us consider a field theory T described by an action $S[\varphi]$ with φ a collection of fields, possibly in a non-trivial representation of the spacetime symmetries. To keep things simple, we will start with d-dimensional Minkowski spacetime. The action can be written then as as integral over such spacetime

$$S[\varphi] = \int d^d x \ \mathcal{L}[\varphi] \ . \tag{2.1}$$

We say that T has the group G as a symmetry if the fields φ are in a representation R of G such that when

$$\varphi \to \varphi' = R(g)\varphi \ , \qquad g \in G \tag{2.2}$$

the action is invariant

$$S[\varphi'] = S[\varphi] \ . \tag{2.3}$$

Such symmetries are usually called internal, to make the distinction with respect to spacetime symmetries (i.e. transformations under the Lorentz group), and global, to make the distinction with respect to local ones, namely when the transformation depends on the spacetime location. The latter are also called gauge symmetries, and are different in nature. Global symmetries tell that two different configurations of the fields are such that the physics of one configuration can be derived from the physics of the other. Local, or gauge symmetries, actually reduce the space of field configurations by identifying the points that are mapped to each other. We will see examples of both symmetries. But it is important to bear in mind that the main object of study of these lectures are global symmetries (which sometimes arise in theories with gauge symmetries—also, gauge symmetries are global symmetries that are made local).

Another important remark is that we have assumed above that the fields transform linearly under the symmetry. This is not always the case, but one can always find some object in the theory (a composite of the fundamental fields φ) that transforms linearly in a representation of G.

To rederive the well-known Noether theorem, we now focus on the case in which G is a continuous (Lie) group. We can then consider transformations that are close to the identity, i.e. such that

$$R(g) = 1 + \alpha \rho_R(X) \tag{2.4}$$

with X an element of the algebra $\mathfrak{g}$ of G, ρ_R the corresponding representation of $\mathfrak{g}$, and α a small continuous parameter. The field transforms as

$$\varphi' = \varphi + \alpha \rho_R(X)\varphi \ . \tag{2.5}$$

It is useful to consider the variation of the field

$$\delta\varphi = \varphi' - \varphi = \alpha \rho_R(X)\varphi = \alpha \Delta\varphi \ . \tag{2.6}$$

The theory is invariant under the symmetry if

$$\delta S = S[\varphi'] - S[\varphi] = 0 \ . \tag{2.7}$$

We have

$$\begin{aligned} \delta S &= \int d^d x \ \delta\mathcal{L}[\varphi] \\ &= \int d^d x \ \left(\frac{\delta\mathcal{L}}{\delta\varphi} \alpha\Delta\varphi + \frac{\delta\mathcal{L}}{\delta\partial_\mu\varphi} \alpha\partial_\mu\Delta\varphi \right) \\ &= \int d^d x \ \alpha\partial_\mu \mathcal{V}^\mu \end{aligned} \tag{2.8}$$

where in the second line we have used the fact that α is a constant in space-time, and the third implements the vanishing of δS by allowing the variation to be proportional to a total derivative, which vanishes if the (Euclidean)

spacetime manifold is compact, or in the non-compact case, upon restricting to field configurations that die-off sufficiently fast at infinity.

Take now α to be no longer a constant but a function of spacetime $\alpha(x)$. The relation above will still be true, up to terms that involve $\partial_\mu \alpha$:

$$\begin{aligned} \delta S &= \int d^d x \ \left(\partial_\mu \alpha \frac{\delta \mathcal{L}}{\delta \partial_\mu \varphi} \Delta\varphi + \alpha \partial_\mu \mathcal{V}^\mu \right) \\ &= \int d^d x \ \alpha \ \partial_\mu \left(-\frac{\delta \mathcal{L}}{\delta \partial_\mu \varphi} \Delta\varphi + \mathcal{V}^\mu \right) \end{aligned} \tag{2.9}$$

where in the second line we have integrated by parts the first term (assuming again the boundary terms vanish). We will call

$$J^\mu = -\frac{\delta \mathcal{L}}{\delta \partial_\mu \varphi} \Delta\varphi + \mathcal{V}^\mu \ , \tag{2.10}$$

the Noether current. Now, we notice that $\delta\varphi = \alpha(x)\Delta\varphi$ is a particular instance of a general (small) variation, of the kind one takes to derive the equations of motion. The latter are satisfied precisely at the stationary points under such variations. It means that when the equations of motion are satisfied, we must have $\delta S = 0$, also for a variation as above, for any $\alpha(x)$. This then implies that we have the local condition

$$\partial_\mu J^\mu = 0 \tag{2.11}$$

which is valid on-shell. J^μ is thus a *conserved current*.

In the generic case where the algebra underlying the symmetry has many dimensions, we can associate to every generator t^a of $\mathfrak{g}$ a conserved current J^μ_a.

Every conserved current leads to a conserved charge

$$Q_a(t) = \int_{\Sigma_t} d^{d-1}x \ J^t_a \tag{2.12}$$

where Σ_t is a constant time slice of the spacetime, which we take to be Minkowski for now. The conservation of the current implies the time independence of the charge

$$\partial_t Q_a(t) = \int_{\Sigma_t} d^{d-1}x \; \partial_t J_a^t = \int_{\Sigma_t} d^{d-1}x \; \partial_i J_a^i = 0 \; . \tag{2.13}$$

This implies that $Q_a(t) = Q_a$ is a constant, namely it does not change if the time slice Σ_t on which it is evaluated is translated (forward or backward) in time. This of course assumes that the equations of motion are satisfied in all the spacetime region swept by the translation of the time slice, namely there are no localized sources.

Charges, like currents, belong to the algebra $\mathfrak{g}$. Group elements are built by exponentiation, namely $e^{i\alpha_a Q_a}$ with α_a some continuous parameters. In the following, we will mostly focus on one-dimensional (abelian) algebras, hence we will drop the index a.

We can reformulate Noether's theorem in a way that makes the topological nature of the charges more manifest. We need the language of differential forms, that we assume to be known. In this language, a conserved current is a 1-form $J = J_\mu dx^\mu$ which satisfies

$$d * J = 0 \tag{2.14}$$

where $*$ is the Hodge star, that in d-dimensional spacetime generically maps p-forms to $(d-p)$-forms. Now the charge, instead of being defined on a constant time slice, is more generally defined as the integral over a $(d-1)$-dimensional submanifold Σ_{d-1} in spacetime of the $(d-1)$-form $*J$:

$$Q(\Sigma_{d-1}) = \int_{\Sigma_{d-1}} *J \; . \tag{2.15}$$

We want to compare the charge defined on Σ_{d-1} with the charge defined on a slightly deformed submanifold Σ'_{d-1}, such that

$$\Sigma'_{d-1} - \Sigma_{d-1} = \partial \mathcal{B}_d \tag{2.16}$$

with $\mathcal{B}_d$ a d-dimensional region of $\mathcal{M}_d$ with the topology of a ball (namely, contractible to a point). In other words, Σ'_{d-1} is just Σ_{d-1} with a bump on it. We then evaluate

$$Q(\Sigma'_{d-1}) - Q(\Sigma_{d-1}) = \int_{\partial\mathcal{B}_d} *J = \int_{\mathcal{B}_d} d * J = 0 \tag{2.17}$$

where we used Stokes theorem in the middle equality. Since $Q(\Sigma_{d-1})$ does not depend on small deformations of Σ_{d-1}, it is a *topological* quantity since it will depend only on global, topological properties of the manifold Σ_{d-1} on which it is defined. For instance, if in order to deform Σ_{d-1} to Σ'_{d-1}, one has to pass through a singular point of $\mathcal{M}_d$, then clearly the charge is allowed to change. Finally, the topological nature of Q also applies to other objects built from it, such as

$$U_\alpha(\Sigma_{d-1}) = e^{i\alpha Q(\Sigma_{d-1})} \ . \tag{2.18}$$

Going now to quantum field theory, the idea is that the classical action $S[\varphi]$ appears in the path integral

$$Z = \int \mathcal{D}\varphi \ e^{iS[\varphi]} \ . \tag{2.19}$$

More generally, the latter is a way to compute (time-ordered) correlation functions of operators in the QFT

$$\langle \hat{\mathcal{O}}_1(x_1) \dots \hat{\mathcal{O}}_n(x_n) \rangle = Z^{-1} \int \mathcal{D}\varphi \ \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) e^{iS[\varphi]} \ . \tag{2.20}$$

$\mathcal{O}_i$ are local composite expressions of the fields φ, and $\hat{\mathcal{O}}_i$ are the associated quantum operators. The classical conserved current J^μ is associated to a quantum operator $\hat{J}^\mu$, and the question is whether the conservation equation holds as an operator identity $\partial_\mu \hat{J}^\mu = 0$. In other words, does an insertion of $\partial_\mu \hat{J}^\mu$ in any n-point function make the latter vanish?

The answer [**Exercise 1**] is a relation between correlation functions called Ward-Takahashi identities (we drop the hats, it is understood that only operators appear in correlation functions, while only classical expressions appear under the path integral)

$$\begin{aligned}\langle \partial_\mu J^\mu(x)\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle =& i\delta^d(x-x_1)\langle \Delta\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle \\ &+\dots+ i\delta^d(x-x_n)\langle \mathcal{O}_1(x_1)\dots\Delta\mathcal{O}_n(x_n)\rangle\end{aligned} \tag{2.21}$$

where under the transformation $\delta\varphi = \alpha\Delta\varphi$, the $\mathcal{O}_i$ transform as $\delta\mathcal{O}_i = \alpha\Delta\mathcal{O}_i$. We observe that the insertion of $\partial_\mu J^\mu(x)$ in a correlator makes it almost vanishing, up to what are called contact terms, namely terms that arise if the insertion point x of $\partial_\mu J^\mu$ collides with the location x_i of one of the other local operators $\mathcal{O}_i$. Moreover, the coefficient of the contact term is a correlator which is non-trivial only if the local operator transforms non-trivially under the symmetry associated to J^μ, namely if $\Delta\mathcal{O}_i \neq 0$.

Let us now assume in all generality that the operator $\mathcal{O}$ is in a given representation ρ_R of the algebra $\mathfrak{g}$. This means that $\Delta\mathcal{O} = i\rho_R(X)\mathcal{O}$. (For an operator of charge q under an abelian algebra $\mathfrak{u}(1)$ we just have $\Delta\mathcal{O} = iq\mathcal{O}$.) We can then evaluate a correlator with an insertion of the operator $U_g(\Sigma_{d-1})$ together with the local insertion of $\mathcal{O}(x)$. More precisely, we can compare the correlators upon a deformation of the surface Σ_{d-1} that crosses the point x [**Exercise 2**]:

$$\langle U_g(\Sigma_{d-1})\mathcal{O}(x)\rangle = \langle R(g)\mathcal{O}(x)U_g(\Sigma'_{d-1})\rangle \ . \tag{2.22}$$

We can think either as Σ_{d-1} being on one side of x and Σ'_{d-1} on the other, or (in a Euclidean setting) as Σ_{d-1} being a closed surface containing x, while Σ'_{d-1} no longer containing it. If Σ'_{d-1} does not contain any other insertion or singular point (i.e. its interior has the topology of a ball), then it can be shrunk to a point and the operator is equivalent to the identity. Then

oftentimes the same relation will be written as

$$\langle U_g(\Sigma_{d-1})\mathcal{O}(x)\rangle = R(g)\langle\mathcal{O}(x)\rangle \ . \tag{2.23}$$

It is important to notice that the above argument proceeds from the fact that U_g (and Q before it) is defined on a $(d-1)$-dimensional surface, in other words a codimension-one surface. It is only because of this that it makes sense to say that x is one side or the other of Σ_{d-1}, or inside or outside of it.

The above is all well-established textbook material, only presented in a particular way. Our first straightforward generalization of the traditional notion of symmetry is to extend the concept of Ward-Takahashi identities (WTI) to discrete symmetries. Indeed in their local form (2.21) they only apply to continuous symmetries, where a Noether conserved current can be defined (recall that the current lives in the algebra, that only exists for continuous symmetry groups). If on the other hand one takes the global form (2.22) of the WTI, which in the continuous case is completely equivalent to the local one, we see that only group elements and their representations appear. Hence, such kind of WTI can directly be applied also to discrete groups that do not have an algebra, nor a Noether conserved current. In other words, we replace the (quantum) conservation of the current with the topological nature of the symmetry operators. Such a property can be applied to discrete and continuous symmetries alike. We will see examples of such topological defects for discrete symmetries.

Let us summarize the main take-away until now: through the Noether theorem and the definition of charges on arbitrary 'time' slices, that in a Euclidean setting can also be closed codimension-one surfaces, we have determined that symmetries acting on physical local operators are implemented by topological operators, or defects, when they pass through, or shrink on, the local operators.

3 Higher-form symmetries

We have mapped all symmetry operators, acting on local operators (namely, observables) to topological defects, defined on codimension-one surfaces Σ_{d-1} in d-dimensional spacetime $\mathcal{M}_d$. What about other topological defects in a given QFT T? A priori, T can be such that there are more general topological defects, i.e. operators that depend only topologically on the surface on which they are defined. Crucially, the dimension of such surface can be anything. Concretely, an operator V defined on a $(d-p-1)$-dimensional surface Σ_{d-p-1} is topological if its insertion in any correlator does not depend on small deformations of Σ_{d-p-1}:

$$\langle V(\Sigma_{d-p-1})\dots\rangle = \langle V(\Sigma'_{d-p-1})\dots\rangle \tag{3.1}$$

if $\Sigma'_{d-p-1} - \Sigma_{d-p-1} = \partial\mathcal{B}_{d-p}$ and there is nothing (in particular no other insertions) inside $\mathcal{B}_{d-p}$.

A simple way to construct such an operator is as follows. If in the theory there is a $(p+1)$-form $J_{(p+1)}$ such that

$$d * J_{(p+1)} = 0 \tag{3.2}$$

namely a higher-form conserved current, then we can build straightforwardly the topological operator

$$V_\alpha(\Sigma_{d-p-1}) = e^{i\alpha \int_{\Sigma_{d-p-1}} * J_{(p+1)}} \ . \tag{3.3}$$

Its topological nature is trivially determined exactly as in the usual $p = 0$ case.

The question is on what such an operator acts. Not on local operators (when $p > 0$): indeed when Σ_{d-p-1} is of codimension higher than one, there is no notion of a local insertion x being on one side or the other of the surface. In other words, the surface can be freely deformed to any other location without ever having to cross the point x.

The objects that $V(\Sigma_{d-p-1})$ cannot go around unhinged are operators W defined on a p-dimensional surface Γ_p. This is because there is a well-defined notion of *linking* between $(d-p-1)$-dimensional and p-dimensional surfaces in d-dimensional spacetime. Two such surfaces either link or not. The WTI then read

$$\langle V_g(\Sigma_{d-p-1})W(\Gamma_p)\rangle = \langle R(g)W(\Gamma_p)V(\Sigma'_{d-p-1})\rangle \tag{3.4}$$

where again W is taken to be in a representation R of the group to which the element g belongs, and Σ_{d-p-1} links Γ_p while Σ'_{d-p-1} does not. Note that the above correlators depend topologically on Σ_{d-p-1} (and Σ'_{d-p-1} respectively), but not necessarily on Γ_p, W is not necessarily a topological operator (in the same way as correlators with insertions of local operators typically do depend on the relative locations of such insertions).

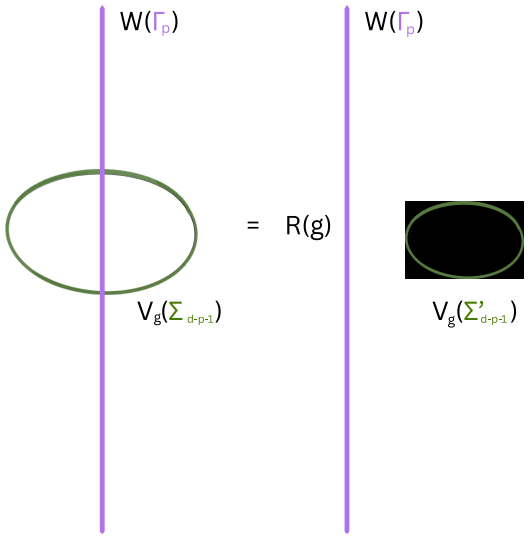

Figure 1: Here we show the action of the topological operator $V_g(\Sigma_{d-p-1})$ (which here appears as a loop for illustrative purposes) on the charged operator $W(\Gamma_p)$. On the left, V_g links with W, while on the right, V_g has now unlinked at the cost of transforming W under the representation of G.

Symmetries acting on p-dimensional objects are called *p-form symmetries*. When $p \geq 1$, these are usually called higher-form symmetries, while the usual textbook symmetries acting on local operators become known as 0-form symmetries.

We can ask now what kind of group G can be generated by such higher codimension topological defects. We then have to consider the group law. The group law determines how group elements can be composed. When group elements are associated to topological defects, one can ask how topological defects can fuse together. That the fusion should map to the group law can be seen by acting repeatedly with two (or more) symmetry defects on a given operator and recalling the basic defining property of representations that mandates $R(g_1)R(g_2) = R(g_1 g_2)$.

