1
Introduction

This two-volume book provides the analog, in quantum field theory, of the deformation quantization approach to quantum mechanics. In this introduction, we start by recalling how deformation quantization works in quantum mechanics.

The collection of observables in a quantum mechanical system forms an associative algebra. The observables of a classical mechanical system form a Poisson algebra. In the deformation quantization approach to quantum mechanics, one starts with a Poisson algebra A^{cl} and attempts to construct an associative algebra A^q, which is an algebra flat over the ring $\mathbb{C}[[\hbar]]$, together with an isomorphism of associative algebras $A^q/\hbar \cong A^{cl}$. In addition, if $a, b \in A^{cl}$, and $\widetilde{a}, \widetilde{b}$ are any lifts of a, b to A^q, then

$$\lim_{\hbar \to 0} \frac{1}{\hbar}[\widetilde{a}, \widetilde{b}] = \{a, b\} \in A^{cl}.$$

Thus, A^{cl} is recovered in the $\hbar \to 0$ limit, i.e., the classical limit.

We will describe an analogous approach to studying perturbative quantum field theory. To do this, we need to explain the following.

- *The structure present on the collection of observables of a* classical *field theory.* This structure is the analog, in the world of field theory, of the commutative algebra that appears in classical mechanics. We call this structure a commutative factorization algebra.
- *The structure present on the collection of observables of a* quantum *field theory.* This structure is that of a factorization algebra. We view our definition of factorization algebra as a differential geometric analog of a definition introduced by Beilinson and Drinfeld. However, the definition we use is very closely related to other definitions in the literature, in particular to the Segal axioms.

- *The additional structure on the commutative factorization algebra associated to a classical field theory that makes it "want" to quantize.* This structure is the analog, in the world of field theory, of the Poisson bracket on the commutative algebra of observables.
- *The deformation quantization theorem we prove.* This states that, provided certain obstruction groups vanish, the classical factorization algebra associated to a classical field theory admits a quantization. Further, the set of quantizations is parametrized, order by order in $\hbar$, by the space of deformations of the Lagrangian describing the classical theory.

This quantization theorem is proved using the physicists' techniques of perturbative renormalization, as developed mathematically in Costello (2011b). We claim that this theorem is a mathematical encoding of the perturbative methods developed by physicists.

This quantization theorem applies to many examples of physical interest, including pure Yang–Mills theory and σ-models. For pure Yang–Mills theory, it is shown in Costello (2011b) that the relevant obstruction groups vanish and that the deformation group is one-dimensional; thus there exists a one-parameter family of quantizations. In Li and Li (2016), the topological B-model with target a complex manifold X is constructed; the obstruction to quantization is that X be Calabi–Yau. Li and Li show that the observables and correlations functions recovered by their quantization agree with well-known formulas. Grady, Li, and Li (2015) describe a one-dimensional σ-model with a smooth symplectic manifold as target and show how it recovers Fedosov quantization. Other examples are considered in Costello (2010, 2011a), Costello and Li (2011), and Gwilliam and Grady (2014).

We will explain how (under certain additional hypotheses) the factorization algebra associated to a perturbative quantum field theory encodes the correlation functions of the theory. This fact justifies the assertion that factorization algebras encode a large part of quantum field theory.

This work is split into two volumes. Volume 1 develops the theory of factorization algebras and explains how the simplest quantum field theories – free theories – fit into this language. We also show in this volume how factorization algebras provide a convenient unifying language for many concepts in topological and quantum field theory. Volume 2, which is more technical, derives the link between the concept of perturbative quantum field theory as developed in Costello (2011b) and the theory of factorization algebras.

1.1 The Motivating Example of Quantum Mechanics

The model problems of classical and quantum mechanics involve a particle moving in some Euclidean space $\mathbb{R}^n$ under the influence of some fixed field. Our main goal in this section is to describe these model problems in a way that makes the idea of a factorization algebra (Section 6.1 in Chapter 6) emerge naturally, but we also hope to give mathematicians some sense of the physical meaning of terms such as "field" and "observable." We will not worry about making precise definitions, since that's what this book aims to do. As a narrative strategy, we describe a kind of cartoon of a physical experiment, and we ask that physicists accept this cartoon as a friendly caricature elucidating the features of physics we most want to emphasize.

