IX

LOCALITY OF HOLOMORPHIC BUNDLES, AND LOCALITY IN QUANTUM FIELD THEORY

Graeme Segal

Dedicated to Nigel Hitchin, on his 60th birthday

Nigel Hitchin has been a close colleague of mine for most of my mathematical life, and I have profited enormously from my contact with him. His wonderful skill in asking the right questions and in obtaining deep results without sacrificing concreteness or simplicity has both inspired me and filled me with envy. Having seen, however, how many others at this meeting are better qualified than I am to talk about Nigel's work, I decided it would be best to keep to my own terrain, and talk about locality in quantum field theory. All the same, many of the ideas involved are well exemplified in the study of bundles on Riemann surfaces which Nigel is famous for, and I shall begin there, especially as the question of locality relates to an aspect of his work that has not been talked about so far at this meeting, namely, its role in so-called 'geometric Langlands theory'.

We can approach the subject by contrasting two opposite ways of looking at holomorphic vector bundles E on a compact Riemann surface Σ. At one extreme, if we remove any finite set σ of points from Σ then E is trivial on the remaining surface $\Sigma - \sigma$, so we can think of all the 'twisting' of the bundle as being concentrated into tiny neighbourhoods of the points of σ. At the other extreme, we can try to spread the twist as evenly as possible over all of Σ.

The classical case of line bundles is very simple – perhaps misleadingly so. On the one hand, any line bundle L can be constructed from a *divisor* $\mathbf{z} = n_1z_1 + \cdots + n_kz_k$ – an element of the free abelian group on the set of points of Σ – by attaching the trivial bundle L_0 on $\Sigma - \{z_1, \ldots, z_k\}$ to trivial bundles L_i in the neighbourhood of each z_i by means of the clutching function $\zeta_i^{n_i}$, where ζ_i is a local parameter at z_i. (The resulting bundle $L(\mathbf{z})$ depends up to *canonical* isomorphism only on the divisor $\mathbf{z}$.) On the other hand, for any choice of Riemannian metric on Σ, any line bundle can be given a unique unitary connection with constant curvature, so that it looks exactly the same in the neighbourhood of any point of Σ. The isomorphism classes of holomorphic line bundles on Σ naturally form a commutative complex Lie group $\mathrm{Pic}(\Sigma)$, and the correspondence between the two ways of looking at a bundle amounts to the classical theorem that $\mathrm{Pic}(\Sigma)$ is – in the holomorphic category – the free abelian group on Σ, traditionally called its 'Albanese variety', that is, that the map

$\Sigma \to \mathrm{Pic}(\Sigma)$ given by $z \mapsto L(z)$ is universal[1] among holomorphic maps from Σ to a commutative group.

We encounter two main obstacles when we try to formulate an analogue of this attractive picture for higher dimensional bundles. The first is that the isomorphism classes of n-dimensional holomorphic bundles on a Riemann surface Σ do not form a nice space; and the second is that the bundles trivialized in the complement of a given point do not form an abelian group. For brevity I shall refer to these as the problems of 'noncommutative geometry' and of 'algebraic structures up to homotopy', respectively, and shall say a little about each in turn.

9.1 Noncommutative geometry

If Σ is the Riemann sphere then every n-dimensional holomorphic bundle E on Σ is a sum of line bundles, of degrees $k_1 \geq k_2 \geq \cdots \geq k_n$ say, and E is determined up to isomorphism by the n-tuple of degrees. Equivalently, E can be obtained by attaching trivial bundles on $\Sigma - \{\infty\}$ and $\Sigma - \{0\}$ by a holomorphic attaching function which is a homomorphism $\lambda : \mathbb{C}^\times \to \mathrm{GL}_n(\mathbb{C})$, unique up to conjugation in $\mathrm{GL}_n(\mathbb{C})$. Now the conjugacy classes of homomorphisms from $\mathbb{C}^\times$ to a complex reductive group G – and hence the isomorphism classes of G-bundles on the sphere – are in one-to-one correspondence with the finite-dimensional holomorphic irreducible representations of the Langlands dual group ${}^LG$. (If G is $\mathrm{GL}_n(\mathbb{C})$ then ${}^LG = G$.) This is the starting point of geometric Langlands theory, but it is not my subject here. Let us notice, however, that whereas the isomorphism classes of representations form a countable discrete set the behaviour of the bundles is quite different. When a holomorphic bundle $E^{(t)}$ on the sphere depends holomorphically on a parameter $t \in T$ we find that – if T is connected – the isomorphism class of $E^{(t)}$ is constant on a dense open subset of the parameter space T, but 'jumps' when t belongs to certain submanifolds: the space T is *stratified* by the isomorphism class of $E^{(t)}$. Thus the countable set $\mathcal{B}_\Sigma$ of isomorphism classes of bundles consists of a sequence of connected components