What is interesting is that in general the order in the group composition matters: a priori $g_1 g_2 \neq g_2 g_1$. Only *abelian* groups are such that $g_1 g_2 = g_2 g_1$ for any pair of group elements. Concerning now the fusion of topological defects, is it always the case that we can define an order? For codimension-one defects, certainly. Indeed, if we take them to be parallel (or concentric if they are closed) one can always determine the order of the defects with respect to the common transverse direction. Then

$$\begin{aligned} \langle U_{g_1 g_2}(\Sigma_{d-1})\mathcal{O}(x)\rangle &= \langle U_{g_1}(\Sigma^L_{d-1})U_{g_2}(\Sigma^R_{d-1})\mathcal{O}(x)\rangle \\ &\neq \langle U_{g_2}(\Sigma^L_{d-1})U_{g_1}(\Sigma^R_{d-1})\mathcal{O}(x)\rangle \\ &= \langle U_{g_2 g_1}(\Sigma_{d-1})\mathcal{O}(x)\rangle \ . \end{aligned} \tag{3.5}$$

On the other hand, when $p \geq 1$, there is no longer a natural ordering in the common transverse space of two parallel (or concentric) defects. One can continuously and smoothly exchange their positions without the two having

ever to touch each other. We then have for $p \geq 1$

$$\begin{aligned}\langle U_{g_1 g_2}(\Sigma_{d-p-1}) W(\Gamma_p)\rangle &= \langle U_{g_1}(\Sigma_{d-p-1}) U_{g_2}(\Sigma'_{d-p-1}) W(\Gamma_p)\rangle \\ &= \langle U_{g_2}(\Sigma_{d-p-1}) U_{g_1}(\Sigma'_{d-p-1}) W(\Gamma_p)\rangle \\ &= \langle U_{g_2 g_1}(\Sigma_{d-p-1}) W(\Gamma_p)\rangle \ . \end{aligned} \tag{3.6}$$

Then $R(g_1 g_2) = R(g_2 g_1)$ implies $g_1 g_2 = g_2 g_1$ for all pairs of group elements, i.e. the group G must be abelian. We can then state the important fact

all higher-form symmetries are abelian.

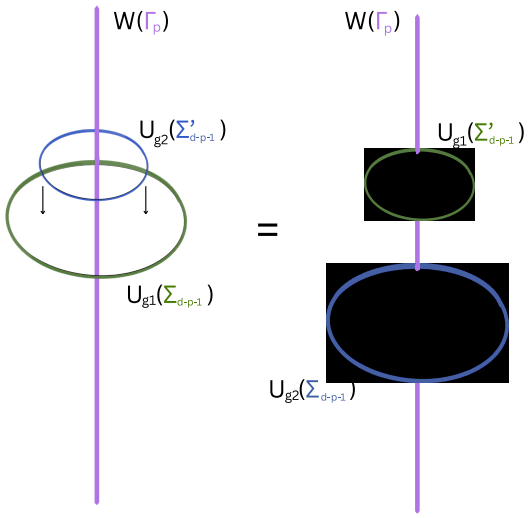

Figure 2: Higher-form symmetries with $p > 1$ have symmetry generators — here $U_{g_{1/2}}$ which are schematically loops — that are free to deform past each other. Thus, this eliminates the notion of ordering.

Until now we have tacitly assumed that symmetries form a group. But we have also declared that any topological defect in a given QFT implements a symmetry. The group nature of the symmetry is actually a constraint.

Indeed groups are defined precisely by the property that the composition of two elements of the group give one, and only one element of the group. Similarly, one asks that any element of the group has an inverse. Since composition of group elements maps to fusion of topological defects, we then have the constraint that the fusion of any two topological defects must give a single topological defect, and that for any topological defect, there must be an inverse defect with which it fuses to the identity (i.e. the trivial defect, namely nothing).

However in all generality, it is not always the case that the fusion of two topological defects yields a single topological defect. Topological defects more generally define a fusion ring, namely the fusion of two defects gives a priori a sum of other topological defects

$$V_{g_1} V_{g_2} = V_{g_1'} + V_{g_2'} + \dots \tag{3.7}$$

Such generalization of symmetry goes under the name of categorical symmetries, because the correct framework to discuss such fusions is the one of the theory of categories. It implies for instance that it is no longer true that any element has an inverse. Some elements behave as projectors, i.e. they have zero eigenvalues. For this reason, in the physics literature such generalized symmetries are dubbed *non-invertible symmetries*.

Categories also encode how symmetries of different form degree combine with each other. When the combination is non-trivial, one speaks of *higher-groups*, for symmetries that behave like groups, or more generally of higher-categorical symmetries.

It is time to see some examples to have a more concrete grasp of all these symmetry structures. We will start with the simplest higher-form symmetries.

4 Maxwell theory and its symmetries

We will study Maxwell theory in d-dimensions. This is the theory of an abelian 1-form gauge connection $A = A_\mu dx^\mu$. The action is (we now set ourselves in a Euclidean spacetime)

$$S[A] = \frac{1}{4e^2}\int d^d x F^{\mu\nu}F_{\mu\nu} = \frac{1}{2e^2}\int_{\mathcal{M}_d} F\wedge *F \tag{4.1}$$

with $F = dA$ the 2-form field strength. Maxwell theory has a gauge symmetry, namely the theory is invariant under the gauge transformation

$$A \to A + d\lambda \tag{4.2}$$

with λ a function on $\mathcal{M}_d$. As already stated, this is not a symmetry relating different field configurations, but a relation identifying them, leading to a redundancy in the description of physical configurations. Another way to see that this is not a symmetry is that its would-be Noether current trivially vanishes, since $\delta S = 0$ also for local transformations.

The (vacuum) Maxwell equations, which encompass both the equations of motion (EOM) and the Bianchi identities (BI), can be written in form notation as

$$d * F = 0 \qquad \text{(EOM)} \tag{4.3}$$

$$d\ F = 0 \qquad \text{(BI)} \tag{4.4}$$

They are then suggestive of conservation equations, as follows

$$\text{(EOM)}: \qquad d * J_e \quad \leftrightarrow \quad J_e \propto F \tag{4.5}$$

$$\text{(BI)}: \qquad d * J_m \quad \leftrightarrow \quad J_m \propto *F \tag{4.6}$$

where the subscripts e and m stand for electric and magnetic respectively, and we recall that $*^2 = \pm 1$ (the sign depends on the degree of the form it acts upon, and on the signature of spacetime).

Since we have conserved, generically higher-form, currents, these are continuous (higher-form) symmetries. We now determine the form degree of such symmetries: the current for the electric symmetry is given by the 2-form F, hence we have a 1-form electric symmetry; the current for the magnetic symmetry is given by $*F$, which in d-dimension is a $(d-2)$-form, hence we have a $(d-3)$-form magnetic symmetry. Note that in 4-dimensions, also the magnetic symmetry is a 1-form symmetry, while in 3-dimensions, it is actually a 0-form symmetry.

Now, p-form symmetries act on p-dimensional objects, which should be physical observables, hence gauge invariant in the present context of a gauge theory. What are the 1-dimensional and $(d-3)$-dimensional observables in Maxwell theory?

Let us first focus on the 1-dimensional objects. These should be gauge invariant line operators. These are known since a long time: they are called Wilson lines and are written as

$$W_n(\Gamma_1) = e^{in\int_{\Gamma_1} A} \ . \tag{4.7}$$

They can be thought as worldlines of electrically charged particles. These charged particles do not appear in the action of Maxwell theory, they are then external, or probe particles, with no dynamics of their own, very heavy in the sense that they move on a prescribed trajectory. Of course in a Euclidean setting, Wilson lines can be (and typically are) defined on closed loops (in this case they are usually called Wilson loops).

We should check that Wilson loops are indeed gauge invariant. Under $A \to A + d\lambda$ we have

$$W_n \to W_n \ e^{in\int_{\Gamma_1} d\lambda} \ . \tag{4.8}$$

It is gauge invariant if we can say that the λ-dependent phase is actually 1. If Γ_1 is a closed loop and $d\lambda$ in indeed the exterior derivative of a function

λ which is periodic along the coordinate parameterizing Γ_1, then by Stokes theorem its integral vanishes and the phase is 1 for any $n \in \mathbb{R}$.

However here we have to open a very important parenthesis. At face value in the Maxwell action only the connection A appears, and it takes value in the gauge algebra, which is $\mathfrak{u}(1)$. There is no notion of gauge group. The latter must be specified when discussing which gauge transformations are allowed. λ is a map from spacetime to the gauge group. Now consider looking at the values of λ restricted to a loop in spacetime, parameterized by θ with periodicity $\theta \sim \theta + 2\pi$. If the gauge group is non compact, namely it is $\mathbb{R}$, then in going around the loop, λ must be a strictly periodic function of θ, $\lambda(\theta + 2\pi) = \lambda(\theta)$. If on the other hand the gauge group is compact, namely it is $U(1)$, then the parameter λ is itself periodic, $\lambda \sim \lambda + 2\pi$. We can then allow for *winding*, namely

$$\lambda(\theta + 2\pi) = \lambda(\theta) + 2\pi w \ . \tag{4.9}$$

In such a situation, the loop integral evaluates to

$$\int_{\Gamma_1} d\lambda = 2\pi w \tag{4.10}$$

and the gauge transformation of the Wilson loop becomes

$$W_n \to W_n \ e^{2\pi i n w} \ . \tag{4.11}$$

A Wilson loop in Maxwell theory with $U(1)$ gauge group is gauge invariant only if the phase is 1 for any value of w. This is achieved if $n \in \mathbb{Z}$.

We thus learn that the label of the Wilson loop depends crucially on the global properties (compactness or not) of the gauge group.

Let us go back to relating the Wilson loop to electric charge. For this we can consider a path integral with the insertion of a Wilson loop

$$\int \mathcal{D}A \ e^{-S[A]} e^{in \int_{\Gamma_1} A} = \int \mathcal{D}A \ e^{-S[A] + in \int_{\mathcal{M}_d} A \wedge \delta(\Gamma_1)} \ , \tag{4.12}$$

where on the r.h.s. we have rewritten the integral of A over the loop as an integral over all of spacetime of A wedged with the Poincaré dual of Γ_1, which is a $(d-1)$-form that localizes (in the Dirac delta-function sense) to Γ_1. The EOM for A in presence of the Wilson loop become

$$\frac{i}{e^2} d * F = n\delta(\Gamma_1) \ . \tag{4.13}$$

We can now evaluate the correlator with a symmetry defect defined on a $(d-2)$-dimensional surface that links the loop Γ_1. First of all let us take as a properly normalized topological electric symmetry defect the following

$$U^e_\alpha(\Sigma_{d-2}) = e^{-\frac{\alpha}{e^2}\int_{\Sigma_{d-2}} *F} \ . \tag{4.14}$$

Then we have the relation

$$\langle U^e_\alpha(\Sigma_{d-2}) W_n(\Gamma_1)\rangle = \langle U^e_\alpha(\Sigma_{d-2})\rangle|_{W_n(\Gamma_1)} \tag{4.15}$$

namely the correlator of the two objects can be evaluated as the expectation value of one in the background of the other. The latter is finally easily evaluated given the modified EOM:

$$\begin{aligned} \langle U^e_\alpha(\Sigma_{d-2})\rangle|_{W_n(\Gamma_1)} &= \langle e^{-\frac{\alpha}{e^2}\int_{\Sigma_{d-2}} *F}\rangle|_{W_n(\Gamma_1)} \\ &= \langle e^{-\frac{\alpha}{e^2}\int_{D_{d-1}} d*F}\rangle|_{W_n(\Gamma_1)} \\ &= e^{i\alpha n \int_{D_{d-1}} \delta(\Gamma_1)} \\ &= e^{i\alpha n} \end{aligned} \tag{4.16}$$

In the second line we have defined D_{d-1} such that $\partial D_{d-1} = \Gamma_{d-2}$, and in the last line we have used the fact that $\int_{D_{d-1}} \delta(\Gamma_1) = 1$ when the surfaces Σ_{d-2} and Γ_1 have unit linking (if two surfaces link, then one must necessarily pierce the filling of the other one).

The label n of the Wilson line is then its charge under the electric 1-form symmetry. If the gauge group is $U(1)$, we recover the quantization of the

electric charge, but we also see that the electric 1-form symmetry *group* is also $U(1)$, since the action of $U^e_{\alpha+2\pi}$ is exactly the same as the one of U^e_α. If on the other hand the gauge group is $\mathbb{R}$, then n is not quantized, and α has no periodicity. We see that in this case the electric 1-form symmetry group is also $\mathbb{R}$.

Let us now turn to the magnetic $(d-3)$-form symmetry. It acts on $(d-3)$-dimensional objects. The topological symmetry defects that measure the charge of the latter are defined on a 2-dimensional surface and can be taken to be

$$U^m_\beta(\Sigma_2) = e^{i\frac{\beta}{2\pi}\int_{\Sigma_2} F} \ . \tag{4.17}$$

Σ_2 must be taken to enclose the insertion of an operator on a $(d-3)$-dimensional surface Γ_{d-3} (otherwise we can trivially shrink it to a point and the defect evaluates to 1). Taking Σ_2 to have the topology of an S^2, can the gauge connection A in the space transverse to Γ_{d-3} be taken such that $\int_{\Sigma_2} F \neq 0$? The answer takes us back to the Dirac monopole and, again, to the global structure of the gauge group. Indeed, one evaluates the integral as follows

$$\int_{S^2} F = \int_{H^2_N} dA_N + \int_{H^2_S} dA_S = \int_{S^1_{\text{eq.}}} (A_N - A_S) = \int_{S^1_{\text{eq.}}} d\lambda \tag{4.18}$$

where in the first equality we have split the evaluation of the integral over the sphere into the northern and southern hemisphere, where we can globally define a connection. However, the two connection may differ by a gauge transformation, so that the magnetic charge eventually is evaluated by integrating the gauge transformation on the equator.

Then, the global structure of the gauge group determines the possible magnetic charges. For a $U(1)$, compact gauge group, we can have winding gauge parameters so that

$$\int_{S^2} F \in 2\pi\mathbb{Z} \ . \tag{4.19}$$

We then have

$$\langle U_\beta^m(\Sigma_2)\rangle|_{H_m(\Gamma_{d-3})} = e^{i\beta m} \tag{4.20}$$

where $H_m(\Gamma_{d-3})$ is exactly defined as performing a path integral with the insertion of a $(d-3)$-dimensional surface such that $\int_{S^2} F = 2\pi m$ when S^2 links Γ_{d-3}: it is called a 't Hooft surface, or also sometimes a disorder operator. Then since magnetic charges are quantized $m \in \mathbb{Z}$, it also follows that $U^m_{\beta+2\pi} = U^m_\beta$ as far as the action on any such operator is concerned, and thus the magnetic $(d-3)$-form symmetry group is also $U(1)$.

For a non-compact $\mathbb{R}$ gauge group, we cannot have winding gauge parameters, and therefore

$$\int_{S^2} F = 0 \tag{4.21}$$

with no exception. In other words, it is not possible to insert non-trivial 't Hooft surface operators. Then all magnetic symmetry defects trivialize to the identity $U^m_\beta = 1$ and the magnetic symmetry is not a faithful symmetry (since it does not act on any operator of the theory). The magnetic symmetry is not a symmetry of non-compact Maxwell then.

What we have seen (and what we will see) in the present Maxwell example generalizes straightforwardly when replacing the 1-form A by a p-form [**Exercise 3**]. There will be a p-form electric symmetry and a $(d-p-2)$-form magnetic symmetry (when the p-form is a $U(1)$ connection). Note that this applies also to the case $p = 0$, namely a the theory of a scalar. In this case the distinction will be between a compact (periodic) and a non-compact target space. The electric 0-form symmetry is usually called a shift, or momentum symmetry, while the magnetic $(d-2)$-form symmetry is usually called winding, or vortex symmetry. We will discuss later this example.

5 Background gauge fields and anomalies

In the path integral formulation of QFT, background fields are introduced so that the partition function depends on them, and becomes a generating functional for correlation functions of the operators to which the background fields linearly couple. If the operator in question is a conserved current J^μ, then the background field must also carry a Lorentz index, and is hence a vector B_μ. The path integral with the linear coupling is then

$$Z[B_\mu] = \int \mathcal{D}\varphi \; e^{-S[\varphi]+i\int d^dx J^\mu B_\mu} \; . \tag{5.1}$$

However the current is conserved $\partial_\mu J^\mu = 0$, hence the background field has to take this into account. Since one out of the d components of J^μ is redundant because of the conservation equation, there must be also a redundancy in the d components of B_μ. This is just a (background) gauge symmetry, indeed

$$\int d^dx J^\mu B_\mu \to \int d^dx J^\mu (B_\mu + \partial_\mu \Lambda) = \int d^dx (J^\mu B_\mu - \Lambda \partial_\mu J^\mu) = \int d^dx J^\mu B_\mu \tag{5.2}$$

i.e. the coupling of the current to the background field is gauge invariant by virtue of the conservation equation. We call B_μ a *background gauge field*. Background means that we do not path integrate over it, contrarily for instance to the case of the Maxwell field in the previous section.

In form notation the coupling reads

$$\int_{\mathcal{M}_d} *J \wedge B \tag{5.3}$$

and it straightforwardly generalizes to higher-form symmetries: for a p-form symmetry, J is a $(p+1)$-form current, and thus B is a $(p+1)$-form connection, namely $(p+1)$-form with a gauge redundancy by a p-form parameter

$$B \to B + d\Lambda \; . \tag{5.4}$$

Let us then come back to Maxwell theory and write its action with the couplings to the background gauge fields for the electric and magnetic currents

$$J_e = \frac{i}{e^2} F \ , \qquad J_m = \frac{1}{2\pi} * F \tag{5.5}$$

we thus need respectively a 2-form B_e and a $(d-2)$-form B_m.