1.1.1 A Particle in a Box

For the general framework we want to present, the details of the physical system under study are not so important. However, for concreteness, we focus attention on a very simple system: that of a single particle confined to some region of space. We confine our particle inside some box and occasionally take measurements of this system. The set of possible trajectories of the particle around the box constitutes all the imaginable behaviors of this particle; we might describe this space of behaviors mathematically as $\text{Maps}(I, \text{Box})$, where $I \subset \mathbb{R}$ denotes the time interval over which we conduct the experiment. We say the set of possible behaviors forms a space of *fields* on the timeline of the particle.

The behavior of our theory is governed by an action functional, which is a function on $\text{Maps}(I, \text{Box})$. The simplest case typically studied is the massless free field theory, whose value on a trajectory $f : I \to \text{Box}$ is

$$S(f) = \int_{t \in I} (f(t), \ddot{f}(t))\, \mathrm{d}t.$$

Here we use $(-,-)$ to denote the usual inner product on $\mathbb{R}^n$, where we view the box as a subspace of $\mathbb{R}^n$, and $\ddot{f}$ to denote the second derivative of f in the time variable t.

The aim of this section is to outline the structure one would expect the observables – that is, the possible measurements one can make of this system – should satisfy.

1.1.2 Classical Mechanics

Let us start by considering the simpler case where our particle is treated as a classical system. In that case, the trajectory of the particle is constrained to be in a solution to the Euler–Lagrange equations of our theory, which is a differential equation determined by the action functional. For example, if the action functional governing our theory is that of the massless free theory, then a map $f : I \to \text{Box}$ satisfies the Euler–Lagrange equation if it is a straight line. (Since we are just trying to provide a conceptual narrative here, we will assume that Box becomes all of $\mathbb{R}^n$ so that we do not need to worry about what happens at the boundary of the box.)

We are interested in the observables for this classical field theory. Since the trajectory of our particle is constrained to be a solution to the Euler–Lagrange equation, the only measurements one can make are functions on the space of solutions to the Euler–Lagrange equation.

If $U \subset \mathbb{R}$ is an open subset, we will let Fields(U) denote the space of fields on U, that is, the space of maps $f : U \to \text{Box}$. We will let

$$\mathrm{EL}(U) \subset \mathrm{Fields}(U)$$

denote the subspace consisting of those maps $f : U \to \text{Box}$ that are solutions to the Euler–Lagrange equation. As U varies, $\mathrm{EL}(U)$ forms a sheaf of spaces on $\mathbb{R}$.

We will let $\mathrm{Obs}^{cl}(U)$ denote the commutative algebra of functions on $\mathrm{EL}(U)$ (the precise class of functions we consider discussed later). We will think of $\mathrm{Obs}^{cl}(U)$ as the collection of observables for our classical system that depend only upon the behavior of the particle during the time period U. As U varies, the algebras $\mathrm{Obs}^{cl}(U)$ vary and together constitute a cosheaf of commutative algebras on $\mathbb{R}$.

1.1.3 Measurements in Quantum Mechanics

The notion of measurement is fraught in quantum theory, but we will take a very concrete view. Taking a measurement means that we have a physical measurement device (e.g., a camera) that we allow to interact with our system for a period of time. The measurement is then how our measurement device has changed due to the interaction. In other words, we *couple* the two physical systems so that they interact, then decouple them and record how the measurement device has modified from its initial condition. (Of course, there is a symmetry in this situation: both systems are affected by their interaction, so a measurement inherently disturbs the system under study.)

The *observables* for a physical system are all the imaginable measurements we could take of the system. Instead of considering all possible observables, we might also consider those observables that occur within a specified time period. This period can be specified by an open interval $U \subset \mathbb{R}$.

Thus, we arrive at the following principle.

Principle 1. For every open subset $U \subset \mathbb{R}$, we have a set $\mathrm{Obs}(U)$ of observables one can make during U.

Our second principle is a minimal version of the linearity implied by, e.g., the superposition principle.

Principle 2. The set $\mathrm{Obs}(U)$ is a complex vector space.

We think of $\mathrm{Obs}(U)$ as being the collection of ways of coupling a measurement device to our system during the time period U. Thus, there is a natural map $\mathrm{Obs}(U) \to \mathrm{Obs}(V)$ if $U \subset V$ is a shorter time interval. This means that the space $\mathrm{Obs}(U)$ forms a precosheaf.