[1] This may seem strange. The set-theoretical free abelian group F_Σ generated by the points of Σ fits into an exact sequence

$$K_\Sigma^\times \to F_\Sigma \to \mathrm{Pic}_\Sigma \to 1,$$

where K_Σ is the field of rational functions on Σ. The group F_Σ is the disjoint union of a sequence of finite-dimensional algebraic varieties $F_\Sigma^{(n)}$, where $F_\Sigma^{(n)}$ consists of all $\Sigma n_k x_k$ such that $\Sigma|n_k| = n$. Furthermore, F_Σ has a natural topology in which the closure of $F_\Sigma^{(n)}$ is compact, and is the union of the $F_\Sigma^{(m)}$ for $m \leq n$. But we cannot say that F_Σ is any kind of algebraic variety: in fact it is easy to see that if U is a neighbourhood of the identity element of F_Σ then any continuous $f : U \to \mathbb{C}$ for which $f|U \cap F_\Sigma^{(n)}$ is holomorphic for each n has to be constant along the orbits of $K_\Sigma^\times$.

indexed by the first Chern class of E, but each connected component, though infinite, is the closure of a single point.

To make better sense of unpromising 'spaces' such as $\mathcal{B}_\Sigma$ a number of different approaches are commonly used. In algebraic geometry the main candidates are

1. To regard $\mathcal{B}_\Sigma$ as a *stack*, that is, to work with the *category* – actually a *groupoid* – of bundles and their isomorphisms rather than just with the set $\pi_0(\mathcal{B}_\Sigma)$ of isomorphism classes of objects
2. the approach of *geometric invariant theory*, which picks out a class of 'stable' bundles whose isomorphism classes do form a nice space, in fact an algebraic variety

Here I am going to talk about the first approach, which is more obviously related to quantum field theory, especially in the treatment of the Langlands theory by Kapustin and Witten (2007). Nigel Hitchin's own main tool, however, was geometric invariant theory.

Stepping back a little from algebraic geometry, one can say that the category of bundles on a space resembles the category of representations of a group. For example, the Hitchin moduli space associated to a surface Σ and a compact group G is – among other things – the space of conjugacy classes of homomorphisms from the fundamental group $\pi_1(\Sigma)$ to the complexified group $G_\mathbb{C}$. The 'space' of irreducible unitary representations of a group Γ is an archetypal example in Connes's theory of noncommutative geometry (Connes 1994). He observes that the set of irreducible representations is the set $\mathrm{Spec}(A_\Gamma)$ of irreducible $*$-representations of a C*-algebra A_Γ associated to Γ. If A is a commutative algebra its irreducible representations are the algebra-homomorphisms $A \to \mathbb{C}$, which form a topological space $\mathrm{Spec}(A)$ on which A is an algebra of continuous complex-valued functions. Connes's idea is to think of the geometry of the set of irreducible representations of the group Γ as *defined* by the noncommutative algebra A_Γ rather than by the – often too small – commutative algebra of continuous functions on $\mathrm{Spec}(A_\Gamma)$, which in fact is the *centre* of A_Γ.

To relate this picture to the stacks or groupoids of algebraic geometry we think of a groupoid as halfway between a group and a space. More precisely, a groupoid $\mathcal{B}$ consists of two sets together with some maps between them: a set $\mathcal{B}_0$ of objects, and a set $\mathcal{B}_1$ of morphisms which is the disjoint union of the sets $\mathcal{B}(x,y)$ of morphisms from x to y, where x and y run through all the objects in $\mathcal{B}_0$. We are concerned here, however, with *topological* groupoids, for which the sets $\mathcal{B}_0$ and $\mathcal{B}_1$ have topologies and the structural maps between them are continuous. At one extreme, if $\mathcal{B}_0$ is a point then we have a topological group; at the other, if $\mathcal{B}_0 = \mathcal{B}_1$, we have no morphisms except identities, and the groupoid is simply a space. As Connes has emphasized, a groupoid $\mathcal{B}$ has a groupoid-algebra $A_\mathcal{B}$, which interpolates between the group-algebra of a group and the commutative algebra of functions on a space. (The algebra $A_\mathcal{B}$ is devised so that an $A_\mathcal{B}$-module is the same thing as a functor from $\mathcal{B}$ to vector spaces. For a discrete groupoid, $A_\mathcal{B}$ has a vector space basis e_f indexed by the set

$\mathcal{B}_1$ of morphisms, and composition is given by $e_f e_g = e_{f \circ g}$ when f and g are composable, and $e_f e_g = 0$ otherwise. For the actual groupoids at hand, we must make a choice of what we mean by the algebra, just as for a Lie group we can define different group algebras depending on the geometric context in which we are working.)