The Maxwell action with the coupling to the background gauge field is

$$\begin{aligned} S[A, B_e, B_m] &= \int_{\mathcal{M}_d} \left(\frac{1}{2e^2} F \wedge *F - i * J_e \wedge B_e - i * J_m \wedge B_m \right) \\ &= \int_{\mathcal{M}_d} \left(\frac{1}{2e^2} F \wedge *F + \frac{1}{e^2} * F \wedge B_e - \frac{i}{2\pi} F \wedge B_m \right) \\ &= \int_{\mathcal{M}_d} \left(\frac{1}{2e^2} (F + B_e) \wedge *(F + B_e) - \frac{i}{2\pi} F \wedge B_m \right) \ . \end{aligned} \tag{5.6}$$

To get to the last line, we note that the first term of the action is invariant under (electric) background gauge transformations $B_e \to B_e + d\Lambda_e$ also off-shell if the Maxwell potential is also shifted as

$$A \to A - \Lambda_e \ . \tag{5.7}$$

Then the first term of the action is invariant under this combined transformation provided one adds the term quadratic in B_e (this is sometimes called the seagull term, in reference of what one does for instance in scalar QED).

Note that a way to see that Maxwell theory has a 1-form symmetry is to notice that the Maxwell action is invariant under shifts of A by a *closed* 1-form, $A \to A - \Lambda_e$, $d\Lambda_e = 0$. Then one finds the current J_e by implementing the same symmetry though with $d\Lambda_e \neq 0$. Invariance of the action is restored by coupling to the (background) gauge field B_e.

Finally, we notice that under the magnetic background gauge transformation $B_m \to B_m + d\Lambda_m$, the action is invariant by virtue of the Bianchi identities. In particular, A does not transform in combination with this gauge transformation.

Let us now consider the path integral. It is now a functional of the background gauge fields

$$Z[B_e, B_m] = \int \mathcal{D}A \; e^{-S[A,B_e,B_m]} \; . \tag{5.8}$$

Since we have argued that background gauge transformations are related to the conservation of the currents, and hence to the symmetries of the theory, we expect the partition function to be itself gauge invariant:

$$Z[B_e + d\Lambda_e, B_m] = Z[B_e, B_m] = Z[B_e, B_m + d\Lambda_m] \; . \tag{5.9}$$

Let us check that this is indeed true. First for a magnetic gauge transformation

$$S[A, B_e, B_m + d\Lambda_m] = S[A, B_e, B_m] - \frac{i}{2\pi} \int F \wedge d\Lambda_m = S[A, B_e, B_m] \tag{5.10}$$

as we had already argued, by integration by parts and then using $dF = 0$. Then for an electric gauge transformation

$$S[A - \Lambda_e, B_e + d\Lambda_e, B_m] = S[A, B_e, B_m] + \frac{i}{2\pi} \int d\Lambda_e \wedge B_m \tag{5.11}$$

where the last arises because $F \to F - d\Lambda_e$. This term is not vanishing even after integration by parts, because for a generic background $dB_m \neq 0$.

Note that we might try to cancel this term by adding to the action a counterterm which depends uniquely on the background fields, namely

$$S_{c.t.} = -\frac{i}{2\pi} \int B_e \wedge B_m \; . \tag{5.12}$$

However this term, while curing the invariance under electric gauge transformations, will generate a non-invariance under the magnetic gauge transformations. It merely displaces the problem. Eventually, the partition function is not invariant

$$Z[B_e + d\Lambda_e, B_m] = Z[B_e, B_m] \; e^{-\frac{i}{2\pi} \int d\Lambda_e \wedge B_m} \; . \tag{5.13}$$

Such non-invariance is called an *anomaly*, more specifically a 't Hooft anomaly, or also a global anomaly.

It is very important to notice that the anomaly manifests itself as a phase that depends only on the background gauge fields and the background gauge transformations. In other words, for vanishing backgrounds, there is no problem. The symmetries are preserved in the QFT. (The situation would be different if the phase would depend on dynamical gauge fields, namely on fields one still has to path integrate over. In this case the symmetry can be indeed broken—more on this later.) The global anomaly has nevertheless an (observable) effect on the correlators of the theory. Indeed recall that the background gauge fields are tools to generate insertions of the currents in correlation functions. Then the anomalous phase signals that some correlators involving J_e and J_m will be unexpectedly non-zero [**Exercise 4**].

An important property of the anomalous phase is the following. It is defined in the spacetime manifold $\mathcal{M}_d$, but if one thinks of the latter as the boundary of $(d+1)$-dimensional manifold $\mathcal{N}_{d+1}$, then using Stokes theorem we can write

$$\frac{i}{2\pi}\int_{\mathcal{M}_d} d\Lambda_e \wedge B_m = \frac{i}{2\pi}\int_{\mathcal{N}_{d+1}} d(d\Lambda_e \wedge B_m) = \frac{i}{2\pi}\int_{\mathcal{N}_{d+1}} d\Lambda_e \wedge dB_m \quad (5.14)$$

This in turn can be written as the gauge variation of a $(d+1)$-dimensional action written purely in terms of background gauge fields (i.e. which does not depend explicitly on the gauge parameters):

$$\frac{i}{2\pi}\int_{\mathcal{N}_{d+1}} d\Lambda_e \wedge dB_m = \delta\left(\frac{i}{2\pi}\int_{\mathcal{N}_{d+1}} B_e \wedge dB_m\right) . \quad (5.15)$$

This means that the anomaly is cancelled if the QFT living on $\mathcal{M}_d$ is coupled to a (background) theory living on $\mathcal{N}_{d+1}$ (the 'bulk'). This phenomenon is called *anomaly inflow*. The coupled system

$$Z[B_e, B_m]\; e^{\frac{i}{2\pi}\int_{\mathcal{N}_{d+1}} B_e \wedge dB_m} \quad (5.16)$$

is completely invariant under gauge transformations. Note that anomaly inflow is a non-local way of compensating the non-invariance of the action.

An important property of the inflow action is that it does not depend on the metric of the bulk manifold $\mathcal{N}_{d+1}$. It has actually the form of a topological action, which we will soon study in more detail.

6 Non-invertible symmetries in Maxwell

We have already mentioned that the paradigm of generalized symmetries mandates that any topological defect in a given QFT implements a symmetry, provided it acts non-trivially on some physical object of the theory. It is not always the case that topological defects fuse according to a group law. We will see now that a simple instance of this fact occurs in a particular version of Maxwell theory.

Let us then consider the usual Maxwell theory, but where we declare that the gauge *group* is $O(2)$ instead of $U(1)$. Recall that as a group, $U(1)$ is isomorphic to $SO(2)$, which is a subgroup of $O(2)$. The latter has an extra element, which is the parity-non-preserving reflection about only one of the coordinates of the plane (when seeing $O(2)$ as the symmetry group of norm-preserving transformations of vectors on the plane $\mathbb{R}^2$). Another way to see the $O(2)$ group structure is to write

$$O(2) = U(1) \rtimes \mathbb{Z}_2 \tag{6.1}$$

where the action by conjugation of the non-trivial element C of $\mathbb{Z}_2$ on the $U(1)$ rotations is to flip the sign of the angle, $CR_\alpha C = R_{-\alpha}$.[1] In Maxwell, the rotation angle α is the gauge transformation, hence the action of C flips its sign, and hence must also flip the sign of the $U(1)$ gauge field A. It then

[1]Since $CR_\alpha = R_{-\alpha}C \neq R_\alpha C$ for generic α, $O(2)$ is a non-abelian group (though its algebra is indeed abelian!).

implements charge conjugation in $U(1)$ Maxwell. $O(2)$ Maxwell is then the usual $U(1)$ Maxwell where we also gauge, or mod out by, charge conjugation.

Note that as exotic as this sounds, it does appear in physical systems, such as for instance as the low-energy theory in a particular phase of $SO(5)$ Yang-Mills theory higgsed by a scalar matter field in the adjoint representation.

We then start with $U(1)$ Maxwell and gauge charge conjugation, which acts as

$$C \ : \ A \to -A \ . \tag{6.2}$$

How do we gauge a $\mathbb{Z}_2$ symmetry such as charge conjugation? It is actually simpler than gauging a continuous symmetry, since no propagating degrees of freedom need to be introduced. We will later discuss how to describe a $\mathbb{Z}_k$ gauge theory, but for the moment we just have to realize that this is a procedure that is routinely done in the context of the world-sheet CFT of string theory, where it is called 'orbifolding,' and it amounts to modding out by the action of a discrete group. In essence, it is very similar to the way one thinks of a quotient group G/H.

All that we will need is to take into account what remains as observables/operators in the theory after imposing that they be (gauge) invariant under charge conjugation. (Note that in principle there are also observable/operators that need to introduced, in the so-called twisted sector. This is familiar from orbifold CFTs but we will not need to discuss them in the present example.)

As far as local operators are concerned, $F_{\mu\nu}$ is a gauge invariant operator in $U(1)$ Maxwell but it is mapped to $-F_{\mu\nu}$ by C, hence it is no longer a gauge invariant operator in $O(2)$ Maxwell. Even powers of $F_{\mu\nu}$ yield gauge-invariant local operators, such as $F_{\mu\nu}F^{\mu\nu}$.

However when discussing Maxwell theory we saw the importance of extended gauge-invariant operators. Take the Wilson line, under C it trans-

forms as

$$C \ : \ W_n = e^{in\int A} \to e^{in\int(-A)} = e^{-in\int A} = W_{-n} \ . \tag{6.3}$$

The Wilson lines that are gauge invariant when the gauge group is $U(1)$ are no longer gauge invariant when the gauge group is $O(2)$ (except of course for the identity $W_0 = 1$).

We can nevertheless build a gauge invariant line operator by taking a linear combination of a line and its charge conjugate:

$$\widetilde{W}_n = W_n + W_{-n} \ . \tag{6.4}$$

Note first of all that $\widetilde{W}_n$ are defined for $n \geq 0$ (the others are redundant). Then notice that $\widetilde{W}_n$ is defined as the *sum* of two operators defined on the same 1-dimensional line (not as their product/fusion, which would trivially give the identity). Finally, we remark that indeed, if we expand $\widetilde{W}_n$ in a power series in A, only even powers will appear.

Let us now consider the topological operators, which are also gauge invariant objects. Let us consider the ones implementing the electric symmetry of $U(1)$ Maxwell, under C they behave as

$$C \ : \ U^e_\alpha = e^{-\frac{\alpha}{e^2}\int *F} \to e^{\frac{\alpha}{e^2}\int *F} = U^e_{-\alpha} \ . \tag{6.5}$$

Only for $\alpha = 0, \pi$ this is invariant (for U^e_π, one uses the periodicity of α so that $U^e_{-\pi} = U^e_{-\pi+2\pi}$). For the other values of α they are not gauge invariant, but the fix is very similar, one just needs to define

$$\widetilde{U}^e_\alpha(\Sigma_{d-2}) = U^e_\alpha(\Sigma_{d-2}) + U^e_{-\alpha}(\Sigma_{d-2}) \ . \tag{6.6}$$

We have emphasized that the two operators on the r.h.s. are evaluated on the same surface. Note that the range of α is now $[0, \pi]$, however this is not a reduced periodicity, it is rather α taking values in $S^1/\mathbb{Z}_2$ where $\mathbb{Z}_2$ is a reflection of the circle about one of its two directions.

Recall now that the U^e_α indeed fuse like a group

$$U^e_\alpha(\Sigma_{d-2})U^e_\beta(\Sigma'_{d-2}) \underset{\Sigma'\to\Sigma}{\to} e^{-\frac{\alpha}{e^2}\int_\Sigma *F} e^{-\frac{\beta}{e^2}\int_\Sigma *F} = e^{-\frac{\alpha+\beta}{e^2}\int_\Sigma *F} = U^e_{\alpha+\beta}(\Sigma_{d-2}) \ . \tag{6.7}$$

This is the same as

$$e^{i\alpha}\cdot e^{i\beta} = e^{i(\alpha+\beta)} \quad \text{with} \quad e^{i\alpha}\ ,\ e^{i\beta}\ ,\ e^{i(\alpha+\beta)}\ \in U(1)\ . \tag{6.8}$$

What about $\widetilde{U}^e_\alpha$, how do they fuse?

$$\begin{aligned}\widetilde{U}^e_\alpha\widetilde{U}^e_\beta &= (U^e_\alpha + U^e_{-\alpha})(U^e_\beta + U^e_{-\beta}) \\ &= U^e_{\alpha+\beta} + U^e_{\alpha-\beta} + U^e_{-\alpha+\beta} + U^e_{-\alpha-\beta} \\ &= \widetilde{U}^e_{\alpha+\beta} + \widetilde{U}^e_{\alpha-\beta}\ .\end{aligned} \tag{6.9}$$

This is not the composition law of a group! (We are glossing over some subtleties that arise for particular values of the angles, and that are not important for the present discussion.)

Note that of the $U(1)$ 1-form symmetry of the $U(1)$ Maxwell theory, only the $\mathbb{Z}_2$ subgroup generated by U^e_0 and U^e_π survives as a group in $O(2)$ Maxwell. The rest remains as a symmetry but of more general, categorical nature.

Indeed, $\widetilde{U}^e_\alpha$ are still topological operators of $O(2)$ Maxwell. The discrete gauging of C cannot affect their topological nature that eventually depends only on the local EOM. Hence they must be considered as implementing symmetries in $O(2)$ Maxwell. More specifically, they implement 1-form symmetries, and hence they must act on (fully gauge invariant) line operators of the same theory. The latter are precisely the operators $\widetilde{W}_n$. Let us see how

this action takes place:

$$\begin{aligned}\widetilde{U}^e_\alpha(\Sigma_{d-2})\widetilde{W}_n(\Gamma_1) &= (U^e_\alpha(\Sigma_{d-2}) + U^e_{-\alpha}(\Sigma_{d-2}))(W_n(\Gamma_1) + W_{-n}(\Gamma_1)) \\ &= e^{i\alpha n}W_n(\Gamma_1) + e^{i\alpha(-n)}W_{-n}(\Gamma_1) \\ &\qquad + e^{i(-\alpha)n}W_n(\Gamma_1) + e^{i(-\alpha)(-n)}W_{-n}(\Gamma_1) \\ &= (e^{i\alpha n} + e^{-i\alpha n})(W_n(\Gamma_1) + W_{-n}(\Gamma_1)) \\ &= 2\cos(\alpha n)\widetilde{W}_n(\Gamma_1). \end{aligned} \tag{6.10}$$

We see that unlike the typical action in an abelian group representation, the operator is not multiplied by a phase, but rather by a cosine. This means in particular that $\widetilde{U}^e_\alpha$ *annihilates* $\widetilde{W}_n$ for $\alpha = \frac{\pi}{2n}(1+2k)$, $k \in \mathbb{Z}$. Then at least some of these topological operators have some zero eigenvalues, and they then do not possess an inverse. It is then a fact that the 1-form symmetry of $O(2)$ Maxwell is non-invertible.

A very similar story applies to 't Hooft surfaces H_m and to the topological defects U^m_β of the $(d-3)$-form magnetic symmetry. It also becomes non-invertible, with very similar fusion rules and action on the charged operators.

What is currently not known is how to write background gauge fields for non-invertible symmetries. As a consequence, it is also more involved to characterize the anomaly between the electric 1-form symmetry and the magnetic $(d-3)$-form symmetry. That such anomaly is still present can be made manifest by restricting to the $\mathbb{Z}_2$ subsymmetries, which are group-like and for which the anomaly can be characterized with some more notions from algebraic topology that we will not indulge in.

7 A topological QFT and its symmetries

We are now going to discuss the symmetries of another QFT, whose interest has many motivations. Its Euclidean action is the following

$$S[B,C] = \frac{ik}{2\pi}\int_{\mathcal{M}_d} B_{(p)} \wedge dC_{(d-p-1)} \ . \tag{7.1}$$

It is usually called BF theory just because the Lagrangian reads BF if one writes $F = dC$. It is a topological theory in the following sense: its action can be written without having to specify the metric of the spacetime $\mathcal{M}_d$ (indeed in components, the integrand can be written as $\epsilon^{\mu_1\cdots\mu_d} B_{\mu_1\ldots\mu_p}\partial_{\mu_{p+1}} C_{\mu_{p+2}\ldots\mu_d}$, i.e. only in terms of the Levi-Civita tensor and the fields B and C—the metric in Maxwell theory written in the language of forms enters through the Hodge star $*$). Then it is obviously independent on the local details of the spacetime manifold it lives on. Another, strictly equivalent, way to see this is that its Hamiltonian (actually the whole stress-energy tensor) vanishes. A topological QFT is abridged as TQFT.