1.1.4 Combining Observables

Measurements (and so observables) differ qualitatively in the classical and quantum settings. If we study a classical particle, the system is not noticeably disturbed by measurements, and so we can do multiple measurements at the same time. (To be a little less sloppy, we suppose that by refining our measuring devices, we can make the impact on the particle as small as we would like.) Hence, on each interval J we have a commutative multiplication map $\mathrm{Obs}(J) \otimes \mathrm{Obs}(J) \to \mathrm{Obs}(J)$. We also have maps $\mathrm{Obs}(I) \otimes \mathrm{Obs}(J) \to \mathrm{Obs}(K)$ for every pair of disjoint intervals I, J contained in an interval K, as well as the maps that let us combine observables on disjoint intervals.

For a quantum particle, however, a measurement typically disturbs the system significantly. Taking two measurements simultaneously is incoherent, as the measurement devices are coupled to each other and thus also affect each other, so that we are no longer measuring just the particle. Quantum observables thus do not form a cosheaf of commutative algebras on the interval. However, there are no such problems with combining measurements occurring at different times. Thus, we find the following.

Principle 3. If U, U' are disjoint open subsets of $\mathbb{R}$, and $U, U' \subset V$ where V is also open, then there is a map

$$\star : \mathrm{Obs}(U) \otimes \mathrm{Obs}(U') \to \mathrm{Obs}(V).$$

If $O \in \mathrm{Obs}(U)$ and $O' \in \mathrm{Obs}(U')$, then $O \star O'$ is defined by coupling our system to measuring device O during the period U and to device O' during the period U'.

Further, there are maps for a finite collection of disjoint time intervals contained in a long time interval, and these maps are compatible under composition of such maps. (The precise meaning of these terms is detailed in Section 3.1 in Chapter 3.)

1.1.5 Perturbative Theory and the Correspondence Principle

In the bulk of this two-volume book, we will be considering perturbative quantum theory. For us, this adjective "perturbative" means that we work over the base ring $\mathbb{C}[[\hbar]]$, where at $\hbar = 0$ we find the classical theory. In perturbative theory, therefore, the space $\mathrm{Obs}(U)$ of observables on an open subset U is a $\mathbb{C}[[\hbar]]$-module, and the product maps are $\mathbb{C}[[\hbar]]$-linear.

The correspondence principle states that the quantum theory, in the $\hbar \to 0$ limit, must reproduce the classical theory. Applied to observables, this leads to the following principle.

Principle 4. The vector space $\mathrm{Obs}^q(U)$ of quantum observables is a flat $\mathbb{C}[[\hbar]]$-module such that modulo $\hbar$, it is equal to the space $\mathrm{Obs}^{cl}(U)$ of classical observables.

These four principles are at the heart of our approach to quantum field theory. They say, roughly, that the observables of a quantum field theory form a factorization algebra, which is a quantization of the factorization algebra associated to a classical field theory. The main theorem presented in this two-volume book is that one can use the techniques of perturbative renormalization to construct factorization algebras perturbatively quantizing a certain class of classical field theories (including many classical field theories of physical and mathematical interest). As we have mentioned, this first volume focuses on the general theory of factorization algebras and on simple examples of field theories; this result is derived in Volume 2.

1.1.6 Associative Algebras in Quantum Mechanics

The principles we have described so far indicate that the observables of a quantum mechanical system should assign, to every open subset $U \subset \mathbb{R}$, a vector space $\mathrm{Obs}(U)$, together with a product map

$$\mathrm{Obs}(U) \otimes \mathrm{Obs}(U') \to \mathrm{Obs}(V)$$

if U, U' are disjoint open subsets of an open subset V. This is the basic data of a factorization algebra (see Section 3.1 in Chapter 3).

It turns out that in the case of quantum mechanics, the factorization algebra produced by our quantization procedure has a special property: it is *locally constant* (see Section 6.4 in Chapter 6). This means that the map $\mathrm{Obs}((a,b)) \to \mathrm{Obs}(\mathbb{R})$ is an isomorphism for any interval (a,b). Let A denote the vector space $\mathrm{Obs}(\mathbb{R})$; note that A is canonically isomorphic to $\mathrm{Obs}((a,b))$ for any interval (a,b).

The product map

$$\mathrm{Obs}((a,b)) \otimes \mathrm{Obs}((c,d)) \to \mathrm{Obs}((a,d))$$

when $a < b < c < d$, becomes, via this isomorphism, a product map

$$m : A \otimes A \to A.$$

The axioms of a factorization algebra imply that this multiplication turns A into an associative algebra. As we will see in Section 4.2 in Chapter 4, this associative algebra is the Weyl algebra, which one expects to find as the algebra of observables for quantum mechanics of a particle moving in $\mathbb{R}^n$.