The groupoids $\mathcal{B}$ we are concerned with all arise from the action of a group G on a space $\mathcal{B}_0$, so that $\mathcal{B}_1 = G \times \mathcal{B}_0$, and the set of morphisms from x to y is $\{g \in G : gx = y\}$. I shall denote this groupoid by $\mathcal{B}_0 /\!/ G$. The algebra $A_{\mathcal{B}}$ in this case is the twisted group algebra $G \ltimes A_0$, where A_0 is the commutative algebra of functions on $\mathcal{B}_0$.

The essential features of a groupoid $\mathcal{B}$ are the (often not very nice) space $\pi_0(\mathcal{B})$ of isomorphism classes of its objects, and the groups $\mathrm{Aut}(x)$ of automorphisms of the objects, which are the obstructions to $\mathcal{B}$'s being a space. These features depend only on the equivalence class of the groupoid (in the sense of category theory), and we are interested in groupoids only up to equivalence. In fact we are content (cf. Segal 1973, 2.8) even with what I shall call 'weak equivalences': functors $T : \mathcal{B} \to \mathcal{B}'$ which induce bijections $\pi_0(\mathcal{B}) \to \pi_0(\mathcal{B}')$ and $\mathrm{Aut}(x) \to \mathrm{Aut}(T(x))$ for each object x, and which are 'covering maps' (in the sense that they have local cross-sections) on the space of objects and morphisms. Thus if G acts freely on $\mathcal{B}_0$, making it a locally trivial principal G-bundle on the space $\mathcal{B}_0/G$, then we are not interested in the difference between the groupoid $\mathcal{B}_0 /\!/ G$ and the space $\mathcal{B}_0/G$. Similarly, whenever we decompose a closed surface Σ into two pieces Σ_1 and Σ_2 intersecting in a curve S we can identify the set of isomorphism classes of n-dimensional bundles on Σ with the (non-Hausdorff) double-coset space

$$G_{\Sigma_1} \backslash G_S / G_{\Sigma_2},$$

where G_{Σ_i} is the group of holomorphic maps from Σ_i to $G = \mathrm{GL}_n(\mathbb{C})$, and G_S is the group of smooth maps $S \to G$, for any bundle can be trivialized on a non-closed surface. We then have three groupoids:

1. $\mathcal{B}^{(1)}$ formed by the action of G_{Σ_1} on the homogeneous space G_S/G_{Σ_2}
2. $\mathcal{B}^{(2)}$ formed by the action of G_{Σ_2} on $G_{\Sigma_1} \backslash G_S$
3. $\mathcal{B}^{(12)}$ formed by the action of $G_{\Sigma_1} \times G_{\Sigma_2}$ on G_S

All three are weakly equivalent, and are weakly equivalent to the groupoids obtained from any other way of decomposing Σ into pieces. According to the 'stack' point of view, any of them can be taken as the 'space' of bundles on Σ.

The relation of equivalence of groupoids translates into *Morita equivalence* of algebras. Two algebras are Morita equivalent if their categories of left modules are equivalent. Recall that if A and B are algebras then an (A, B)-bimodule F defines an additive functor $F_* : \mathcal{M}_B \to \mathcal{M}_A$, where $\mathcal{M}_A$ and $\mathcal{M}_B$ are the categories of left A and B modules, by $F_*(M) = F \otimes_B M$. In particular, for the

(A, A)-bimodule A we have $A_* = \mathrm{id}_{\mathcal{M}_A}$, and if G is a (B, C)-bimodule we have

$$F_* \circ G_* = (F \otimes_B G)_* : \mathcal{M}_C \to \mathcal{M}_A.$$

On the level of discrete groupoids, at least, there is a dictionary

$$\begin{array}{ccc} \text{groupoid } \mathcal{B} & \longmapsto & \text{algebra } A_{\mathcal{B}} \\ \text{functor } T : \mathcal{B} \to \mathcal{B}' & \longmapsto & (A_{\mathcal{B}}, A_{\mathcal{B}'})\text{-bimodule } F_T \\ \text{natural transformation } \Phi : T \to T' & \longmapsto & \text{isomorphism } \Phi_* : F_T \to F_{T'}. \end{array}$$

In the light of the previous discussion we see that a 'noncommutative space' is defined not by an algebra but by a Morita equivalence class of algebras, and so it is more naturally described by a *linear category* – the category of left modules for the algebra. In holomorphic geometry, however, the algebras that arise, even when they are commutative, are not semisimple, and their module categories do not have very convenient properties. It is therefore better to go further, and replace the module categories by the categories of cochain complexes of modules. That is why the geometric Langlands correspondence is stated in terms of the linear categories of complexes of coherent sheaves[2] – or of $\mathcal{D}$-modules – on the moduli spaces of bundles.

Yet another version – slightly more general still – of the notion of noncommutative space arises from quantum field theory, and it is perhaps the most natural one in the geometric Langlands theory. I shall sketch it below. Before leaving the present discussion, however, it may be worth making another remark.