This action is of the same kind that appears in the inflow mechanism (5.16), though now in its own respect, and since it is now a QFT, we indeed path integrate over B and C. With respect to the anomaly inflow action, here we have a parameter k which can be considered as the (inverse) coupling. However it cannot take any value. B and C are in fact connections, namely they are defined up to gauge transformations

$$B \to B + d\lambda_B \ , \qquad C \to C + d\lambda_C \ . \tag{7.2}$$

Let us for the moment assume both B and C are actually $U(1)$ connections. This allows a periodicity in the gauge parameters that generalizes the winding that we discussed for the (scalar) gauge parameter of the Maxwell theory. The condition we impose on the gauge transformations is

$$\int_{\Sigma_p} d\lambda_B \in 2\pi\mathbb{Z} \ , \qquad \int_{\Sigma_{d-p-1}} d\lambda_C \in 2\pi\mathbb{Z} \ , \tag{7.3}$$

where Σ_p and Σ_{d-p-1} are non-trivial cycles (i.e. closed submanifolds that cannot be shrunk to a point) in $\mathcal{M}_d$. When the above integrals are non-zero one speaks of large gauge transformations. By a similar reasoning as for the Dirac monopole, the existence of large gauge transformations is in one-to-one correspondence with the possibility of having (quantized) magnetic flux:

$$\int_{\Sigma_{p+1}} dB \in 2\pi\mathbb{Z} \ , \qquad \int_{\Sigma_{d-p}} dC \in 2\pi\mathbb{Z} \ . \tag{7.4}$$

We can now check whether the action of the BF theory is indeed gauge invariant:

$$S[B + d\lambda_B, C + d\lambda_C] = S[B, C] + \frac{ik}{2\pi} \int_{\mathcal{M}_d} d\lambda_B \wedge dC \ . \tag{7.5}$$

For $\mathcal{M}_d$ of simple enough topology (e.g. S^d or $\mathbb{R}^d$) we can simply integrate by parts and set the extra term to zero. However we would like to insist that the theory be gauge invariant on *any* manifold $\mathcal{M}_d$ (this is a common requirement, motivated by the ultimate goal of considering QFTs that are consistent with quantum gravity [**Exercise 6**]). Let us then consider $\mathcal{M}_d = S^p \times S^{d-p}$, this gives

$$\int_{\mathcal{M}_d} d\lambda_B \wedge dC = \int_{S^p} d\lambda_B \int_{S^{d-p}} dC \in (2\pi)^2\mathbb{Z} \ . \tag{7.6}$$

Thus the action is not really gauge invariant, we have

$$S[B + d\lambda_B, C + d\lambda_C] = S[B, C] + 2\pi ik\mathbb{Z} \ . \tag{7.7}$$

Note however that all of the physics is determined by the path integral

$$\int \mathcal{D}B\mathcal{D}C \ e^{-S[B,C]} \ , \tag{7.8}$$

hence the above shift of the action does not affect the path integral as long as $k \in \mathbb{Z}$. We thus conclude that the coupling of this TQFT is itself quantized, namely it is another topological feature of the theory (topological in the sense

that we cannot implement small modifications of k, in particular it cannot run due to radiative corrections—even if the theory interacted with any other sector).

The equations of motion of the theory are simply

$$dB = 0 \ , \qquad dC = 0 \ . \tag{7.9}$$

They imply that on-shell, B and C are both pure gauge, hence there are no propagating physical degrees of freedom in this theory. Does this mean it is completely trivial? For now, we can assert that it describes the deep low-energy physics of a gapped phase of matter. We will see that such a gapped phase can have topologically non-trivial properties.

Similarly as in Maxwell theory, we can build Wilson surfaces for the gauge connections B and C. They are

$$W_B^n(\Sigma_p) = e^{in\int_{\Sigma_p} B} \ , \qquad W_C^m(\Sigma_{d-p-1}) = e^{im\int_{\Sigma_{d-p-1}} C} \ . \tag{7.10}$$

Because of the $U(1)$ large gauge transformations of B and C, these Wilson operators are fully gauge invariant only if $n, m \in \mathbb{Z}$. What is interesting about the Wilson surfaces in the present theory is that due to the EOM, and in contrast to Maxwell theory, they are actually topological operators. Indeed

$$W_B^n(\Sigma_p') = W_B^n(\Sigma_p) e^{in\int_{\Sigma_p'-\Sigma_p} B} = W_B^n(\Sigma_p) e^{in\int_{\mathcal{B}_{p+1}} dB} = W_B^n(\Sigma_p) \ , \tag{7.11}$$

where $\partial\mathcal{B}_{p+1} = \Sigma_p' - \Sigma_p$. This should not come as a surprise since in a TQFT, any operator should be topological. In fact, they are precisely the topological operators one would build after interpreting the equations of motion as conservation laws for two currents proportional to $*B$ and $*C$, respectively. We thus learn that in this TQFT, the Wilson surface operators actually coincide with what we would call the electric symmetry operators.

At this point we could ask if it is possible to define in this theory also the magnetic operators, from the ever-present Bianchi identities and thus from conserved currents proportional to dB and dC respectively. However the charges $\int dB$ and $\int dC$ are both trivially zero on the EOM, this means that topological operators built from such charges all reduce to the identity. In other words they generate a trivial (non-faithful) symmetry.

Let us go back to the Wilson surface. Since they are topological, they must generate symmetries in the BF theory. Such symmetries are faithful (and thus true symmetries of the theory) only if there are objects in the theory that are indeed acted upon. Then first of all let us establish the form degree of these symmetries. W_B is defined on a p-dimensional surface, hence it has codimension $(d-p)$, and it then defines a $(d-p-1)$-form symmetry, with a natural action on $(d-p-1)$-dimensional objects. On the other hand, W_C is defined on a $(d-p-1)$-dimensional surface, and by a similar argument it implements a p-form symmetry. We thus observe that W_B and W_C are perfectly suited to act on each other! They are both symmetry operators, and charged operators. Again, this is a mandatory feature for a TQFT, since any operator is topological, and hence any charged operator must also implement a symmetry of its own.

Let us compute then the correlator of two such operators, defined on surfaces that link. We will evaluate it as before, as the VEV of one operator in the background of the other

$$\langle W_B^n(\Sigma_p) W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle = \langle W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle|_{W_B^n(\Sigma_p)} \ . \tag{7.12}$$

The insertion of $W_B^n(\Sigma)$ in the path integral is

$$\int \mathcal{D}B\mathcal{D}C \ e^{-\frac{ik}{2\pi}\int B\wedge dC} e^{in\int B\wedge\delta(\Sigma)} \ , \tag{7.13}$$

where again $\delta(\Sigma_p)$ is the Poincaré dual $(d-p)$-form that localizes the integral

on Σ_p. It leads to a modification of the EOM for B, that become

$$-\frac{ik}{2\pi}dC + in\delta(\Sigma_p) = 0 \quad \leftrightarrow \quad dC = \frac{2\pi n}{k}\delta(\Sigma_p) \ . \tag{7.14}$$

We can now compute the VEV of W_C:

$$\begin{aligned}\langle W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle|_{W_B^n(\Sigma_p)} &= \langle e^{im\int_{\widetilde{\Sigma}_{d-p-1}} C}\rangle|_{W_B^n(\Sigma_p)} \\ &= \langle e^{im\int_{\widetilde{D}_{d-p}} dC}\rangle|_{W_B^n(\Sigma_p)} \\ &= e^{\frac{2\pi imn}{k}\int_{\widetilde{D}_{d-p}}\delta(\Sigma_p)} \\ &= e^{\frac{2\pi imn}{k}}\end{aligned} \tag{7.15}$$

where $\partial\widetilde{D}_{d-p} = \widetilde{\Sigma}_{d-p-1}$, and in going to the last line, we have used the fact that Σ_p and $\widetilde{\Sigma}_{d-p-1}$ are supposed to have unit linking. This is summarized as

$$\langle W_B^n(\Sigma_p)W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle = e^{\frac{2\pi imn}{k}} \ . \tag{7.16}$$

When the phase is non-trivial, it is usually said that W_B^n and W_C^m have a non-trivial *braiding*, this terminology being obviously inspired by the case $d = 3$, $p = 1$. This relation can be read in both ways, hence as already remarked, W_B^n and W_C^m are both symmetry defects and charged operators with respect to each other.

Now we should ask what is the symmetry that is being generated. It is a group since we trivially have $W_B^n W_B^m = W_B^{n+m}$, and similarly for W_C. But at face value, it would seem to be a group where both the group elements, and the representations, are labeled by integers. Now integers label representations of $U(1)$, whose elements are labeled by a continuous periodic parameter. A group with labels in the integers is $\mathbb{Z}$ but its representations have labels that take values in $U(1)$ [**Exercise 7**]. The way out of this problem is by observing that the braiding phase is trivial for some values of n and m. Indeed

$$\langle W_B^k(\Sigma_p)W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle = e^{\frac{2\pi ikm}{k}} = e^{2\pi im} = 1 \qquad \forall m \ . \tag{7.17}$$

This means that W_B^k effectively behaves as the identity, and the same can be said of W_C^k. The topological operators then have periodic labels

$$W_B^{n+k} = W_B^n \ , \qquad W_C^{m+k} = W_C^m \ . \tag{7.18}$$

In other words

$$n, m \in \mathbb{Z}_k \tag{7.19}$$

and the symmetry generated by W_B and W_C is $\mathbb{Z}_k^{(p)} \times \mathbb{Z}_k^{(d-p-1)}$, where we have indicated the form degree of the symmetries to be more precise. This is now consistent with representation theory, since the representations of $\mathbb{Z}_k$ are indeed labeled in $\mathbb{Z}_k$ itself [**Exercise 7**].

The fact that the symmetry generators are themselves charged, can be considered as a mixed anomaly. Indeed, one can find an anomaly inflow action for the symmetries of this BF theory [**Exercise 8**].

Another way to see that the holonomies of B and C are reduced to take values in $\mathbb{Z}_k$ is by performing directly the path integral on either B or C. Let us do it on C, and split its field configurations into $C = c + C_m$ where $\int dc = 0$, namely c are globally well-defined connections, while C_m are representatives of non-trivial configurations with a non-zero magnetic flux $\int C_m = 2\pi m$. The path integral then becomes

$$\begin{aligned} \int \mathcal{D}B\mathcal{D}C \ e^{-\frac{ik}{2\pi}\int B\wedge dC} &= \int \mathcal{D}B\mathcal{D}c \sum_m \ e^{-\frac{ik}{2\pi}\int B\wedge dc - \frac{ik}{2\pi}\int B\wedge dC_m} \\ &= \int \mathcal{D}B \left(\mathcal{D}c \ e^{-\frac{ik}{2\pi}\int B\wedge dc}\right) \sum_m \ e^{-\frac{ik}{2\pi}\int B\wedge dC_m} \ . \end{aligned} \tag{7.20}$$

The path integral over c just gives the local EOM $dB = 0$. The sum over the flux representatives on the other hand gives

$$\sum_{m\in\mathbb{Z}} e^{-ikm\int B} = \sum_{n\in\mathbb{Z}} \delta\left(k\int B - 2\pi n\right) \tag{7.21}$$

where we have used a version of the Poisson resummation formula, and we have been sloppy about unimportant prefactors. Note that the magnetic fluxes need in principle to specify on which cycles they occur, and consequently the integrals of B (i.e. its holonomies) are taken on the cycles that intersect the previous ones exactly once. We gloss over these details, by assuming for instance that $\mathcal{M}_d = S^p \times S^{d-p}$.

We thus see that the remaining integral over field configurations of B localizes on the set of flat connections, which moreover have holonomies constrained to be

$$\int B \in \frac{2\pi}{k}\mathbb{Z} \tag{7.22}$$

i.e. they must take values in $\mathbb{Z}_k$ (given that large gauge transformations already implement a 2π-periodicity).

The above constraint on the holonomy implies that $W_B^k = 1$, namely that the symmetry generated by W_B is $\mathbb{Z}_k$. The same can be concluded for W_C, reversing the roles of B and C in the argument.

Because of the constraints on its holonomies (7.22), we say that B is a $\mathbb{Z}_k$ gauge field, and that it is precisely the BF topological coupling, proportional to k, that makes it so. Hence a BF theory with coupling (also called *level*) k is also called a $\mathbb{Z}_k$ gauge theory. Note that this refers equivalently to a p-form or a $(d-p-1)$-form gauge theory (but not both at the same time—the other field is the one that makes the first one discrete).

We can now finally be more precise in how to proceed when one wants to gauge a discrete symmetry such as $\mathbb{Z}_k \subset U(1)$ in a theory with a partition function Z. One first introduces a (background) gauge field (generically a p-form) for the $U(1)$ symmetry, so that the partition function now depends on it, $Z[B]$. Then we make B dynamical by path integrating over it, however instead of adding a Maxwell-like kinetic term for it, we add another dynamical $(d-p-1)$-form connection C which couples to it with a BF term

proportional to k

$$\int \mathcal{D}B\mathcal{D}C \; Z[B] \; e^{-\frac{ik}{2\pi}\int B\wedge dC} \; . \tag{7.23}$$

In case one wants to gauge a discrete symmetry that is not (manifestly) embedded in a $U(1)$ group, the above trick does not directly apply, and one has to leave the realm of differential forms to enter the one of cocycles (one of the basic notions of algebraic topology). In those terms one can consider directly cocycles $\hat{B}$ in $\mathbb{Z}_k$. Then the path integral over $\hat{B}$ is just a sum

$$\sum_{\hat{B}\in\mathbb{Z}_k} Z[\hat{B}] \; . \tag{7.24}$$

We will discuss the gauging procedure in more detail in Section 9.

Let us conclude this section by observing that a TQFT such as BF theory is indeed gapped in the sense that there are no local propagating degrees of freedom, however there is non-trivial topological content, in the form of topological objects that represent non-trivial symmetries. From this point of view it is sometimes called *topological order*, in view of applications in condensed matter physics, such as topological insulators and the fractional quantum Hall effect.

8 Non-compact TQFTs and their symmetries

As we did for Maxwell theory, we can also consider a BF theory where one or both fields are *non-compact*, namely they are connections in $\mathbb{R}$ rather than in $U(1)$. This means that their gauge transformations are globally well-defined $\int d\lambda = 0$ (large gauge transformations are not allowed), and in turn there is no magnetic flux $\int F = 0$ even off-shell.

Before writing the action for such theories, let us notice that in absence of configurations with magnetic flux, the normalization of the connections is

completely arbitrary (in the compact case, one cannot change the normalization without spoiling the quantization $\int F \in 2\pi\mathbb{Z}$). Hence, as soon as one of the two connections in the BF theory is in $\mathbb{R}$, the coefficient of the BF action can be rescaled at will. We will then set this immaterial coupling to unity.

Let us start with the case where both connections are in $\mathbb{R}$. To emphasize the difference we will write them in small letters b and c. The action reads

$$S[b,c] = \frac{i}{2\pi}\int b \wedge dc \ . \tag{8.1}$$

Then most of the previous analysis goes through as before, except that the action is now strictly gauge invariant since there are no large gauge transformations. As for the Wilson surfaces, they can be written in a similar way

$$W_b^x(\Sigma_p) = e^{ix\int_{\Sigma_p} b} \ , \qquad W_c^y(\Sigma_{d-p-1}) = e^{iy\int_{\Sigma_{d-p-1}} c} \ , \tag{8.2}$$

but now both of them are gauge invariant for any $x, y \in \mathbb{R}$.

To compute the braiding, one performs exactly the same steps (one only uses local EOM), to arrive at

$$\langle W_b^x(\Sigma_p) W_c^y(\widetilde{\Sigma}_{d-p-1})\rangle = e^{2\pi i x y} \ . \tag{8.3}$$

It is still true that W_b and W_c are at the same time symmetry defects and charged operators for each other. They also fuse as a group. Their labels are in $\mathbb{R}$, and there is no periodicity, since there is no non-trivial value of x for which the phase $e^{2\pi i x y}$ is trivial for any y. Now, the representations of the (additive) group $\mathbb{R}$ take values in $\mathbb{R}$ [**Exercise 7**], hence we can finally say that the symmetry of this fully non-compact TQFT is $\mathbb{R}^{(p)} \times \mathbb{R}^{(d-p-1)}$. Such a theory can also be seen as a gauge theory of flat connections (either b or c) in $\mathbb{R}$.

Let us finally look at the mixed case, where b is non-compact while C is compact (we allow for large gauge transformations $\int d\lambda_C \in 2\pi\mathbb{Z}$). The

action reads

$$S[b,C] = \frac{i}{2\pi}\int b \wedge dC \ , \tag{8.4}$$

and it is strictly gauge invariant. The Wilson surfaces are

$$W_b^x(\Sigma_p) = e^{ix\int_{\Sigma_p} b} \ , \qquad W_C^m(\Sigma_{d-p-1}) = e^{im\int_{\Sigma_{d-p-1}} C} \ , \tag{8.5}$$

where now gauge invariance imposes $m \in \mathbb{Z}$, while still keeping $x \in \mathbb{R}$. The braiding is now given by

$$\langle W_b^x(\Sigma_p) W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle = e^{2\pi i x m} \ . \tag{8.6}$$

There is no non-trivial value of m for which the braiding trivializes for any x, however for any $x \in \mathbb{Z}$, the braiding trivializes for any m. This implies that the label x has unit periodicity $W_b^{x+1} = W_b^x$. Writing $x = \alpha/2\pi$ we see that the W_b surfaces have a label that takes values in $U(1)$, hence they generate a symmetry with this group. The total symmetry of this theory is then $\mathbb{Z}^{(p)} \times U(1)^{(d-p-1)}$. This is consistent with the representation of $U(1)$ being labeled by $\mathbb{Z}$, and vice-versa. (The fact that the representations of an abelian group form themselves an abelian group goes under the name of Pontryagin duality. The Pontryagin dual of a group G is denoted $\widehat{G}$. The following isomorphisms are then verified: $\widehat{\mathbb{R}} \simeq \mathbb{R}$, $\widehat{\mathbb{Z}_k} \simeq \mathbb{Z}_k$, $\widehat{\mathbb{Z}} \simeq U(1)$ and $\widehat{U(1)} \simeq \mathbb{Z}$ [**Exercise 7**].)