This kind of geometric interpretation of algebra should be familiar to topologists: associative algebras are algebras over the operad of little intervals in $\mathbb{R}$, and this is precisely what we have described. As we explain in Section 6.4 in Chapter 6, this relationship continues and so our quantization theorem produces many new examples of algebras over the operad E_n of little n-discs.

An important point to take away from this discussion is that *associative algebras appear in quantum mechanics because associative algebras are connected with the geometry of* $\mathbb{R}$. There is no fundamental connection between associative algebras and any concept of "quantization": associative algebras appear only when one considers one-dimensional quantum field theories. As we will see later, when one considers topological quantum field theories on n-dimensional space–times, one finds a structure reminiscent of an E_n-algebra instead of an E_1-algebra.

Remark: As a caveat to the strong assertion in the preceding (and jumping ahead of our story), note that for a manifold of the form $X \to \mathbb{R}$, one can push forward a factorization algebra Obs on $X \times \mathbb{R}$ to a factorization algebra $\pi_* \mathrm{Obs}$ on $\mathbb{R}$ along the projection map $\pi : X \times \mathbb{R} \to \mathbb{R}$. In this case, $\pi_* \mathrm{Obs}((a,b)) = \mathrm{Obs}(X \times (a,b))$. Hence, a quantization of a higher dimensional theory will produce, via such pushforwards to $\mathbb{R}$, deformations of associative algebras, but knowing only the pushforward is typically insufficient to reconstruct the factorization algebra on the higher dimensional manifold. ◊

1.2 A Preliminary Definition of Prefactorization Algebras

Below (see Section 3.1 in Chapter 3) we give a more formal definition, but here we provide the basic idea. Let M be a topological space (which, in practice, will be a smooth manifold).

Definition 1.2.1 A prefactorization algebra $\mathcal{F}$ on M, taking values in cochain complexes, is a rule that assigns a cochain complex $\mathcal{F}(U)$ to each open set $U \subset M$ along with

(i) A cochain map $\mathcal{F}(U) \to \mathcal{F}(V)$ for each inclusion $U \subset V$.
(ii) A cochain map $\mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_n) \to \mathcal{F}(V)$ for every finite collection of open sets where each $U_i \subset V$ and the U_i are disjoint.
(iii) The maps are compatible in a certain natural way. The simplest case of this compatibility is that if $U \subset V \subset W$ is a sequence of open sets, the map $\mathcal{F}(U) \to \mathcal{F}(W)$ agrees with the composition through $\mathcal{F}(V)$).

Remark: A prefactorization algebra resembles a precosheaf, except that we tensor the cochain complexes rather than taking their direct sum. ◊

The observables of a field theory, whether classical or quantum, form a prefactorization algebra on the space–time manifold M. In fact, they satisfy a kind of local-to-global principle in the sense that the observables on a large open set are determined by the observables on small open sets. The notion of a factorization algebra (Section 6.1 in Chapter 6) makes this local-to-global condition precise.

1.3 Prefactorization Algebras in Quantum Field Theory

The (pre)factorization algebras of interest in this book arise from perturbative quantum field theories. We have already discussed in Section 1.1 how factorization algebras appear in quantum mechanics. In this section we will see how this picture extends in a natural way to quantum field theory.

The manifold M on which the prefactorization algebra is defined is the space–time manifold of the quantum field theory. If $U \subset M$ is an open subset, we will interpret $\mathcal{F}(U)$ as the collection of observables (or measurements) that we can make that depend only on the behavior of the fields on U. Performing a measurement involves coupling a measuring device to the quantum system in the region U.

One can bear in mind the example of a particle accelerator. In that situation, one can imagine the space–time M as being of the form $M = A \times (0, t)$, where A is the interior of the accelerator and t is the duration of our experiment.

In this situation, performing a measurement on an open subset $U \subset M$ is something concrete. Let us take $U = V \times (\varepsilon, \delta)$, where $V \subset A$ is some small region in the accelerator and where (ε, δ) is a short time interval. Performing a measurement on U amounts to coupling a measuring device to our accelerator in the region V, starting at time ε and ending at time δ. For example, we could imagine that there is some piece of equipment in the region V of the accelerator, which is switched on at time ε and switched off at time δ.

1.3.1 Interpretation of the Prefactorization Algebra Axioms

Suppose that we have two different measuring devices, O_1 and O_2. We would like to set up our accelerator so that we measure both O_1 and O_2.