A generalized space defined by a topological groupoid has a *homotopy type*, just like an ordinary space. For – like any topological category (cf. Segal 1968) – the groupoid $\mathcal{B}_0//G$ has a 'realization' $|\mathcal{B}_0//G|$ as a space, and equivalent groupoids have homotopy-equivalent realizations. A generalized space defined by a noncommutative ring – or, better, by a linear category $\mathcal{C}$ – has no such homotopy type. The best one can do is consider the *stable* homotopy type (or *spectrum*) $\mathbb{K}_{\mathcal{C}}$ defined by applying the usual K-theory construction to $\mathcal{C}$: this is (cf. Segal 1977) the 'space' whose homotopy groups are the K-groups of the category.

In the case of the stack $\mathcal{B}_\Sigma$ of holomorphic G-bundles on a Riemann surface Σ the space $|\mathcal{B}_\Sigma|$ has the same homotopy type as the realization of the corresponding topological groupoid $\mathcal{B}_\Sigma^{sm}$ of smooth bundles, which in turn has the homotopy type of the space $\mathrm{Map}(\Sigma; BG)$ of continuous maps from Σ to the classifying space of G. (This follows at once from the fact that for a non-closed Riemann surface Σ_i the space of holomorphic maps from Σ_i to G has the homotopy type of the space of continuous maps.) On the other hand, we know from the beautiful work of Atiyah and Bott 1982 that the moduli space of holomorphic bundles in the sense of geometric invariant theory can be identified with the minimum level of the Yang–Mills functional on the space of smooth bundles.

[2] *Sheaves* rather than modules, because the algebraic varieties are not affine. But that is a standard technicality.

9.2 Algebraic structures up to homotopy

Let us consider the space $\hat{\mathcal{B}}_\Sigma$ of pairs (E, e) consisting of a holomorphic vector bundle E on Σ equipped with a meromorphic trivialization e – that is, E is trivialized, in the complement of a finite subset σ of Σ, by n holomorphic sections $e_1, \ldots, e_n$ which extend to meromorphic sections on Σ. A pair (E, e) is completely determined by giving σ and, for each $z \in \sigma$, the local information consisting of a bundle in a neighbourhood of z trivialized away from z. The items of local information can be prescribed independently, so $\hat{\mathcal{B}}_\Sigma$ can be regarded as a *labelled configuration space*. The group $\mathrm{GL}_n(K_\Sigma)$ acts on $\hat{\mathcal{B}}_\Sigma$ by changing the meromorphic trivialization, and the orbit space is the space of isomorphism classes of holomorphic bundles. (The isotropy group of any pair (E, e) is just the group of holomorphic automorphisms of E.) Let us now recall a few aspects of the theory of labelled configuation spaces.

For an d-dimensional manifold M let $\check{C}(M)$ denote the manifold of all finite unordered subsets of M, with its natural topology in which it is the disconnected union $\coprod_{k \geq 0} C_k(M)$ of the spaces $C_k(M)$ of subsets with exactly k elements. If P is an arbitrary auxiliary space, we can also form $\check{C}(M; P) = \coprod C_k(M; P)$, whose points are the finite subsets σ of M with each point $z \in \sigma$ 'labelled' by a point p_z of P. The space $C_k(M; P)$ is fibred over $C_k(M)$ with fibre P^k.

The configuration spaces we are interested in, however, have a topology which allows the points of a configuration σ to move continuously into coincidence and the labels at the same time to amalgamate – or 'add' – in some sense. For example, the free abelian group F_Σ is the configuration space of Σ labelled by the group $\mathbb{Z}$, with the usual addition when points merge. We can define such a 'configuration space with amalgamation' whenever the labelling space P has a composition-law which is sufficiently associative and commutative to make it what is called in homotopy theory an *d-fold loop space* – in fact the existence of the amalgamated configuration space $C(M; P)$ for all d-manifolds M is a good way of *defining* an d-fold loop space.[3] To be precise, I shall say, in the style of Beilinson and Drinfeld, that an *amalgamated configuration space* is any space $C(M; P)$ to each of whose points is associated a finite subset σ of M called its *support*, and which is equipped with a map $i_\mathbf{z} : P^k \to C(M; P)$, for each sequence $\mathbf{z} = \{z_1, \ldots, z_k\}$ of distinct points of M, with the properties:

(i) the image of $i_\mathbf{z}$ consists of configurations with support $\mathbf{z}$, and
(ii) if U is the disjoint union of k small open balls $U_1, \ldots, U_k$ such that U_i contains z_i, and $C_U(M; P)$ denotes the part of $C(M; P)$ with supports in U, then $i_\mathbf{z} : P^k \to C_U(M; P)$ is a homotopy equivalence.

[3] The conventional way to define the structure of an d-fold loop space on a space P is to give the amalgamation maps

$$C_k(\mathbb{R}^d; P) \to P$$

for each $k > 0$; but these must of course satisfy various compatibilities.