Performing the path integral over C sets the holonomies of b to be

$$\int b \in 2\pi\mathbb{Z} \ . \tag{8.7}$$

We are thus describing a discrete, though non-compact gauge theory, namely a $\mathbb{Z}$ gauge theory. If instead we perform the path integral over b, there are no constraints on the holonomies of C (besides their periodicity due to the large gauge transformations), however the EOM of b restrict C to be flat. Hence this is a theory of flat $U(1)$ connections. It is a way to gauge a $U(1)$ symmetry *without* introducing a gauge field with propagating degrees of freedom. We will see a use of this possibility later on.

9 Gauging the symmetries of a QFT

If a given QFT has a global symmetry G, then we can attempt to *gauge* it by making it local. We have already commented on this distinction: global symmetries tell us that two different configurations related by a symmetry transformation have equivalent physics, namely the observables in one configuration can be directly mapped from the observables in the other configuration; for local symmetries, two configurations related by a gauge transformation are considered to be one and only physical configuration, as a result physical observables should be completely invariant under gauge transformations.

Gauging a continuous global symmetry is very familiar from the way particle physics models are built. Typically one starts from a theory of matter fields (i.e. fermions) with some global symmetries, and then gauges some of those global symmetries by introducing dynamical gauge fields. Take for instance QED. One starts with a free Dirac fermion ψ. In presence of a mass term, such a theory has a single $U(1)$ (0-form) symmetry that rotates with the same phase the two chiralities that compose the Dirac fermion (only such rotations preserve the mass term). In order to get to QED, we gauge this symmetry by coupling the Noether current of the $U(1)$ symmetry above to a gauge field A, over which we will now path-integrate, and that will have degrees of freedom whose propagation is dictated by a Maxwell term.

What are actually the (continuous) global symmetries of QED? The 0-form symmetry of the Dirac fermion is no longer in the list of global symmetries, since we have gauged it, and as a consequence any gauge invariant observable has to be neutral under its transformations. However in the process of gauging, we have added a Maxwell field and hence we have potentially its electric and magnetic 1-form symmetries (we stick to 4 dimensions in this real-physics-inspired example). The electric 1-form symmetry is however not present (or in a sense explicitly broken) because of the presence of the Dirac

fermions. Indeed, Wilson lines can now end on the fermionic operators. For instance a Wilson line can be defined on the following open line between two points

$$\bar{\psi}(x)e^{i\int_{\Gamma_{x,x'}}A}\psi(x') \ . \tag{9.1}$$

This makes it impossible to define a (topological) conserved charge, because the surface on which the latter is defined could always be unlinked by sliding it through the endpoint of the line, without ever having to cross it. In other words, because of the possibility to create (and annihilate) charged fermions, there is no conserved 1-form symmetry charge.

On the other hand, the magnetic 1-form symmetry is still present, since we did not introduce dynamical magnetic monopoles, and hence 't Hooft lines cannot end. The magnetic charge is conserved. Hence QED has a continuous symmetry which is the 1-form symmetry generated by $J_m \propto *F$.

This story can be generalized as follows, to a generic QFT in d-dimensions that possesses a $U(1)$ p-form symmetry. Suppose that we gauge such symmetry by introducing a $(p+1)$-form connection A and making it dynamical, with a Maxwell-like kinetic term. In the gauged theory, the original $U(1)^{(p)}$ is no longer present, being gauged. The electric $U(1)^{(p+1)}$ is also not present, being broken by the objects charged under the $U(1)^{(p)}$ of the original theory. One remains then with the magnetic $U(1)^{(d-p-3)}$, that acts on the 't Hooft surfaces with magnetic flux for $F = dA$. Thus generically, gauging a $U(1)$ p-form symmetry with a dynamical connection, lands us in a new theory, which has a $U(1)$ $(d-p-3)$-form symmetry.

For instance, in $d = 3$ QFTs with a $U(1)$ 0-form symmetry, gauging of the latter yields another 3-dimensional QFT with a $U(1)$ 0-form symmetry. Hence gauging can be seen as an action on the set of 3-dimensional QFTs with a $U(1)$ 0-form symmetry.

Note also that further gauging the magnetic $U(1)^{(d-p-3)}$, one obtains a

theory with a $U(1)$ of form degree $d-(d-p-3)-3=p$, as the original one. It is however not always true that the 'doubly gauged' theory is identical to the ungauged one, because the dynamical gauge fields one introduces in the double gauging process can affect the dynamics of the whole theory [**Exercise 9**].

Let us now discuss the gauging of discrete symmetries. It is actually simpler since the gauge fields one introduces are flat. Gauging a discrete p-form symmetry, such as $\mathbb{Z}_k$, can be thought as in the orbifolding procedure (in string theory), or as in constructing a quotient (in group theory), namely modding by all the possible actions of the symmetry one is gauging. In terms of the topological defects that implement the symmetries, one is summing over all possible insertions of such defects:

$$Z_{\text{gauged}} = \sum_{\Sigma_{d-p-1}} \int \mathcal{D}\varphi e^{-S[\varphi]} U(\Sigma_{d-p-1}) \tag{9.2}$$

The sum is over $\Sigma_{d-p-1} \in H_{d-p-1}(\mathcal{M}_d, \mathbb{Z}_k)$, namely all topologically distinct closed surfaces of co-dimension $p+1$, with the understanding that wrapping the same cycle k times gives back the identity. One can trade the sum over the $(d-p-1)$-cycles with the sum over their Poincaré dual $(p+1)$-cocycles $\hat{B}$:

$$Z_{\text{gauged}} = \sum_{\hat{B} \in H^{p+1}(\mathcal{M}_d, \mathbb{Z}_k)} Z[\hat{B}] \tag{9.3}$$

where $Z[\hat{B}]$ is the partition function of the ungauged theory coupled to the background gauge fields $\hat{B}$. This coupling can be easily seen when one can write $U(\Sigma_{d-p-1}) = e^{i \int_{\Sigma_{d-p-1}} *J}$, with $\int_{\Sigma_{d-p-1}} *J \in \frac{2\pi}{k}\mathbb{Z}$. Indeed then we can rewrite

$$U(\Sigma_{d-p-1}) = e^{i \int_{\Sigma_{d-p-1}} *J} = e^{i \int_{\mathcal{M}_d} *J \wedge \hat{B}} = U(\hat{B}) \tag{9.4}$$

with $\hat{B}$ the Poincaré dual of Σ_{d-p-1} (we are being a bit sloppy in mixing the language of algebraic topology with the one of differential geometry, we hope

this will not offend/confuse the readers—many references are more precise about this). Note that $\hat{B}$ is flat, $d\hat{B} = 0$, because all Σ_{d-p-1} are closed.

As in the previous example, the p-form symmetry is no longer present in the gauged theory because one has summed over all insertions of its defects. Inserting one more does not change the sum. However, we have a new field in the theory, over which we are summing, which is the same as path-integrating as far as a field taking discrete, finite values is concerned. We can then consider Wilson surfaces of this gauge field:

$$V_n(\Sigma_{p+1}) = e^{in \int_{\Sigma_{p+1}} \hat{B}} \ . \tag{9.5}$$

$\hat{B}$ being a $\mathbb{Z}_k$ gauge field, its holonomies are constrained as $\int_{\Sigma_{p+1}} \hat{B} \in \frac{2\pi}{k}\mathbb{Z}$. Moreover the flatness of $\hat{B}$ implies that the V_n are topological defects. Hence, given their dimension, they implement a $\mathbb{Z}_n$ $(d-p-2)$-form symmetry.

We thus see that also for discrete symmetries, after gauging we obtain a new symmetry, though its form degree differs from the case of the dynamical continuous symmetries. As an example, we see that in $d = 2$, as in the world-sheet of string theory, gauging (i.e. orbifolding by) a discrete 0-form symmetry yields another 0-form symmetry, often called the quantum symmetry of the orbifold.

The same actually applies to the gauging of discrete but non-finite symmetries like $\mathbb{Z}$, and also to the gauging of $U(1)$ and $\mathbb{R}$ but with flat connections. We will see some examples later.

Finally, let us mention that for discrete symmetries, gauging the quantum symmetry yields back exactly the original theory. Indeed, let us start from the partition function of the original theory $Z[\hat{B}]$ coupled to a background $(p+1)$-cocycle field for its finite p-form symmetry. The partition function of the gauged theory is then

$$Z_{\text{gauged}}[\hat{B}'] = \sum_{\hat{B}} Z[\hat{B}] e^{i \int \hat{B} \wedge \hat{B}'} \ , \tag{9.6}$$

where now it depends on the background $(d-p-1)$-cocycle field $\hat{B}'$ for the finite $(d-p-2)$-form quantum symmetry. Repeating the procedure for the gauging of the quantum symmetry we have

$$\begin{aligned} Z_{\text{gauged}^2}[\hat{B}''] &= \sum_{\hat{B}'} Z_{\text{gauged}}[\hat{B}'] e^{i\int \hat{B}'\wedge\hat{B}''} = \sum_{\hat{B}'}\sum_{\hat{B}} Z[\hat{B}] e^{i\int \hat{B}\wedge\hat{B}'+\hat{B}'\wedge\hat{B}''} \\ &= \sum_{\hat{B}} Z[\hat{B}] \sum_{\hat{B}'} e^{i\int \hat{B}'\wedge(\hat{B}''-\hat{B})} = \sum_{\hat{B}} Z[\hat{B}]\delta(\hat{B}-\hat{B}'') = Z[\hat{B}''] \end{aligned} \tag{9.7}$$

where in one-before-the-last equality we have (schematically) used a $\mathbb{Z}_k$ version of Poisson resummation. The doubly gauged theory has the same partition function as the original one.

Until now, we have essentially assumed that the symmetry that is being gauged is not anomalous by itself. Also, we have tacitly assumed that it does not participate in a mixed anomaly. Indeed, in the latter case, gauging the symmetry usually affects drastically also the symmetry with whom it has the mixed anomaly. This effect is familiar in particle physics in the context of the Adler-Bardeen-Jackiw anomaly (of the axial symmetry), but here we will discuss it in the context of the discrete symmetries of the (compact) BF theory.

The starting point of this discussion is the braiding relation

$$\langle W_B^n(\Sigma_p) W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle = e^{\frac{2\pi i m n}{k}} \ . \tag{9.8}$$

In this relation all the topological symmetry operators appear. Gauging some symmetry is equivalent, as we have seen above, to summing over all possible insertions of the operators that generate it. This effectively renders transparent such operators/defects. Thus one can only sum over a set of defects that have trivial braiding among each other, otherwise in taking the sum, there would be incontrollable phases appearing in the various configurations one is summing over. Of the remaining defects, one should then throw out those

that are not gauge invariant, namely those that have a non-trivial braiding with the defects that have become transparent.

Looking at the braiding above, we see that generic W_B^ns have a non-trivial braiding with generic W_C^ms. On the other hand, W_B^ns have trivial braiding among themselves (often just because of their dimensionality), and the same is true for the W_C^ms.

It is then consistent to sum over all of the W_B^ns, since they have trivial braiding among themselves. However, all W_C^ms are charged under the W_B^ns, hence we have to throw them all out. Eventually, the gauged theory is empty, in the sense that there is only the identity. We call this kind of theory a trivially gapped theory. This is actually true when spacetime is compact. If we allow for boundaries, we will see that a trivially gapped theory can nevertheless be non-trivial in the latter setting, and in this case we more generally talk of a Symmetry Protected Topological phase, or SPT for short. Of course, summing all W_C^ms leads to the same conclusion. Hence, gauging a compact BF theory by either of its $\mathbb{Z}_k$ symmetries, leads to an SPT. We cannot gauge more because of the mixed anomaly between the two $\mathbb{Z}_k$ symmetries.

We can also consider intermediate situations. Suppose the level of the BF theory can be factorized as

$$k = rs \ . \tag{9.9}$$

Then we can imagine gauging only the subset of the symmetries generated by $W_B^{n'r}$. In other words we are gauging the subgroup $\mathbb{Z}_{k/r} = \mathbb{Z}_s \subset \mathbb{Z}_k^{(d-p-1)}$. Not all of the W_C^ms have non-trivial braiding with the above operators:

$$\langle W_B^{n'r}(\Sigma_p) W_C^m(\widetilde{\Sigma}_{d-p-1})\rangle = e^{\frac{2\pi i m n' r}{k}} = e^{\frac{2\pi i m n'}{s}} = 1 \quad \Leftrightarrow \quad m = m's \ . \tag{9.10}$$

This means we can keep the operators $W_C^{m's}$, which generate the subgroup $\mathbb{Z}_{k/s} = \mathbb{Z}_r \subset \mathbb{Z}_k^{(p)}$. Together with the remaining W_B^n operators, with $0 \leq n <$

r, they result in a theory with $\mathbb{Z}_r^{(p)} \times \mathbb{Z}_r^{(d-p-1)}$ symmetry, with braiding

$$\langle W_B^n(\Sigma_p) W_C^{m's}(\widetilde{\Sigma}_{d-p-1})\rangle = e^{\frac{2\pi i m' n}{r}} \ . \tag{9.11}$$

This is exactly the symmetry content of a BF theory at level r. Hence gauging a BF theory at level $k = rs$ by a $\mathbb{Z}_s$ subgroup of either symmetries, leads to a BF theory at level $k/s = r$.

Note that in the previous example, since the $W_C^{m's}$ braid trivially with all the $W_B^{n'r}$, we can actually also gauge the theory with respect to them. This is a maximal set of operators we can gauge, hence no other operators (besides the identity) survives after this. Hence, we can say that gauging the level $k = rs$ BF theory by the $\mathbb{Z}_r^{(p)} \times \mathbb{Z}_s^{(d-p-1)} \subset \mathbb{Z}_k^{(p)} \times \mathbb{Z}_k^{(d-p-1)}$ results in an SPT.

A set of defects that, upon gauging, leads to an SPT, is often called a *Lagrangian subalgebra*. For the BF theory at hand, one can classify Lagrangian subalgebras straightforwardly: there is one for each factorization of k, including the trivial ones $k = k \cdot 1 = 1 \cdot k$ (we have to consider the order of the factors).

Let us briefly consider the case of the non-compact BF theory, we will discuss only the one with symmetry $\mathbb{R}^{(p)} \times \mathbb{R}^{(d-p-1)}$ (see [**Exercise 10**]) and braiding

$$\langle W_b^x(\Sigma_p) W_c^y(\widetilde{\Sigma}_{d-p-1})\rangle = e^{2\pi i x y} \ . \tag{9.12}$$

Here again we can directly gauge by all W_b^x or all W_c^y, namely by a whole $\mathbb{R}$, leaving behind an SPT. Note that here by gauging we mean flat gauging, indeed we are summing over insertion of closed topological defects, which are associated to flat Poincaré dual connections.

More interestingly, we can gauge by a subset $W_b^{n/R}$. This set generates a $\mathbb{Z} \subset \mathbb{R}$, hence the remaining W_b^x operators generate a $\mathbb{R}/\mathbb{Z} \simeq U(1)$ (recall that flat gauging is really the same as taking a quotient of a group). The

surviving W_c^y operators are determined by

$$\langle W_b^{n/R}(\Sigma_p) W_c^y(\widetilde{\Sigma}_{d-p-1})\rangle = e^{\frac{2\pi i n y}{R}} = 1 \quad \Leftrightarrow \quad y = mR \ . \tag{9.13}$$

They generate a $\mathbb{Z}$ symmetry. We are then left with a BF theory with symmetry $\mathbb{Z}^{(p)} \times U(1)^{(d-p-1)}$, this is the symmetry content of the 'mixed' compact/non-compact BF theory. Finally, gauging $\mathbb{Z}^{(p)} \times \mathbb{Z}^{(d-p-1)} \subset \mathbb{R}^{(p)} \times \mathbb{R}^{(d-p-1)}$ leads to an SPT.

All the latter examples have given us abundant evidence that in the presence of anomalies, the symmetry content of a theory can indeed be reduced upon gauging. It can also increase, as we can imagine by 'undoing' the gaugings performed above. In order to get intuition about this, we need to understand in which way an SPT can be different from a really empty, trivially gapped theory.