There are two ways we can do this. We can insert O_1 and O_2 into disjoint regions V_1, V_2 of our accelerator. Then we can turn O_1 and O_2 on at any times we like, including for overlapping time intervals.

If the regions V_1, V_2 overlap, then we cannot do this. After all, it doesn't make sense to have two different measuring devices at the same point in space at the same time.

However, we could imagine inserting O_1 into region V_1 during the time interval (a, b); and then removing O_1, and inserting O_2 into the overlapping region V_2 for the disjoint time interval (c, d).

These simple considerations immediately suggest that the possible measurements we can make of our physical system form a prefactorization algebra. Let $\mathrm{Obs}(U)$ denote the space of measurements we can make on an open subset $U \subset M$. Then, by combining measurements in the way outlined in the preceding text, we would expect to have maps

$$\mathrm{Obs}(U) \otimes \mathrm{Obs}(U') \to \mathrm{Obs}(V)$$

whenever U, U' are disjoint open subsets of an open subset V. The associativity and commutativity properties of a prefactorization algebra are evident.

1.3.2 The Cochain Complex of Observables

In the approach to quantum field theory considered in this book, the factorization algebra of observables will be a factorization algebra of cochain complexes. That is, Obs assigns a cochain complex $\mathrm{Obs}(U)$ to each open U. One can ask for the physical meaning of the cochain complex.

We will repeatedly mention observables in a gauge theory, as these kinds of cohomological aspects are well known for such theores.

It turns out that the "physical" observables will be $H^0(\mathrm{Obs}(U))$. If $O \in \mathrm{Obs}^0(U)$ is an observable of cohomological degree 0, then the equation $\dot{O} = 0$ can often be interpreted as saying that O is compatible with the gauge symmetries of the theory. Thus, only those observables $O \in \mathrm{Obs}^0(U)$ that are closed are physically meaningful.

The equivalence relation identifying $O \in \mathrm{Obs}^0(U)$ with $O + \dot{O}'$, where $O' \in \mathrm{Obs}^{-1}(U)$, also has a physical interpretation, which will take a little more work to describe. Often, two observables on U are physically indistinguishable (that is, they cannot be distinguished by any measurement one can perform). In the example of an accelerator outlined earlier, two measuring devices are equivalent if they always produce the same expectation values, no matter how we prepare our system, or no matter what boundary conditions we impose.

As another example, in the quantum mechanics of a free particle, the observable measuring the momentum of a particle at time t is equivalent to that measuring the momentum of a particle at another time t'. This is because, even at the quantum level, momentum is preserved (as the momentum operator commutes with the Hamiltonian).

From the cohomological point of view, if $O, O' \in \mathrm{Obs}^0(U)$ are both in the kernel of the differential (and thus "physically meaningful"), then they are equivalent in the sense described previously if they differ by an exact observable.

It is a little more difficult to provide a physical interpretation for the other cohomology groups $H^n(\mathrm{Obs}(U))$. The first cohomology group $H^1(\mathrm{Obs}(U))$ contains anomalies (or obstructions) to lifting classical observables to the quantum level. For example, in a gauge theory, one might have a classical observable that respects gauge symmetry. However, it may not lift to a quantum observable respecting gauge symmetry; this happens if there is a nontrivial anomaly in $H^1(\mathrm{Obs}(U))$.

The cohomology groups $H^n(\mathrm{Obs}(U))$, when $n < 0$, are best interpreted as symmetries, and higher symmetries, of observables. Indeed, we have seen that the physically meaningful observables are the closed degree 0 elements of $\mathrm{Obs}(U)$. One can construct a simplicial set, whose n-simplices are closed and degree 0 elements of $\mathrm{Obs}(U) \otimes \Omega^*(\Delta^n)$. The vertices of this simplicial set are observables, the edges are equivalences between observables, the faces are equivalences between equivalences, and so on.

The Dold–Kan correspondence (see Theorem A.2.7 in Appendix A) tells us that the nth homotopy group of this simplicial set is $H^{-n}(\mathrm{Obs}(U))$. This allows us to interpret $H^{-1}(\mathrm{Obs}(U))$ as being the group of symmetries of the trivial observable $0 \in H^0(\mathrm{Obs}(U))$, and $H^{-2}(\mathrm{Obs}(U))$ as the symmetries of the identity symmetry of $0 \in H^0(\mathrm{Obs}(U))$, and so on.