Although it is not strictly necessary, we may as well assume that $C_U(M;P)$ is an open subset of $C(M;P)$ when U is an open subset of M, and that it can be identified with $C(U;P)$. That is certainly true in our examples, where $C(M;P)$ is simply $\check{C}(M;P)$ with a coarser topology.

In fact we need a slight generalization to include *bundles* of d-fold loop spaces. We shall allow the label of a point $z \in M$ to lie not in a fixed space P but rather in the fibre at z of a bundle P on the manifold M. It is clear how the definition of $C(M;P)$ should be adapted to this case.

When we have a single d-fold loop space P we define its *d-fold classifying space* – or dth 'delooping' – as the space

$$B^d P = C(U;P)/C_V(U;P),$$

where U is an open ball – say $U = \{z \in \mathbb{R}^d : \| z \| < 1\}$ – and V is the annular region $\{z \in U : 1/2 \leq \| z \| < 1\}$, and the notation means that the subspace $C_V(U;P)$ of $C(U;P)$ is collapsed to a single point, which is a natural base point in $B^d P$.

Similarly, when P is a bundle of d-fold loop spaces on M we can define a bundle $B^d P$ whose fibre at $z \in M$ is constructed using a neighbourhood U of z in M.

The main theorem about labelled configuration spaces is the following.

Theorem 9.1 *There is a natural map*

$$C(M;P) \longrightarrow \Gamma_{\text{cpt}}(M;B^d P),$$

where Γ_{cpt} *denotes the space of cross-sections which are equal to the base point outside of a compact subset of* M*. If the composition law of* P *makes the set of components* $\pi_0(P)$ *into a group then the map is a homotopy equivalence.*

Notice that applying the theorem when M is an open ball tells us that P is homotopy-equivalent to the d-fold based loop space of $B^d P$.

The equivalences are defined by the *scanning map* (see McDuff 1975, 1977 and Segal 1979) which associates to an element $c \in C(M;P)$, the section of $B^d P$ whose value at $z \in M$ is the image of c in $C(M;P)/C_{M-W}(M;P) = C(U;P)/C_V(U;P)$, where $W \subset U$ are two concentric open balls around z, and $V = U - W$. The theorem is very easy to prove, and not at all deep, in the form I have stated it here: in applications the main difficulty may be to verify the hypotheses.

In the application to holomorphic vector bundles on a surface, at a point z on the surface, with neighbourhood U, the labelling space P is the quotient $\mathcal{G}_{\check{U}}/\mathcal{G}_U$, where $\mathcal{G}_{\check{U}}$ is the group of holomorphic maps $U - \{z\} \to \mathrm{GL}_n(\mathbb{C})$ which extend meromorphically over U, and $\mathcal{G}_U$ is the group of holomorphic maps $U \to \mathrm{GL}_n(\mathbb{C})$. This can be identified with $\mathrm{GL}_n(\mathbb{C}(t))/\mathrm{GL}_n(\mathbb{C}[t])$. It is (see Pressley and Segal 1986) the union of a sequence of compact finite-dimensional algebraic varieties, and it has the homotopy type of the based loop space of $\mathrm{GL}_n(\mathbb{C})$. To define the topology of $\hat{\mathcal{B}}_\Sigma$ we first define a topology on the part $\hat{\mathcal{B}}_{\Sigma,U}$ with support in a disjoint union U of open discs by identifying it with $\mathcal{G}_{\partial U}/\mathcal{G}_U$, where $\mathcal{G}_{\partial U}$ is the group of smooth maps $\partial U \to \mathrm{GL}_n(\mathbb{C})$ which are boundary values of meromorphic maps $U \to \mathrm{GL}_n(\mathbb{C})$, and then we give $\hat{\mathcal{B}}_\Sigma$ the finest topology compatible with these. Having checked the hypotheses of the theorem, it tells us that $\hat{\mathcal{B}}_\Sigma$ has the homotopy type of the space of continuous maps from Σ to $B^2P \simeq \mathrm{BGL}_n(\mathbb{C})$, that is, the homotopy type of the space of smooth bundles, as we expect from the stack picture.

Before leaving this topic, let us mention the converse question to the one answered by the theorem. If we are given a space Q, can we model the mapping spaces $\mathrm{Map}(M; Q)$ for varying d-manifolds M – at least up to homotopy – by labelled configuration spaces $C(M; P)$? The answer, clearly, is: if and only if the space Q is d-connected; for the delooping B^dP of any d-fold loop space P is d-connected. If we want to model spaces of maps into less highly connected spaces Q we would have to allow not just 'particles' but also configurations of higher dimensional submanifolds – presumably, up to dimension m if Q is $(d-m)$-connected. (One way to see this is to realize Q as an open manifold with a Morse function with critical points of indices $\geq d-m$, and to make maps $M \to Q$ flow downwards along the gradient flow.)