10 BF theory with a twist and SPTs

The BF theories that we have considered until now are rather simple TQFTs, but by no means they exhaust all the possibilities for such theories, neither they display the most general symmetry structure. In this section we will see how to make more complicated theories out of the BF ones, just by adding a little 'twist'. In fact, twist is the technical term, though it is used a little loosely, to denote theories that sometimes go under the general name of Dijkgraaf-Witten theories. We are not going to discuss those in all generality, we are going to focus on one example in 4 dimensions, which could also be referred to as the Kapustin-Seiberg theory. One starts from the BF theory of a 2-form B and a 1-form C

$$S_{BF} = \frac{ik}{2\pi} \int_{\mathcal{M}_4} B \wedge dC \ . \tag{10.1}$$

In 4 dimensions a topological action is obtained integrating a 4-form (that does not involve the Hodge star $*$). What other 4-forms can we write out of B and C, besides $B \wedge dC$ (and $C \wedge dB$ that is trivially obtained by integration by parts)? $C \wedge C = 0$ by antisymmetry of the wedge product on 1-forms, $dC \wedge dC$ is a total derivative, so we are left only with $B \wedge B$, which is exactly a 4-form. Hence the most general action is

$$S_{KS} = \frac{ik}{2\pi} \int_{\mathcal{M}_4} \left(B \wedge dC + \frac{p}{2} B \wedge B \right) \ . \tag{10.2}$$

Considering only small gauge transformations for the moment, we see that S_{KS} is gauge invariant only if we modify the gauge transformations as follows

$$B \to B + d\lambda_B \ , \qquad C \to C + d\lambda_C - p\lambda_B \ , \tag{10.3}$$

the 1-form gauge parameter of B also enters, with no differential, in the gauge transformation of C. This part of δC precisely compensates the gauge variation of the $B \wedge B$ term.

Considering now also large gauge transformations, we obtain

$$S_{KS} \to S_{KS} + \frac{ik}{2\pi} \int_{\mathcal{M}_4} \left(d\lambda_B \wedge dC - \frac{p}{2} d\lambda_B \wedge d\lambda_B \right) \ . \tag{10.4}$$

For the path integral to be invariant under gauge transformations, we need the extra term to be in $2\pi i\mathbb{Z}$ for any non-trivial fluxes $\int d\lambda_B \in 2\pi\mathbb{Z}$, $\int dC \in 2\pi\mathbb{Z}$. In fact, $\int d\lambda_B \wedge d\lambda_B$ is the second Chern class of the bundle associated to λ_B, and it takes values in $4\pi^2\mathbb{Z}$ in general, and in $8\pi^2\mathbb{Z}$ on manifolds $\mathcal{M}_4$ that support a spin structure (also called spin manifolds). Thus, asking the path integral to be invariant in the most general transformations of the most general configuration, we must have $k \in \mathbb{Z}$ and $kp \in \mathbb{Z}$ for spin manifolds (or $kp \in 2\mathbb{Z}$ more generally). In fact, for consistent flux quantization of dC after a gauge transformation, p must be an integer. Hence this theory is specified by two integers k and p, with no further constraints if we restrict to spin manifolds (which we will do in the following to simplify the presentation).

The gauge invariant extended operators are determined by the gauge transformations (10.3). The Wilson surfaces of B are given as usual by

$$W_B^n(\Sigma) = e^{in\int_\Sigma B} \ , \tag{10.5}$$

where Σ is a closed 2-surface and $n \in \mathbb{Z}$ because of the large gauge transformations. In the pure BF theory we would then have Wilson lines for C, however in the twisted BF theory these are no longer gauge invariant, because of the shift involving λ_B:

$$\tilde{W}_C^m(\Gamma) = e^{im\int_\Gamma C} \to \tilde{W}_C^m(\Gamma) e^{-imp\int_\Gamma \lambda_B} \tag{10.6}$$

where Γ is a closed 1-line and we have taken $m \in \mathbb{Z}$ to ensure invariance under large gauge transformations of the parameter λ_C. The additional phase can be canceled by an operator $W_B^{mp}(\Sigma)$ if the latter is defined on an *open* surface Σ whose boundary is Γ, $\partial\Sigma = \Gamma$:

$$W_B^{mp}(\Sigma) \to W_B^{mp}(\Sigma) e^{imp\int_\Sigma d\lambda_B} = W_B^{mp}(\Sigma) e^{imp\int_\Gamma \lambda_B} \ . \tag{10.7}$$

Eventually, the gauge invariant operator is defined on a closed 1-line $\partial\Sigma$ *and* the open surface Σ of which it is a boundary, as follows

$$W_C^m(\partial\Sigma, \Sigma) = \tilde{W}_C^m(\partial\Sigma) W_B^{mp}(\Sigma) = e^{im\int_{\partial\Sigma} C + imp\int_\Sigma B} \ . \tag{10.8}$$

Both the operators $W_B^n(\Sigma)$ and $W_C^m(\partial\Sigma, \Sigma)$ are topological by virtue of the EOM derived from (10.2):

$$dB = 0 \ , \qquad dC + pB = 0 \ . \tag{10.9}$$

This is perhaps more subtle for $W_C^m(\partial\Sigma, \Sigma)$. For fixed $\partial\Sigma$, the topological dependence on the extension Σ is determined by the topological nature of $W_B^n(\Sigma)$. The topological dependence on $\partial\Sigma$ can be argued as follows. Consider a deformation of $\partial\Sigma$ to $\partial\Sigma'$ such that Σ' is entirely contained in Σ,

namely $\Sigma - \Sigma' = D$ is a disk. We can always take it this way because of the independence on the extension. Then we also have $\partial\Sigma - \partial\Sigma' = \partial D$, so that

$$\begin{aligned} W_C^m(\partial\Sigma, \Sigma) &= W_C^m(\partial\Sigma', \Sigma') e^{im \int_{\partial D} C + imp \int_D B} \\ &= W_C^m(\partial\Sigma', \Sigma') e^{im \int_D (dC + pB)} \\ &= W_C^m(\partial\Sigma', \Sigma') \ . \end{aligned} \tag{10.10}$$

We see that this theory has line operators which are not *genuine*, in the sense that they require to be attached to an open surface in order to be gauge invariant.

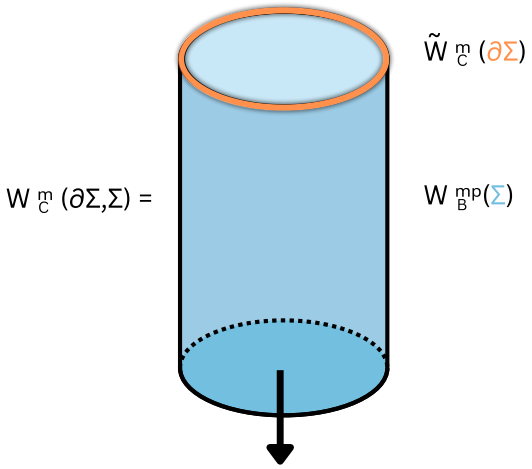

Figure 3: Due to the twist term in the action, only the total, non-genuine operator $W_C^m(\partial\Sigma, \Sigma)$ is gauge invariant and well-defined.

As in the pure BF theory, the path integral over the non-trivial fluxes of B and C yields constraints on the respective holonomies, which eventually imply

$$W_B^k(\Sigma) = 1 \ , \qquad W_C^k(\partial\Sigma, \Sigma) = 1 \ . \tag{10.11}$$

As an interesting consequence, this means that some W_C^m operators can be genuine after all. Indeed, if $mp \in k\mathbb{Z}$, the surface part of the operator

trivializes and the latter is defined purely on a line. This happens when

$$g = \gcd(k, p) > 1 \tag{10.12}$$

then we see that

$$W_C^{k/g}(\partial\Sigma, \Sigma) = \tilde{W}_C^{k/g}(\partial\Sigma) W_B^{k(p/g)}(\Sigma) = \tilde{W}_C^{k/g}(\partial\Sigma) \tag{10.13}$$

since p/g is an integer.

The fact that some line operators are non-genuine alters the content of the theory with respect to the pure BF theory. Indeed, any non-genuine line operator can be eventually shrunk to a point along its extension Σ. Conversely, any closed surface operator, whose open counterpart appears in a non-genuine operator, can be opened by the creation of a line boundary, and eventually also shrunk to a point. Hence we have to take out of the spectrum of extended operator both the non-genuine lines, and the surface operators that appear in the definition of the non-genuine line operators.

For generic g, this means we keep only the g line operators $W_C^{m(k/g)}(\partial\Sigma)$, and the first g surface operators $W_B^n(\Sigma)$ (since p has a factor of g, neither of the operators with $0 \leq n < g$ will ever appear in a surface attached to a non-genuine line). In the extreme case where $g = 1$, for instance for $p = 1$, then only the identity is left, and the twisted BF theory is an SPT. It is fairly obvious that though we have just seen that such a theory is trivial in a sense, it is not in many others. We will see that the non-triviality of such an SPT manifests itself when the spacetime has boundaries.

Let us see how the objects in the present TQFT braid with each other. In fact, we can define the braiding (or more generally the correlation functions) of the operators including the non-genuine ones. We find [**Exercise 11**]

$$\langle W_B^n(\Sigma) W_C^m(\partial\Sigma', \Sigma') \rangle = e^{\frac{2\pi i n m}{k}} \tag{10.14}$$

where Σ links $\partial\Sigma'$ once. Since in the definition of W_C^m there is both a line and a surface, the correlator between two such objects can be non-trivial

(depending on the configuration)

$$\langle W_C^m(\partial\Sigma,\Sigma)W_C^{m'}(\partial\Sigma',\Sigma')\rangle = e^{\frac{2\pi i m m' p}{k}} \tag{10.15}$$

with Σ and Σ' intersecting at one point.

When $g > 1$, we can see that the physical operators have the braiding relations of a BF theory at level g: first of all for $m, m' \in (k/g)\mathbb{Z}$ the phase of the correlator between two line operators is correctly trivial; then we have that the phase from braiding W_B^n with $W_C^{m(k/g)}$ is precisely $e^{\frac{2\pi i n m}{g}}$. Hence the symmetry content of a 4d BF theory at level k with twist p is given by $\mathbb{Z}^{(1)}_{\gcd(k,p)} \times \mathbb{Z}^{(2)}_{\gcd(k,p)}$, with the usual mixed anomaly.

One can twist also BF theories that involve non-compact fields. Keeping our set-up of a 1-form and a 2-form in 4 dimensions, we have three possibilities: both B and C in $\mathbb{R}$; B in $\mathbb{R}$ and C in $U(1)$; and B in $U(1)$ and C in $\mathbb{R}$. In all such cases one can operate some rescalings to set k and/or p to a standard value. The discussion is rather straightforward but the final symmetry content can be surprising [**Exercise 12**].

In spacetime dimension different than 4, twists of BF theories can also be defined but they will be slightly different, due to the need to be able to write a non-vanishing twist term. The 2-dimensional case is very similar to the one discussed above. In 3 dimensions the situation is different because the twist term is of Chern-Simons type, leading to qualitatively different consequences. We will not discuss these interesting topics here.

11 TQFTs with boundaries and the SymTFT

In the previous sections, we have always implicitly assumed that spacetime was not only Euclidean, but also compact and closed, i.e. with no boundaries. We are now going to discuss how to cope with spacetimes with boundaries. This is usually addressed by imposing boundary conditions on the various

fields of the given theory, in such a way that the variational principle is well-defined. In the case of gauge theories, one can also impose that the boundary conditions do not break gauge invariance. We are going to discuss first the simple compact BF theory, and we will see that boundary conditions have an effect very similar to gauging.

The relation is the following. A boundary can be seen more generally as a codimension-one interface between a given QFT and the empty QFT. In the case of our simple BF TQFT, we know that the empty QFT is obtained by gauging a Lagrangian subalgebra of its topological defects. Thus a BF theory on a spacetime with a boundary can be seen as the same BF theory on a larger spacetime with an interface, where on the other side of the interface we gauge a Lagrangian subalgebra. The boundary, or interface, acts then as a locus where the defects that we gauge become transparent. This means in particular that such defects can *end* on the boundary (since the part that lies on the boundary can be considered as the identity).

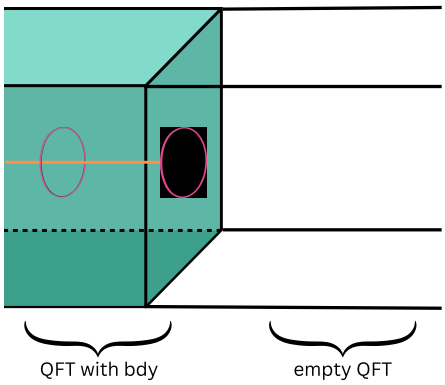

Figure 4: On the left in green, there is a QFT with a boundary. The yellow line represents a representative of the Lagrangian algebra, which is gauged. The pink loops are the operators which are projected out by the boundary, and can exist either in the QFT or on the boundary itself.

The other defects, which are not gauge-invariant by the very definition of

Lagrangian subalgebra, are projected out beyond the interface (i.e. outside of the boundary), but on the boundary they can still survive since they can link the endpoints of the defects that we are gauging. Let us illustrate this with our BF theory example.

We rewrite the action

$$S[B,C] = \frac{ik}{2\pi}\int_{\mathcal{M}_{d+1}} B_{(p+1)} \wedge dC_{(d-p-1)} \ , \tag{11.1}$$

where B and C are two $U(1)$ connections, and for later convenience we have now formulated our BF theory on a spacetime of dimension $(d+1)$, so that its boundary is a d-dimensional manifold. The form degrees are chosen so that in the bulk (or in a spacetime without boundaries) the symmetry content of the theory is $\mathbb{Z}_k^{(p+1)} \times \mathbb{Z}_k^{(d-p-1)}$.

Let us start by choosing the W_B^n $(p+1)$-dimensional surfaces as a Lagrangian subalgebra. They can then end on the boundary, and they will define there p-dimensional loci. The W_C^m $(d-p-1)$-dimensional surfaces can be pushed to lie on the boundary (we do not gauge over them so that they do not become the identity), where they can still link the end-loci of the W_B surfaces. Then, looking only at the boundary, we have a situation similar to a $\mathbb{Z}_k$ p-form symmetry at work. We will make this link more precise in the following.

Note that choosing instead the W_C^m as Lagrangian subalgebra, we invert the roles, so that we have on the boundary the $(p+1)$-dimensional W_B defects that can link the $(d-p-2)$-dimensional boundaries of the W_C defects. This is reminiscent of a $\mathbb{Z}_k$ $(d-p-2)$-form symmetry. We notice that $(d-p-2)$ is precisely the form-degree of the quantum symmetry one expects after gauging a discrete p-form symmetry (and vice-versa). So the different choices of Lagrangian subalgebras in the bulk TQFT seem to be related to different gaugings of discrete symmetries of a would-be QFT on the boundary.

The more general Lagrangian subalgebras that exist when k can be fac-

torized as $k = rs$ lead to a mix of W_B and W_C defects being able to end or not on the boundary, with the ones staying intact eventually generating a $\mathbb{Z}_r^{(p)} \times \mathbb{Z}_s^{(d-p-2)}$ symmetry, i.e. exactly what one would get by gauging a $\mathbb{Z}_s$ subgroup of the $\mathbb{Z}_k^{(p)}$ symmetry of a QFT.

Relating boundary conditions of a TQFT to its Lagrangian subalgebras is a very efficient way to obtain a complete classification. Given that the TQFT we are discussing has a nice explicit action, we would like to translate the boundary conditions that we have just discussed in more mundane terms such as Dirichlet or Neumann boundary conditions, and possibly impose them at the level of the action by coupling the theory to (boundary) edge modes. The latter also ensure invariance under the gauge transformations of B and C in presence of a boundary.

We immediately see that the choice of the Lagrangian subalgebra generated by the W_B is equivalent to choosing Dirichlet boundary conditions for B, namely $B = 0$ at the boundary, while keeping free (Neumann) boundary conditions for C. At the opposite end, Dirichlet for C amounts to choosing the W_C as Lagrangian subalgebra. In this language it is however slightly more involved to discuss boundary conditions that match the 'mixed' Lagrangian subalgebras.

Let us then go back to the action and verify its gauge variation in presence of a boundary $\mathcal{M}_d = \partial \mathcal{M}_{d+1}$

$$\delta S = \frac{ik}{2\pi} \int_{\mathcal{M}_{d+1}} d\lambda_B \wedge dC = \frac{ik}{2\pi} \int_{\mathcal{M}_d} \lambda_B \wedge dC \ , \tag{11.2}$$

where we have used Stokes' theorem to reduce the variation to the boundary. We see then that when the spacetime in which the BF theory lives has a boundary, then the action is not gauge-invariant if we do not impose some conditions there. Instead of freezing the gauge transformations, what we can do is to introduce edge modes, i.e. dynamical modes that are restricted to live on the boundary, and couple them to the bulk fields in such a way that

the coupled system is gauge invariant.

Let us introduce a $U(1)$ p-form connection ϕ that lives only on $\mathcal{M}_d$, with an action that boils down to a (topological) coupling of ϕ to the bulk fields, actually one of them

$$S_\partial = -\frac{ir}{2\pi}\int_{\mathcal{M}_d} \phi \wedge dC \ . \tag{11.3}$$

r must be an integer as usual. We also assume that under the gauge transformations of the bulk fields, ϕ transforms as

$$\phi \to \phi + s\lambda_B \ , \tag{11.4}$$

where it is understood that the variation above is non-trivial only if all the indices of λ_B lie along the boundary.

The variation of the boundary term then gives

$$\delta S_\partial = -\frac{irs}{2\pi}\int_{\mathcal{M}_d} \lambda_B \wedge dC \ , \tag{11.5}$$

and it cancels the contribution from the bulk variation if $k = rs$. We see that we have a set of distinct edge mode actions that exactly matches the Lagrangian subalgebras of the BF theory. We are going to see that they indeed reproduce the same physics.