9.3 Quantum field theory

All the mathematical phenomena I have been discussing play an important role in quantum field theory. In particular, noncommutative geometry enters in two somewhat opposite ways, first because the moduli spaces of theories are noncommutative spaces, and also because field theories can be used to give a new formulation of noncommutative geometry. It is only the second aspect that I am going to talk about. But let us begin at the beginning. . . .

In quantum mechanics a system is described at any time by giving an algebra $\mathcal{A}$ of 'observables' and a linear map $\theta : \mathcal{A} \to \mathbb{C}$ called the 'state'.[4] In quantum field theory, in space-time of dimension d, this picture is enriched by supposing that the observables are spread out over a given space-time manifold M. More precisely, for each $x \in M$ there is given a sub-vector space $\mathcal{O}_x$ of $\mathcal{A}$ formed by the observables which can be measured in the neighbourhood of x. Quantum field theory assumes that – up to some global topological effects[5] to which I shall return at the end of this talk – the complete information about the system is contained in the maps

$$\Theta_k : \mathcal{O}_{x_1} \otimes \cdots \otimes \mathcal{O}_{x_k} \longrightarrow \mathbb{C},$$

[4] It is perhaps more usual to say that $\mathcal{A}$ is an algebra of operators in a Hilbert space $\mathcal{H}$, and that the state is a unit vector ψ in $\mathcal{H}$ related to θ by $\theta(a) = \langle \psi, a\,\psi \rangle$. But the equivalent description by $(\mathcal{A}, \theta)$ seems more satisfactory to me.

[5] The essential example is the Bohm–Aharonov phenomenon, when an electromagnetic field in a non-simply connected region is described by a flat connection which is undetectable in any simply connected subregion.

for each finite set $\{x_1, \ldots, x_k\}$ of distinct points of M, which are got by multiplying in $\mathcal{A}$ and composing with θ. The Θ_k are traditionally called *vacuum expectation values.*

It is not easy to say what properties the vector spaces $\mathcal{O}_x$ and the functions Θ_k must have for them to constitute a quantum field theory. A first attempt at an answer can be given by defining a d-dimensional theory as a rule which

1. associates a complex topological vector space $\mathcal{H}_Y$ to each compact-oriented Riemannian manifold Y of dimension $d-1$, functorially with respect to diffeomorphisms $Y \to Y'$, and
2. associates a trace-class operator $U_X : \mathcal{H}_{Y_0} \to \mathcal{H}_{Y_1}$ to each oriented Riemannian cobordism X from Y_0 to Y_1.

These data are constrained to satisfy two axioms:

(a) *Concatenation*:

$$U_{X' \circ X} = U_{X'} \circ U_X$$

when $X' \circ X$ is the cobordism from Y_0 to Y_2 obtained by concatenating X from Y_0 to Y_1 with X' from Y_1 to Y_2.

(b) *Tensoring*: We are given associative natural isomorphisms

$$\mathcal{H}_Y \otimes \mathcal{H}_{Y'} \xrightarrow{\cong} \mathcal{H}_{Y \sqcup Y'}$$

$$U_X \otimes U_{X'} = U_{X \sqcup X'}$$

when we have disjoint unions $Y \sqcup Y'$ or $X \sqcup X'$ of $(d-1)$-manifolds or cobordisms.

Notice that it follows from property (a) that $\mathcal{H}_Y = \mathbb{C}$ if Y is the empty $(d-1)$-manifold.

When we have a theory in this sense we can reconstruct the local observables and their expectation values. We define the vector space $\mathcal{O}_x$ of observables for each point x in a closed d-manifold M by

$$\mathcal{O}_x = \lim_{\leftarrow} \mathcal{H}_{\partial D},$$

where the inverse limit is over the ordered set of all closed balls D in M which are neighbourhoods of x, ordered by

$$D' > D \Longleftrightarrow D' \subset \overset{\circ}{D},$$

in which case we have a canonical map $U_{D - \overset{\circ}{D'}} : \mathcal{H}_{\partial D'} \to \mathcal{H}_{\partial D}$ defined by the annular cobordism.

If $x_1, \ldots, x_k$ are distinct points of M, and $D_1, \ldots, D_k$ are disjoint discs with x_i in the interior of D_i, let M_0 denote the manifold with boundary obtained by deleting from M the interiors of the discs D_i. We regard M_0 as a cobordism from $\coprod \partial D_i$ to the empty manifold, and define

$$\Theta_k : \mathcal{O}_{x_1} \otimes \cdots \otimes \mathcal{O}_{x_k} \longrightarrow \mathbb{C}$$

as the inverse limit of the maps

$$U_{M_0} : \mathcal{H}_{\partial D_1} \otimes \cdots \otimes \mathcal{H}_{\partial D_k} \to \mathbb{C}$$

as the discs D_i shrink to points around the points x_i.

If the points x_i are all contained in the interior of a disc D then the map Θ_k clearly factorizes through $\mathcal{H}_{\partial D}$. We can interpret this as saying that the local observables have the structure of an *algebra* which is associative and commutative up to homotopy, for if D is a small neighbourhood of a point $x \in M$ then we can regard $\mathcal{H}_{\partial D}$ as a *completion* $\hat{\mathcal{O}}_x$ of $\mathcal{O}_x$ (for the maps $\mathcal{H}_{\partial D'} \to \mathcal{H}_{\partial D}$ in the system defining $\mathcal{O}_x$ are always injective with dense image), and we have a map

$$\mathcal{O}_{x_1} \otimes \cdots \otimes \mathcal{O}_{x_k} \to \hat{\mathcal{O}}_x$$

for any family $x_1, \ldots, x_k$ of distinct points sufficiently close to x. This is an exact linear analogue of the d-fold loop space structures discussed above, and it is what in traditional quantum field theory is called the *operator product expansion.*

It is still not clear, however, that we have made our definition of a quantum field theory sufficiently rigid. One feels that the vector space $\mathcal{H}_Y$ associated to a $(d-1)$-manifold Y should be constructed *locally* from Y, and perhaps more axioms are needed to ensure this. A natural second approximation to the definition is the notion of a *three-tier theory*, which gives an additional layer of structure to allow $(d-1)$-manifolds to be cut into pieces. In a three-tier theory

1. to each compact oriented Riemannian $(d-2)$-manifold Z there is associated a linear category $\mathcal{C}_Z$,
2. to each $(d-1)$-dimensional Riemannian cobordism Y from Z_0 to Z_1 there is associated an additive functor $\mathcal{H}_Y : \mathcal{C}_{Z_0} \to \mathcal{C}_{Z_1}$, and
3. to each d-dimensional Riemannian cobordism X from Y to Y', where Y and Y' are cobordisms from Z_0 to Z_1, there is associated a transformation of functors $U_X : \mathcal{H}_Y \to \mathcal{H}_{Y'}$.

As with the earlier definition, the data are required to satisfy two axioms of concatenation and tensoring. A theory in the earlier two-tier sense is obtained from the three-tier structure by restricting to closed $(d-1)$-manifolds, which can be regarded as cobordisms from the empty $(d-2)$-manifold $\emptyset$ to itself. The tensoring axiom implies that $\mathcal{C}_\emptyset$ is the category of vector spaces, and since any additive functor $\mathcal{C}_\emptyset \to \mathcal{C}_\emptyset$ is given by tensoring with a vector space, we can identify $\mathcal{H}_Y$ with a vector space when Y is closed.

In the form I have just stated, the three-tier definition is too vague to be of much use. In this talk I shall not try to elaborate it, as my purpose is to make just two points. The first is that a three-tier two-dimensional field theory seems to have a good claim as a candidate definition of a 'noncommutative manifold'. For, schematically at least, a natural way to give the data of a three-tier theory is to associate an algebra $\mathcal{A}_Z$ to each $(d-2)$-manifold, and to take $\mathcal{C}_Z$ to be the category of left $\mathcal{A}_Z$-modules; then to a cobordism Y is associated an $(\mathcal{A}_{Z_1}, \mathcal{A}_{Z_0})$-bimodule $\mathcal{H}_Y$, which defines a functor $\mathcal{C}_{Z_0} \to \mathcal{C}_{Z_1}$ by

$$M \mapsto \mathcal{H}_Y \otimes_{\mathcal{A}_{Z_0}} M;$$

and to a cobordism between cobordisms is associated a homomorphism of bimodules. When $d = 2$ this simply means that we have a dual pair of linear categories – the left and right modules for an algebra – associated to a point with its two orientations, while the one-dimensional data expresses the categorical duality, and the two-dimensional data gives us 'trace' or 'integration' maps. The field theory even leads one naturally from categories of modules to categories of cochain complexes of modules, if we assume that the field theory is supersymmetric.

The idea that two-dimensional theories should replace manifolds is, of course, the central proposal of string theory, which models space-time by a two-dimensional conformal field theory, with the category of D-branes in space-time as the category which the field theory associates to a point. It is also what arises in the Kapustin–Witten treatment of geometric Langlands duality. There, one begins from the maximally supersymmetric four-dimensional Yang–Mills theory associated to a compact group G, and observes that, for any compact surface Σ, a four-dimensional theory gives a two-dimensional theory by dimensional reduction along Σ – that is, by composing with the functor $M \mapsto \Sigma \times M$ from i-manifolds to $(i+2)$-manifolds. The two-dimensional theory obtained from Yang–Mills theory for G by reducing along Σ is supposed to associate to a point the category of $\mathcal{D}$-modules on the moduli space of holomorphic $G_{\mathbb{C}}$-bundles on Σ.