In order to do so, we have to pay attention to the EOM at the boundary, and the effect of performing the path integral also over ϕ. The EOM of ϕ will just impose $dC = 0$ on the boundary as in the bulk, but the sum over its fluxes is restricting the holonomy of C on the boundary

$$\int_{\Sigma\subset\mathcal{M}_d} C \in \frac{2\pi}{r}\mathbb{Z} \ . \tag{11.6}$$

This means that some surface operators W_C become trivial at the boundary

$$W_C^r(\Sigma \subset \mathcal{M}_d) = e^{ir\int_{\Sigma\subset\mathcal{M}_d} C} = 1 \ . \tag{11.7}$$

In particular, they become all trivial for $r = 1$ (hence this corresponds to Dirichlet for C), while none besides W_C^k, which is already the identity, become trivial for $r = k$ (hence this is Neumann for C).

There is another contribution to the boundary EOM, which comes from the variation of C. Indeed when varying the bulk action with respect to C, we produce a boundary term by an integration by parts. This combines with the variation of the edge mode action, giving the boundary EOM

$$kB = rd\phi \ . \tag{11.8}$$

This in turn has an implication for the holonomies of B on the boundary

$$\int_{\Sigma \subset \mathcal{M}_d} B = \frac{r}{k} \int_{\Sigma \subset \mathcal{M}_d} d\phi \in \frac{2\pi}{s}\mathbb{Z} \ , \tag{11.9}$$

so that

$$W_B^s(\Sigma \subset \mathcal{M}_d) = 1 \ . \tag{11.10}$$

For $s = 1$ ($r = k$) we have that all W_B surfaces are trivialized on the boundary, hence we have Dirichlet for B, while for $s = k$ ($r = 1$) they all survive, as for Neumann boundary conditions for B. The benefit of the edge mode action is that the boundary conditions are implemented in a gauge invariant way, and all the intermediate cases appear on the same footing.

We should now note that the boundary that we have introduced is more specifically a topological boundary, in the sense that local deformations of the boundary do not have an effect on the coupled bulk-boundary system. This can be seen in several ways. An interface between a TQFT and the same TQFT in which we have gauged a discrete symmetry is necessarily topological because of the topological nature of the discrete gauging. This can also be seen from the perspective of the edge mode action, in which the edge mode ϕ does not have an action (of its own and in its coupling to

the bulk fields) involving the Hodge $*$. Finally, the same can be said about Dirichlet or Neumann boundary conditions.

This is in general not the case: the boundary $\mathcal{M}_d$ could very well host a full-blown QFT with local dynamics and local or extended non-topological operators. Or at the very least it could implement boundary conditions that involve the (boundary) Hodge $*$, thus leading to a dependence of the bulk-boundary system on small deformations of the boundary. Note that this general QFT living on the boundary will still have in its spectrum some topological operators, namely those that implement its symmetries.

We are thus naturally led to the Symmetry TQFT paradigm, or SymTFT for short, which consists of the following construction. We have a TQFT in a $(d+1)$-dimensional bulk which is a 'slab' between two boundaries. On one boundary, the *physical* one, we couple a generic QFT, with its topological and non-topological operators. Generically, the bulk-boundary coupling should imply that topological operators of the boundary QFT can be freely moved to the bulk, while non-topological operators of the boundary QFT can provide anchors on which bulk topological operators can end. On the other boundary, the *topological* one, we impose topological boundary conditions as described above.

Since the topological boundary can be deformed freely, in particular (if the topology of the slab is trivial, i.e. contractible in the codimension of the boundaries) we can push the topological boundary towards the physical one, performing a slab compactification of sorts. In the SymTFT language, sometimes the QFT that one couples to the bulk TQFT is called a *relative* QFT, while the one that is obtained after slab compactification is called an *absolute* QFT. The reason is that in the latter, the information about the choice of the topological boundary conditions is an integral part to its definition.

The important, and useful feature of the SymTFT framework is that it

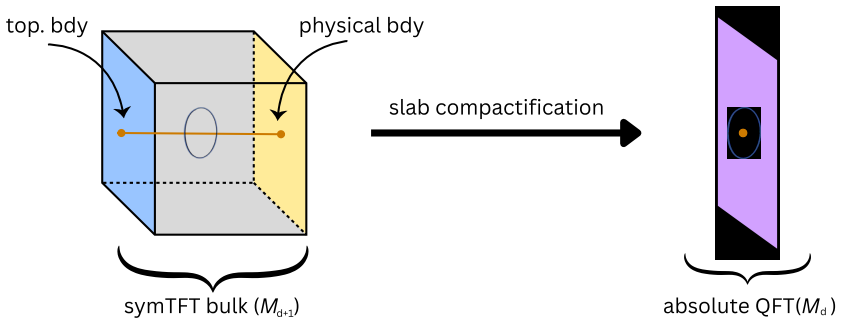

Figure 5: On the left, the symTFT setup, with the topological boundary condition in blue, the physical boundary condition (or relative QFT) in yellow, and the topological bulk which spans between them in gray. After shrinking the interval to slab compactify, one combines the topological and physical boundaries to produce the absolute QFT in d dimensions. Charged operators are those that can end on the topological boundary, shown in orange, while the other (blue) operators remain topological and become the symmetry generators in the absolute QFT.

allows to define families of QFTs which are related by the fact that they are defined by the same SymTFT, but with different topological boundary conditions. As we have already argued, a different choice of boundary conditions yields two QFTs which are related by a (discrete) gauging of some of their symmetries. Indeed, only the topological operators of the bulk TQFT that are not trivialized by the topological boundary conditions survive in the absolute QFT, while the only genuine operators of the latter, are those that are attached precisely to bulk topological operators that can end on the topological boundary. We thus see that the picture of topological defects acting on the end-loci was indeed representing the action of a symmetry in a QFT.

We now briefly go over the non-compact case

$$S = \frac{i}{2\pi} \int_{\mathcal{M}_{d+1}} b_{(p+1)} \wedge dc_{(d-p-1)} \ . \tag{11.11}$$

This is actually an important extension of the formalism, because as we have seen (at least for the BF theory), the compact case can only describe discrete finite symmetries. Now since the aim of the SymTFT paradigm is to encode symmetries of QFTs at large, and since in many of the latter we have continuous symmetries, we are in need of bulk TQFTs that have themselves continuous symmetries. As we have seen in the previous sections, this is a feature of only the non-compact TQFTs (at least for those of the BF kind).

The topological boundary conditions are simply determined. Here we use the edge mode action (we already discussed the Lagrangian subalgebras previously). The gauge variation

$$\delta S = \frac{i}{2\pi} \int_{\mathcal{M}_d} \lambda_b \wedge dc \ , \tag{11.12}$$

is compensated by the boundary action

$$S_\partial = -\frac{iR}{2\pi} \int_{\mathcal{M}_d} \phi \wedge dc \ , \tag{11.13}$$

with R a real number and ϕ again a $U(1)$ p-form connection (we stress that it is $U(1)$-valued even though b and c are $\mathbb{R}$-valued) that transforms as

$$\phi \to \phi + \frac{1}{R} \lambda_b \ . \tag{11.14}$$

The gauge variations of the bulk and of the boundary then cancel.

The sum over fluxes of ϕ and the boundary EOM of c imply the following respective conditions on the holonomies

$$\int_{\Sigma \subset \mathcal{M}_d} c \in \frac{2\pi}{R} \mathbb{Z} \ , \qquad \int_{\Sigma \subset \mathcal{M}_d} b \in 2\pi R \ \mathbb{Z} \ . \tag{11.15}$$

These in turn imply that the following operators are trivialized at the boundary

$$W_b^{n/R} = e^{i\frac{n}{R}\int b} = 1 \ , \qquad W_c^{mR} = e^{imR\int c} = 1 \ , \qquad n, m \in \mathbb{Z} \ . \tag{11.16}$$

These are precisely the Lagrangian subalgebras that we had derived previously, parametrized by the real number R.

12 Scalar in 2d: SymTFT and duality defect

One way to *derive* the theory of a (compact) scalar in 2 dimensions is through the SymTFT. Let us take in the 3-dimensional bulk the non-compact TQFT discussed at the end of the previous section, where both b and c are 1-form connections. The whole set-up of the SymTFT is then given by

$$S^{\text{tot}} = S_\partial^{\text{top}} + S^{\text{bulk}} + S_{\partial'}^{\text{phys}} \tag{12.1}$$

where

$$S^{\text{bulk}} = \frac{i}{2\pi}\int_{\mathcal{M}_3} b \wedge dc \ , \qquad S_\partial^{\text{top}} = -\frac{iR}{2\pi}\int_{\mathcal{M}_2} \phi \ dc \tag{12.2}$$

and we need to pick a $S_{\partial'}^{\text{phys}}$. As already mentioned, it is possible to introduce boundary conditions at the physical boundary $\mathcal{M}_2'$ without introducing physical degrees of freedom there, but just some boundary conditions on the bulk fields, that are no longer required to be topological. Since in the bulk we have two 1-forms, on the boundary they can be (boundary) Hodge dual to each other. We can then add a boundary term that implements the condition

$$c|_{\mathcal{M}_2'} = i *_2 b|_{\mathcal{M}_2'} \ . \tag{12.3}$$

The boundary action that implements this condition is

$$S_{\partial'}^{\text{phys}} = \frac{1}{4\pi}\int_{\mathcal{M}_2'} c \wedge *_2 c \ . \tag{12.4}$$

One can check that the condition (12.3) arises as the the coefficient of δc in the boundary term at the physical boundary of the variation of $S^{\text{bulk}} + S^{\text{phys}}_{\partial'}$. These are sometimes called conformal boundary conditions because they do not introduce any scale.

We can now perform slab compactification to obtain the (absolute) QFT that this set-up is describing. We then push $\mathcal{M}_2$ on $\mathcal{M}_2'$ and we integrate out the bulk modes. Since one of the bulk EOM is $dc = 0$, the on-shell bulk action just gives a vanishing contribution, so that the total action reduces to the sum of the two boundary actions (note that the two boundaries of the slab have to have the opposite orientation, so that $\mathcal{M}_2 = -\mathcal{M}_2'$)

$$S^{\text{tot}} = S^{\text{top}}_{\partial} + S^{\text{phys}}_{\partial'} = \frac{iR}{2\pi} \int_{\mathcal{M}_2'} \phi \; dc + \frac{1}{4\pi} \int_{\mathcal{M}_2'} c \wedge *c \tag{12.5}$$

where however the boundary EOM are imposed

$$b = R d\phi \; , \qquad c = i * b \; , \tag{12.6}$$

and from now on the $*$ is always the 2-dimensional one. We integrate by parts the first term and substitute c for b to get

$$S^{\text{tot}} = \frac{R^2}{2\pi} \int_{\mathcal{M}_2'} d\phi \wedge *b + \frac{1}{4\pi} \int_{\mathcal{M}_2'} *b \wedge b \tag{12.7}$$

where we used $*^2 = -1$ on 1-forms in 2 Euclidean dimensions. We finally eliminate b in favor of ϕ to get

$$S^{\text{tot}} = \frac{R^2}{4\pi} \int_{\mathcal{M}_2'} d\phi \wedge *d\phi \; . \tag{12.8}$$

This a very interesting action, though it represents a free QFT. It describes a compact scalar, and the overall (dimensionless) coupling R^2 can be thought of as related to the *radius* of the circle parametrized by ϕ.

Indeed, note first of all that a $U(1)$-valued 0-form is nothing less and nothing more than a scalar with periodicity

$$\phi \sim \phi + 2\pi \; . \tag{12.9}$$

It is then manifest that ϕ parametrizes a circle. The periodicity is fixed by the normalization of the fluxes $\int d\phi \in 2\pi\mathbb{Z}$, that for a compact scalar are called *windings*. That R^2 is related to the radius of the target S^1 is easily seen by writing $\hat{\phi} = R\phi$, so that $\hat{\phi}$ is canonically normalized and its periodicity is now $\hat{\phi} \sim \hat{\phi} + 2\pi R$.

The action (12.8) actually describes a conformal field theory (CFT) since as already mentioned the 'coupling' is dimensionless, moreover the theory is free, hence the classical scale invariance cannot be spoiled at the quantum level. We are describing a family of CFTs parametrized by R^2, this is called a *conformal manifold*. Note finally that this is nothing else than the generalization to 0-forms in 2 dimensions of the Maxwell action.

What is the content of this theory, both for non-topological and topological operators? Let us start with the ones that are not necessarily topological. These are the end-points of the bulk lines that are trivialized at the topological boundary. They read

$$W_b^{n/R}(\Sigma_x) = e^{i\frac{n}{R}\int_{\Sigma_x} b} = e^{in\phi(x)} \ , \qquad W_c^{mR}(\Sigma_x) = e^{imR\int_{\Sigma_x} c} = e^{im\tilde{\phi}(x)} \ , \tag{12.10}$$

where $x \in \mathcal{M}_2'$, $n, m \in \mathbb{Z}$ and $\tilde{\phi}$ is the field that quantizes the holonomies of c at the boundary, and that by virtue of (12.3) is related non-locally to ϕ by

$$\frac{1}{R}\, d\tilde{\phi} = iR * d\phi \ . \tag{12.11}$$

These are clearly local operators in the compact scalar CFT, and they are not topological as it is possible to find, using standard QFT/CFT techniques, for instance that their 2-point function has a non-trivial dependence on the distance between the two insertions.

The line operators that remain topological are the following ones

$$W_b^x = e^{ix\int b} = e^{ixR\int d\phi} \quad \text{with} \quad x \sim x + \frac{1}{R} \ . \tag{12.12}$$

We can switch to parametrizing these compact set of operators by $\beta = 2\pi R x$ so that $\beta \sim \beta + 2\pi$, and we get

$$W_b^{\frac{\beta}{2\pi R}} = e^{\frac{i\beta}{2\pi}\int d\phi} \equiv U_\beta^w \ . \tag{12.13}$$

This is what in (generalized) Maxwell we called the generator of the magnetic symmetry, but we call the winding symmetry here. It is the symmetry related to the conservation law that comes from $d^2\phi = 0$.

The other line operator is

$$W_c^y = e^{iy\int c} = e^{-y\int *b} = e^{-yR\int *d\phi} \quad \text{with} \quad y \sim y + R \ . \tag{12.14}$$

We take in this case $\alpha = \frac{2\pi}{R} y$ so that $\alpha \sim \alpha + 2\pi$, to get

$$W_c^{\frac{\alpha R}{2\pi}} = e^{-\frac{\alpha R^2}{2\pi}\int *d\phi} \equiv U_\alpha^s \ . \tag{12.15}$$

This is the equivalent of the generator of the electric symmetry, that in the present context is called shift (or momentum) symmetry, since it implements $\phi \to \phi + \alpha$.

As the 2-dimensional compact scalar is in all similar to the generalized Maxwell theories, it also follows that the shift and the winding symmetries have a mixed anomaly, whose inflow is given by (5.16)

$$S^{\text{inflow}} = \frac{i}{2\pi}\int_{\mathcal{M}_3} B_s \wedge dB_w \ . \tag{12.16}$$

Note that the inflow action is of the same form as the bulk action of the SymTFT, albeit with $U(1)$-valued fields in the inflow action, and $\mathbb{R}$-valued fields in the SymTFT.

An important comment is the following. It is clear in the SymTFT that the roles of b and c can be interchanged keeping the theory as it is. In other words there is a $\mathbb{Z}_2$ symmetry of the theory that operates this exchange. The consequence of operating this switch is that, after slab compactification, R

is exchanged with $1/R$. This means that the theory of a compact scalar of radius R is perfectly equivalent to the theory of a compact scalar of radius $1/R$. This is the statement of *T-duality* often encountered in string theory.

As a last example in these lectures, we are going to see that T-duality actually tells us that there are non-invertible symmetries in the theory of the compact scalar.

In order to see this symmetry, we first have to discuss (discrete) gaugings of the model. First recall that given the SymTFT set-up describing the theory, we know that theories defined by different topological boundary conditions, namely compact scalars with different radii, should be related to each other by topological manipulations. The latter are precisely the discrete gaugings.

Let us start with gauging a $\mathbb{Z}_q$ subgroup of the $U(1)$ winding symmetry. First, we can observe what it does on the spectrum. Of all the winding states, only those with charge a multiple of q, $e^{imq\tilde{\phi}}$, are gauge invariant. As for the momentum states, let us notice that previous to the gauging, we have a set of non-genuine operators of 'momentum' n/q that, in order to cure their fractional quantization (i.e. to restore gauge invariance), have to be attached to a line, which is an open winding symmetry topological defect:

$$e^{i\frac{n}{q}\phi(x)+i\frac{n}{q}\int_{\Sigma_x}d\phi} = e^{i\frac{n}{q}\phi(x)}U^{w}_{\frac{2\pi n}{q}}(\Sigma_x)\ . \tag{12.17}$$

(Here we momentarily focus on one end of the line and neglect the fact that there has to be another operators at the other end—or we consider the spacetime to be non-compact.) We see that the open line is precisely among the generators of the $\mathbb{Z}_q$ subgroup we are gauging, hence after gauging, it becomes transparent, and the operator above becomes genuinely local. As a result, this gauging introduces in the spectrum such momentum operators. This spectrum reflects a new periodicity for the scalar, $\phi \sim \phi + 2\pi q$, which can be reabsorbed by a rescaling of the radius $R \to Rq$.