My second point is even vaguer. The obvious fear, if one starts to study three-tier theories, is that one will be impelled to believe that a d-dimensional theory should really mean a $(d+1)$-tier d-dimensional theory, which associates a two-category to a manifold of dimension $d-3$, and even worse things to lower dimensional manifolds, until one gets to a $(d-1)$-category associated to a point.

This is not completely mad, as it works well in the one famous example afforded by three-dimensional Chern–Simons theory for a compact group G at a given 'level'. There, the category associated to a circle S is the category of positive energy representations of the loop group of maps $S \to G$ at the specified level, and the two-category associated to a point is the same thing, but remembering its tensor structure coming from the *fusion* of loop group representations. (We think of this as a two-category with just one object, whose linear category of endomorphisms is the category of loop group representations, with fusion as its composition law.)

I cannot believe, however, that genuine – non-topological – quantum field theory will be advanced by higher categories. The algebras-up-to-homotopy formed by the field operators look much more promising. It is interesting that, in the context of homotopy theory, a d-fold loop space P does indeed give rise to a precise analogue of the structure of a $(d+1)$-tier field theory, as follows.

1. To a closed d-manifold X we associate the space $Q_X = C(X;P)$.
2. To a closed $(d-1)$-manifold Y we associate the 'group' – that is, one-fold loop space – $\mathcal{G}_Y = C(I \times Y; P)$, where I is an open interval.
3. To a closed d-manifold X with boundary we associate the $\mathcal{G}_{\partial X}$-space $Q_X = C(X^\circ; P)$, where X° is the interior of X, noticing that if $X = X_1 \cup_Y X_2$ then $Q_X \simeq Q_{X_1} \times_{\mathcal{G}_Y} Q_{X_2}$. And, more generally, to a cobordism from Y_0 to Y_1 we can associate a $(\mathcal{G}_{Y_0} \times \mathcal{G}_{Y_1})$-space, and hence a functor from $\mathcal{G}_{Y_0}$ spaces to $\mathcal{G}_{Y_1}$ spaces.
4. To a closed $(d-2)$-manifold Z we associate the two-fold loop space $\mathcal{G}_Z = C(I \times I \times Z; P)$, noticing that if Y is a $(d-1)$-manifold with boundary then $\mathcal{G}_Y = C(I \times Y^\circ; P)$ is a one-fold loop space on which the two-fold loop space $\mathcal{G}_{\partial Y}$ acts by maps of one-fold loop spaces.

And so on – I shall not spell out all the details.

I hope that a linear and analytical version of this picture is the model for quantum field theory. One point where the analogy may help is in considering whether one should expect that the d-fold algebra of field operators defined in a small d-dimensional ball should determine the theory on an arbitrary space-time, in the way that the d-fold loop space P determines all the spaces $C(X;P)$. We saw that a mapping space $\mathrm{Map}(X;Q)$ is of the form $C(X;P)$ only when the target space Q is d-connected, and I would guess that an analogous distinction applies to field theories: just as the mapping space cannot be modelled by particles, but requires m-dimensional 'objects', if Q is only $(d-m)$-connected, so we know that a field theory in general has non-local observables that can be seen only in topologically non-trivial regions of space-time. It seems, however, that the non-local observables are 'topological' in the sense that they contribute only a finite number of degrees of freedom to the infinite-dimensional physical system that the field theory describes.

References

Atiyah, M. F. and Bott, R. (1982). The Yang-Mills equations over Riemann surfaces. *Phil. Trans. Roy. Soc. London* **308A**, 523–615.

Connes, A. (1994). *Noncommutative Geometry.* Academic Press, San Diego, CA.

Kapustin, A. and Witten, E. (2007). Electro-magnetic duality and the geometric Langlands program. *Commun. Number Theory Phys.* **1**, 1–236. hep-th/0610210.

McDuff, D. (1975). Configuration spaces of positive and negative particles. *Topology* **14**, 91–107.

McDuff, D. (1977). Configuration spaces. In *K-Theory and Operator Algebras*. Proceedings Georgia 1975, Springer Lecture Notes **575**, pp. 88–95.

Pressley, A. and Segal, G. (1986). *Loop Groups*. Oxford University Press, Oxford.

Segal, G. (1968). Classifying spaces and spectral sequences. *Proc. Math. IHES Paris* **34**, 105–12.

Segal, G. (1973). Configuration spaces and iterated loop-spaces. *Invent. Math.*, **21**, 213–21.

Segal, G. (1977). K-homology theory and algebraic K-theory. In *K-Theory and Operator Algebras*. Proceedings Georgia 1975, Springer Lecture Notes **575**, pp. 113–27.

Segal, G. (1979). The topology of spaces of rational functions. *Acta Math* **143**, 39–72.