This can be seen also at the level of the action. To gauge the $\mathbb{Z}_q$ subgroup of the winding symmetry, we have to couple a gauge field to the winding current and then impose through a BF term that the gauge field is $\mathbb{Z}_q$-valued:

$$S_{\text{gauged winding}} = \frac{R^2}{4\pi}\int d\phi \wedge *d\phi + \frac{i}{2\pi}\int d\phi \wedge A - \frac{iq}{2\pi}\int d\phi' \wedge A \quad (12.18)$$

where ϕ' is a compact scalar that imposes the $\mathbb{Z}_q$-valued nature of the $U(1)$ connection A through the sum over its windings. Integrating out A (i.e. perfoming the path integral over it) yields the relation

$$d\phi = q\ d\phi' \quad (12.19)$$

so that after substitution we get

$$S_{\text{gauged winding}} = \frac{(Rq)^2}{4\pi}\int d\phi' \wedge *d\phi' \ , \quad (12.20)$$

namely we get a new theory defined by the radius $R' = Rq$, as already argued before.

Let us now gauge a $\mathbb{Z}_p$ subgroup of the $U(1)$ shift symmetry instead. At the level of the spectrum, the roles of the winding and momentum states are exchanged, we keep only the momentum states $e^{inp\phi}$, while the fractionally quantized winding states $e^{i\frac{m}{p}\tilde{\phi}}$ become genuine because the lines $U^s_{\frac{2\pi n}{p}}$ to which they are attached are transparent after gauging. Thus we expect the gauged theory to describe a compact scalar at radius R/p.

At the Lagrangian level, we proceed as before, paying attention that the coupling of a gauge field to the shift symmetry current requires the addition of the seagull term for gauge invariance at higher order. We get

$$S_{\text{gauged shift}} = \frac{R^2}{4\pi}\int (d\phi - A) \wedge *(d\phi - A) - \frac{ip}{2\pi}\int d\tilde{\phi}' \wedge A \ , \quad (12.21)$$

where again $\tilde{\phi}'$ is a compact scalar with canonical periodicity of 2π. Integrating out A now gives

$$R^2 * (d\phi - A) = ip\ d\tilde{\phi}' \ . \quad (12.22)$$

Upon substitution, and neglecting boundary terms, we get

$$S_{\text{gauged shift}} = \frac{p^2}{4\pi R^2} \int d\tilde{\phi}' \wedge *d\tilde{\phi}' \ . \tag{12.23}$$

What we get is then the theory of a compact scalar at radius $R' = p/R$, in other words it is the T-dual theory with respect to the one we expected. This is not entirely a surprise because the relation betweem ϕ and $\tilde{\phi}'$ involves the Hodge $*$, and hence momentum and winding operators are exchanged along the substitution.

Another way to see this is to consider the case $p = 1$. Gauging a $\mathbb{Z}_1$ subgroup of the shift symmetry is a trivial operation, and it should do nothing to the theory. But implementing this trivial gauging at the level of the action we see that what we get is the theory at radius $R' = 1/R$. This is the statement of T-duality: the two theories are to be considered strictly equivalent, i.e. we have to identify the theories related by the inversion of the radii. The conformal manifold is thus a half-line, $R \geq 1$.

Now, if $R' = p/R = R$, namely when $R^2 = p \in \mathbb{Z}$, this gauging operation gives back the same theory, and it is thus a symmetry of the latter. We can actually find the topological defect that implements this symmetry. This is done as follows: we split spacetime along a 1-dimensional interface, and we perform the gauging only on one side of the interface (for simplicity, we are going to consider again spacetime to be non-compact so that the interface separates left from right). We start from

$$S = \frac{p}{4\pi} \int_L d\phi_L \wedge *d\phi_L + \frac{p}{4\pi} \int_R d\phi_R \wedge *d\phi_R + \frac{i}{2\pi} \int_I (\phi_L - \phi_R) d\phi \tag{12.24}$$

where the last term is just the trivial interface (the identity) that sets $\phi_L = \phi_R$ up to gauge transformations (periodicity). We now gauge $\mathbb{Z}_p \subset U(1)_s$ only

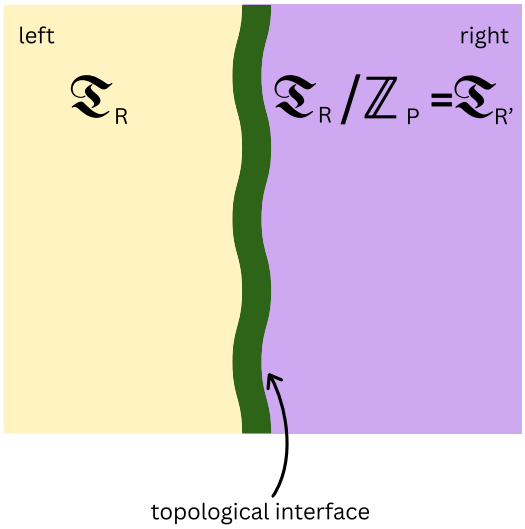

Figure 6: The left-hand side, shows the original, ungauged scalar theory with radius R, denoted as $\mathcal{T}$. On the right, we have gauged $\mathbb{Z}_p \subset U(1)$ to yield the new radius $R' = p/R$. When the original radius $R = p$, the topological interface is a non-invertible symmetry operator.

on the right side

$$\begin{aligned} S =& \frac{p}{4\pi}\int_L d\phi_L \wedge *d\phi_L + \frac{p}{4\pi}\int_R (d\phi_R - A)\wedge *(d\phi_R - A) \\ & - \frac{ip}{2\pi}\int_R d\tilde{\phi}_R \wedge A + \frac{i}{2\pi}\int_I (\phi_L - \phi_R) d\phi \ . \end{aligned} \tag{12.25}$$

Note that we set $A = 0$ at the interface, so that A is really a field defined only on the right of the interface. Integrating it out as we did before, we have to pay attention that we produce an extra term on the interface because of integration by parts

$$-\frac{ip}{2\pi}\int_R d\tilde{\phi}_R \wedge d\phi_R = \frac{ip}{2\pi}\int_I \phi_R d\tilde{\phi}_R \ . \tag{12.26}$$

Eventually the action of the theory with the interface is

$$\begin{aligned} S =& \frac{p}{4\pi}\int_L d\phi_L \wedge *d\phi_L + \frac{p}{4\pi}\int_R d\tilde{\phi}_R \wedge *d\tilde{\phi}_R \\ &+ \frac{i}{2\pi}\int_I \Big((\phi_L - \phi_R)d\phi + p\phi_R d\tilde{\phi}_R\Big) \ . \end{aligned} \tag{12.27}$$

In the first line we have the bulk action we started with, hence the second line represents an interface between two copies of the same theory. Moreover it is manifestly topological (since it does not involve the $*$), hence it is indeed a symmetry defect.

We note that the interface action involves two edge modes, i.e. modes that are defined only on the defect, ϕ and ϕ_R (of the latter, only the modes on I survive while the ones on R have been integrated out along with A). Relabeling $\phi_R \to \psi$ and $\tilde{\phi}_R \to \phi_R$, the interface action reads

$$S_I = \frac{i}{2\pi}\int_I ((\phi_L - \psi)d\phi + p\psi d\phi_R) \ . \tag{12.28}$$

We can actually integrate out ϕ and ψ, to get the equivalent expression

$$S_I = \frac{ip}{2\pi}\int_I \phi_L d\phi_R \ . \tag{12.29}$$

Looking at the boundary EOM at the interface (in the formulation with edge modes), we find the following relations

$$\begin{aligned} &\text{EOM}(\phi): && d\phi_L = d\psi \\ &\text{EOM}(\psi): && d\phi = p d\phi_R \\ &\text{EOM}(\phi_L): && ip * d\phi_L = d\phi \\ &\text{EOM}(\phi_R): && ip * d\phi_R = p d\psi \end{aligned} \tag{12.30}$$

These relations are all consistent with $i * d\phi_L = d\phi_R$, which is what we would get as boundary EOM using the action without edge modes.

We observe that the quantization of the windings $\int d\phi$, $\int d\psi \in 2\pi\mathbb{Z}$ is compatible with the quantization of the momentum and winding charge on the left, $\int ip * d\phi_L$, $\int d\phi_L \in 2\pi\mathbb{Z}$, but on the right we get

$$\int ip * d\phi_R \in 2\pi p\mathbb{Z} \ , \quad \int d\phi_R \in \frac{2\pi}{p}\mathbb{Z} \ . \tag{12.31}$$

This is what we knew, that gauging $\mathbb{Z}_p$ in the shift symmetry takes out of the spectrum momentum modes which are not multiples of p and introduces winding modes which have fractional numbers by p.

However the interface is telling us that if we bring a winding mode from the left that is *not* a multiple of p, then it does not match to any momentum mode within the spectrum on the right. The interface acts as a projector for this mode! The same can be said of a fractional winding mode which is in the spectrum on the right, but is no longer in the spectrum as a fractional momentum mode on the left. The interface, or topological symmetry defect, being a projector for some modes in the spectrum, is then inherently non-invertible. We have thus shown that T-duality at $R^2 = p > 1$ is a non-invertible symmetry.

The argument above can be easily extended to a situation where $R^2 = p/q$, by also gauging a $\mathbb{Z}_q$ subgroup of winding symmetry beforehand. What we are going to show as a final act of this section is that such a non-invertible symmetry exists at any radius R.

Let us perform the steps in all of spacetime first. We start with the theory at radius R, and we can actually *decompactify* it by gauging the entire $U(1)$ winding symmetry, with flat connections. This indeed removes all winding states, and renders genuine momentum operators with any fractional or even irrational momenta. At the level of the action we implement this as

$$S_{\text{gauged winding}} = \frac{R^2}{4\pi}\int d\phi \wedge *d\phi + \frac{i}{2\pi}\int d\phi \wedge A - \frac{i}{2\pi R}\int d\varphi \wedge A \ . \tag{12.32}$$

The field φ is now a non-compact scalar, which imposes $dA = 0$ without imposing any constraint on its holonomies (indeed $\int d\varphi = 0$ and there is no

sum over windings in its path integral). Since the normalization of a non-compact scalar is immaterial, we have chosen a convenient normalization for the last term.

Integrating out A we get $d\phi = \frac{1}{R} d\varphi$ and the action becomes

$$S_{\text{gauged winding}} = \frac{1}{4\pi} \int d\varphi \wedge *d\varphi \ . \tag{12.33}$$

This action only has a $\mathbb{R}$ shift symmetry. We can gauge a $\mathbb{Z}$ subgroup of it, in order to recompactify the scalar. Now, we have a free parameter in doing that, which is the new radius, and we can take it such that the final theory is the same as the one we started with. We then set

$$S_{\text{gauged momentum}} = \frac{1}{4\pi} \int (d\varphi - a) \wedge *(d\varphi - a) - \frac{iR}{2\pi} \int d\tilde{\phi} \wedge a \ , \tag{12.34}$$

where a is an $\mathbb{R}$-valued connection (since it is gauging a $\mathbb{R}$ symmetry) and $\tilde{\phi}$ is a compact scalar that imposes that a is flat and it has holonomies $\int a \in \frac{2\pi}{R}\mathbb{Z}$, i.e. it is effectively a $\mathbb{Z}$-valued gauge field. Integrating out a we get

$$S_{\text{gauged momentum}} = \frac{R^2}{4\pi} \int d\tilde{\phi} \wedge *d\tilde{\phi} \ , \tag{12.35}$$

so that eventually all the steps together implement a symmetry.

As before, we can again introduce the identity interface, and perform the steps just described only on its right. The step of gauging winding will just impose the identification $\phi_R = \frac{1}{R}\varphi$ in the interface action, while the step of gauging momentum produces the boundary term involving φ and $\tilde{\phi}_R$. We finally get

$$S_I = \frac{i}{2\pi} \int_I \left((\phi_L - \frac{1}{R}\varphi) d\phi + R\varphi d\phi_R \right) \tag{12.36}$$

where we relabeled $\tilde{\phi}_R \to \phi_R$. Note that here the action with the edge modes is necessary because, were we to integrate out ϕ or φ to get an action without edge modes, we would get

$$S_I = \frac{iR^2}{2\pi} \int_I \phi_L d\phi_R \tag{12.37}$$

which is *not* gauge invariant, in the sense that, for instance, under the periodic identification of ϕ_L it fails to give something proportional to $2\pi i\mathbb{Z}$, for generic R^2.

The boundary EOM are now

$$\begin{aligned} &\text{EOM}(\phi): && d\phi_L = \frac{1}{R} d\varphi \\ &\text{EOM}(\varphi): && d\phi = R^2 d\phi_R \\ &\text{EOM}(\phi_L): && iR^2 * d\phi_L = d\phi \\ &\text{EOM}(\phi_R): && iR^2 * d\phi_R = R d\varphi \end{aligned} \tag{12.38}$$

In principle, also the above EOM yield the identification $i * d\phi_L = d\phi_R$. However we see that from the left, all the winding modes are killed, while from the right, all momentum modes are killed, and winding modes with irrational momenta $\int d\phi_R \in \frac{2\pi}{R^2}\mathbb{Z}$ are introduced. Eventually, this means that this symmetry defect projects out all the non-trivial operators of theory, except the identity. It is the most non-invertible an operator can be, while preserving the vacuum.

We observe that the nature of this kind of non-invertible symmetries, is to rearrange genuine and non-genuine operators in such a way that the action and spectrum of the theory remains the same. It acts as a projector once we enforce that non-genuine operators are *not* part of the physical spectrum.

13 Further topics to explore

In these lecture notes we could not cover all aspects of generalized symmetries. There are many further topics that the curiosity-driven student could explore. We list here an obviously incomplete (and obviously biased) set of such topics, as appetizers (or desserts).

- We did not discuss spontaneous symmetry breaking. Suffice it to say

here that for higher-form symmetries, the order parameters are the extended charged objects. As for spontaneous breaking of non-invertible symmetries, it can lead to degeneracy of vacua with different physical properties (unlike for ordinary symmetries!).

- An interesting topic (and arguably the reason why higher-form symmetries were developed) is the generalized symmetries of Yang-Mills-like theories. Due to the non-abelian gauge group, it is considerably more involved than in Maxwell. The 1-form symmetry being necessarily abelian, it can be shown eventually to coincide with the center of the gauge group. If the gauge group is not simply-connected, then there is also a magnetic symmetry which is related to the fundamental group of the gauge group. Confinement is related to the spontaneous breaking, or not, of these higher-form symmetries.

- We have not used as a SymTFT the twisted BF theory that we discussed in these lectures. We can do it, and we find that it describes the symmetry of a 3d theory on its boundary that can be described as Yang-Mills-Chern-Simons. The non-trivial linking of lines in the 3d theory is related to the presence of the open tubular defects in the bulk.

- Going to 4 dimensions and to theories of (vaguely) phenomenological interest, we can have more fun with gauging symmetries that have mixed anomalies, because the anomalies (for continuous symmetries) are typically cubic. Here we discover that some situations lead to 2-groups, which are theories where 0-form and 1-form symmetries mix non-trivially (also at the level of the Ward-Takahashi identities). Other situations are in principle very well-known as the ABJ anomaly. What one can show is that the ABJ anomaly sometimes does not destroy the axial symmetry, but makes it non-invertible. This is achieved by dressing the topological defect with a non-trivial TQFT.

- We only discussed QFTs defined in the continuum. It is also interesting to consider QFTs defined on a lattice, either a full spacetime lattice, or a space-like lattice with continuous time evolution determined by a Hamiltonian (a.k.a. a quantum mechanics). Most of the symmetries we discussed can be found on the lattice too, sometimes at the price of introducing for the purpose some tweaked lattice models. The lattice can be a good starting point to consider more complicated issues in the continuum, such as generalized spacetime symmetries (i.e. symmetries like translations and rotations). Some TQFT are also better defined on lattices (or triangulations). Finally, this approach can be of help if one wants to make spacetime dynamical eventually, in a quantum gravity perspective.

Exercises

1. Derive the Ward-Takahashi identities (WTI) for correlators of local operators using the path integral formulation of a QFT with an abelian symmetry.

2. From the WTI derived above, derive the global version of the WTI, namely the action of topological symmetry defects on local charged operators.

3. Derive the symmetries, their action on charged objects, and their anomaly for the general case of p-form Maxwell. Discuss the particular case $p = 0$.

4. Identify and evaluate the 2-point function that is non-zero (at vanishing background fields) because of the mixed anomaly of p-form Maxwell.

5. Work out (with analytical expressions) the computation of the magnetic charge of the Dirac monopole. In other words, what is the field configuration such that $\int F = 2\pi$?

6. Give your arguments as to why a QFT should be defined on any spacetime manifold (or not).

7. Determine the Pontryagin duals of $\mathbb{R}$, $U(1)$, $\mathbb{Z}$ and $\mathbb{Z}_k$, namely the set of their irreducible representations, that is also a group.

8. Determine the anomaly inflow action for the (compact) BF theory.

9. Discuss the (dynamical) gauging of the 0-form shift symmetry of a scalar in d dimensions, and then the (dynamical) gauging of the $(d-2)$-form symmetry of the gauged theory.

10. Discuss and classify the possible gaugings in the mixed BF theory with symmetry $\mathbb{Z}^{(p)} \times U(1)^{(d-p-1)}$.

11. Derive the correlators of the generic surface and line operators of the twisted BF theory.

12. Discuss the effects of the twist in the 4-dimensional BF theories with non-compact connections.

13. Relate the fact that gauging a whole $\mathbb{Z}_k$ symmetry of a BF theory leads to an SPT, to the fact that the gauged theory can be seen as coupling two BF theories with a twist.